

Linked Open Usable Data for Cultural Heritage: Perspectives on Community Practices and Semantic Interoperability

Dissertation

zur Erlangung der Würde eines Doktors der Philosophie

vorgelegt der Philosophisch-Historischen Fakultät
der Universität Basel

von

Julien Antoine Raemy

aus

Plaffeien/Freiburg

Bern, 2024

Copyprint Bern Simab

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Historischen Fakultät der Universität Basel, auf Antrag
von Prof. Dr. Peter Fornaro, Prof. Dr. Walter Leimgruber und Dr. Robert Sanderson.

Basel, den 18. November 2024
Der Dekan Prof. Dr. Martin Lengwiler

Acknowledgements

While a thesis is often perceived as an individual effort, I could not have succeeded without the ideas, encouragement, and support of those around me, and the opportunity to ask questions, no matter how naive they may have been. I would like to extend my sincere thanks to all the individuals and collectives mentioned here, and apologise if I have inadvertently missed anyone.

To begin with, I would like to thank Peter Fornaro, my first supervisor, who contacted me shortly after my Master's thesis defence in August 2020 to possibly offer me a PhD position at the University of Basel's Digital Humanities Lab (DHLab) within the Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) research project. I am also deeply grateful to Walter Leimgruber, my second supervisor, who, together with Peter, interviewed me in Basel at the end of 2020 and officially welcomed me into the research project. Secondly, I would like to thank Robert Sanderson, who kindly agreed to become my third supervisor in early 2022. His help with technical aspects and our many insightful discussions proved very beneficial. I am also grateful for his hospitality during my fortnight stay at Yale University in New Haven in May 2023.

I was fortunate to receive financial support for most of my PhD through my work at the DHLab, as PIA was supported by the Swiss National Science Foundation (SNSF) under the Sinergia funding programme (contract number CRSII5_193788)¹. Additionally, I had the opportunity to work at the Swiss National Data and Service Center for the Humanities (DaSCH), an infrastructure primarily funded by the SNSF. This support not only provided the financial means to pursue my research but also granted me access to a collaborative and well-supported environment. It enabled me to engage with current tools and methodologies, attend key conferences, and collaborate with esteemed colleagues. These opportunities have been crucial to the advancement of my work and my professional development.

I would like to thank everyone involved in the PIA project and the Cultural Anthropology Switzerland (CAS) photo archives. All those directly and indirectly involved are to be properly acknowledged: Walter Leimgruber, Nicole Peduzzi, Regula Anklin, Murielle Cornut, Birgit Huber, Fabienne Lüthi, Véronique Hilfiker, Miriam Kull, Laura Citaku, Andrea Suter, Charlotte Nachtsheim, Alexandra Tschakert, Peter Fornaro, Heiko Schuldt, Vera Chiquet, Adrian Demleitner, Fabian Frei, Christoph Rohrer, Florian Spiess, Ulrike Felsing, Daniel Schoeneck, Lukas Zimmer, Max Frischknecht, and Tobias Hodel. I am especially grateful to the fellow PhD candidates and software developers with whom I have forged close friendships, whether through *Traubenlese* at Fabienne's, 'Dumpling Days' in Biel and Basel, our retreat in Alpstein, looking after my cats while I was on holiday, or having regular lunches together.

My gratitude also goes to the DHLab team, including head and professor Lukas Rosenthaler, as well as the administrative support provided by Irène Kälin Heiniger and Gabriela Kuster Vettiger. Many thanks to past and present fellow PhD students from the DHLab, particularly Béatrice Gauvain, Cristina Loi, Victoria G. D. Landau, Marian Clemens Manz, and Christian Weber. I would also like to thank the staff of DaSCH, particularly Rita Gautschy, for her tireless support, which enabled us to organise a IIF 3D event and working meeting in Basel in October 2023.

¹ PIA - Grant number 193788: <https://data.snf.ch/grants/grant/193788>

Many thanks to Julia Büchele, coordinator of the University of Basel’s Graduate School of Social Sciences (G3S), for her support and organisation of fruitful writing days. From the University of Basel, I also want to thank Alain Müller for inspiring me to read Bruno Latour’s works, and both him and Marion Schulze for their enlightening seminar on Donna Haraway. Special mention goes to Ivan Giangreco and the collaborators and students from the University of Basel’s Department of Mathematics and Computer Science who created and maintain the Basilea L^AT_EX template² that I used and slightly amended to compile the PDF version of this thesis.

I am thankful as well to the Swiss Federal Archives, particularly Audray Sauvage and Stefan Nellen, for placing their trust in me and hiring me as a Data Scientist before the completion of my doctorate. Their flexibility allowed me to work three days a week, and later four, enabling me to finish writing my thesis.

I would also like to acknowledge the following individuals I met at seminars, workshops, or conferences, who inspired me and with whom I was able to discuss anything and everything: Yumeng Hou, Giacomo Allinata, Valentine Bernasconi, Dario Negueruela del Castillo, Rémi Petitpierre, Rahaf Orabi, Shabnam Pasandideh, Pauline Junginger, Karina Rodriguez Echavarria, Zoe Schubert, Lozana Rossenova, Raphaël Grolimund, Baptiste de Coulon, Sol Martinez Demarco, Jasmin Schmidlin, Brittny Lapierre, Tugce Karatas, Moritz Mähr, Rebecca Rochat, Anna Janka, Margherita Schellino, Immacolata Iaccarino, Elina Leblanc, Pauline Jacsont, Marie Barras, Nicola Carboni, Simon Gabay, Anaïs Guillem, Antoine Gros, Francesco Beretta, David Knecht, Elias Kreyenbühl, John Kunze, Ben Brumfield, Sara Brumfield, David Newbury, Rashmi Singhal, Lise Foket, Glen Robson, Caitlin Perry, Erin Canning, Kelly Davis, Alyssa Loera, John McQuaid, Bree Midavaine, George Bruseker, Emmanuelle Delmas-Glass, David Beaudet, Edward Anderson, Tanya Gray, Kevin Page, Claire Knowles, Daniel Sissman, and Tom Crane.

My sincere gratitude is due to everyone participating in the International Image Interoperability Framework (IIIF) and Linked Art communities, especially those who responded to the survey I conducted in 2023. I have always felt welcomed by both communities and it is a privilege to be involved in them. This sense of community was also reflected in the warm welcome I received at Yale, and I would like to thank all those who agreed to be interviewed in the context of the LUX initiative. A special mention goes to the ‘Maddhoughna’ group, with whom we held so many Zoom calls, especially during the COVID-19 pandemic. I would therefore like to express my profound appreciation to the following people: Rebecca Hirsch, Anne McLaughlin, Stacey Redick, Caylin Smith, Aliya Reich, Andy Irving and Richard Palmer.

At this point, I would particularly like to extend my thanks to the people who inspired me and believed in me before I set off on this PhD journey, and who gave me the courage to do it. I am thinking in particular of the people I knew when I was working in Oxford between 2008 and 2009, long before I started my studies, as well as René Schneider.

Finally, I would like to express my heartfelt thanks to my family and friends. Their genuine interest in my PhD progress and timely encouragement provided the perfect balance of support and distraction, allowing me to take necessary breaks and recharge. Last but certainly not least, I extend my deepest thanks to my beloved partner, Davina, for her unwavering support and constant presence throughout this journey. Her encouragement and understanding have been truly invaluable.

² L^AT_EX Template for Bachelor and Master thesis (University of Basel, Department of Mathematics and Computer Science): <https://github.com/ivangiangreco/basilea-latex>

Abstract

Digital technologies have fundamentally transformed the ways in which Cultural Heritage (CH) collections are accessed and engaged with. Linked Open Usable Data (LOUD) specifications, including the International Image Interoperability Framework (IIIF) Presentation API 3.0, Linked Art, and the Web Annotation Data Model (WADM), have emerged as web standards to facilitate the description and dissemination of these valuable resources. Despite the widespread adoption of IIIF, the implementation of LOUD specifications, especially in combination, remains challenging. This is particularly true in the development and assessment of infrastructures, or sites of assemblage, that support these standards.

The research is guided by two perspectives, namely community practices and semantic interoperability. The former aims to assess how organisations, individuals, and apparatuses engage with and contribute to the consensus-making processes surrounding LOUD. By examining community practices, the social fabrics of the LOUD ecosystem can be better understood. The latter perspective explores how data can be made meaningful to machines in a standardised and interoperable approach that promotes the exchange of well-formed information. This work is grounded as part of the SNSF-funded research project Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) (2021-2025), which aims to develop a citizen science platform around three photographic collections from the Cultural Anthropology Switzerland (CAS) archives. The theoretical framework is firmly rooted in Actor-Network Theory (ANT), as one of the aims of the thesis is to describe the collaborative structures of the LOUD ecosystem and to highlight the importance of non-human actors.

Beyond the implementation of LOUD standards within the PIA research project, the empirical research includes an analysis of the social fabrics of the IIIF and Linked Art communities, as well as an investigation of LUX, Yale Collections Discovery platform. The research aims to identify the socio-technical requirements for developing specifications consistent with LOUD design principles. It also seeks to understand how the implementation of LOUD standards within PIA contributes to recognising their potential benefits and limitations in terms of facilitating data reuse and wider participation. The thesis further examines the implementation strategies, challenges and outcomes of Yale University's large-scale deployment of LOUD standards, focusing on how ensuring the consistency of Linked Art and IIIF resources within the LUX platform contributes to the CH domain.

The core methodology of this thesis is an actor and practice-centred inquiry, focusing on the detailed examination of specific cosmologies within LOUD-driven communities, PIA, and LUX. This approach aims to unravel the intricate web of cultural processes and constellations through a micro-perspective, rich in empirical evidence.

Key empirical findings indicate that LOUD improves the discoverability and integration of data in CH, which requires community-driven consensus on model interoperability. Significant challenges include engaging marginalised groups, sustaining long-term engagement, and balancing technological and social considerations. The strategic use of technology and the capture of digital materiality are critical, yet LOUD presents challenges in terms of resource investment, data consistency and the wider implementation of complex patterns.

LOUD should clearly lead efforts to improve the accessibility and usability of CH data. The community-driven methodology of IIF and Linked Art inherently promotes collaboration and transparency, making these standards critical tools in the ongoing evolution of data management. Even for projects and institutions that do not adopt these specifications, LOUD's socio-technical practices provide crucial insights into effective digital stewardship and community engagement strategies.

Keywords: Actor-Network Theory; Community of Practice; Cultural Anthropology Switzerland; Cultural Heritage; Digital Infrastructure; International Image Interoperability Framework; Knowledge Practices; Linked Art; Linked Data; LUX; Participatory Archives; Photographic Archives; Semantic Interoperability; Web Annotation Data Model

Acronyms

3D	three-dimensional
AAC	American Art Collaborative
AACR	Anglo-American Cataloguing Rules
AAT	Art & Architecture Thesaurus
AHRC	Arts and Humanities Research Council
AI	Artificial Intelligence
ABox	Assertion Box
ANT	Actor-Network Theory
API	Application Programming Interface
AR	Augmented reality
ARK	Archival Resource Key
AS	Activity Streams
ASR	Automatic speech recognition
ASV	The Atlas of Swiss Folklore
A/V	Audio/Video
BibFrame	Bibliographic Framework
BnF	Bibliothèque nationale de France
BO	Boundary Object
CARE	Collective Benefit, Authority to Control, Responsibility, Ethics
CAS	Cultural Anthropology Switzerland
CERN	European Organization for Nuclear Research
CCO	Cataloging Cultural Objects
CDWA	Categories for the Description of Works of Art
CH	Cultural Heritage
CHAOS	Community Health Analytics in Open Source Software
CHI	Cultural Heritage Institution
CHIT	Yale's Cultural Heritage Information Technology
CLI	Command-line interface

CLIR	Council on Library and Information Resources
CIDOC	ICOM International Committee for Documentation
CIDOC-CRM	CIDOC Conceptual Reference Model
CNI	Coalition for Networked Information
CoCo	IIIF Coordinating Committee
CORS	Cross-Origin Resource Sharing
CoP	Communities of Practice
CRM-SIG	CIDOC-CRM Special Interest Group
CSC	Yale's Collections and Scholarly Communications
CSV	Comma-separated values
DAM	Digital asset management
DaSCH	Swiss National Data and Service Center for the Humanities
DBIS	Databases and Information Systems
DH	Digital Humanities
DHLab	University of Basel's Digital Humanities Lab
DICOM	Digital Imaging and Communications in Medicine
DL	Description logic
DLF	Digital Library Federation
DMSTech	Digital Manuscript Technical Working Group
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
DSP	DaSCH Service Platform
EAD	Encoded Archival Description
EDM	Europeana Data Model
EODEM	Exhibition Object Data Exchange Model
ETL	Extract, Transform, Load
Exec	IIIF Executive Committee
FAIR	Findable, Accessible, Interoperable, Reusable
FDO	FAIR Digital Object
FOL	First-order logic
FORTH	Foundation for Research & Technology – Hellas
FLOSS	Free/Libre and Open Source Software
FRBR	Functional Requirements for Bibliographic Records

GAFAM	Google, Apple, Facebook, Amazon, Microsoft
GB	Gigabytes
GND	Gemeinsame Normdatei
GLAM	Galleries, Libraries, Archives, and Museums
GUI	Graphical user interface
G3S	University of Basel's Graduate School of Social Sciences
HAL	JSON Hypertext Application Language
HIMMS	Healthcare Information and Management Systems Society
HKB	Bern Academy of the Arts
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
HTR	Handwritten Text Recognition
ICA	International Council on Archives
ICBS	International Committee of the Blue Shield
ICOM	International Council on Museums
ICOMOS	International Council on Monuments and Sites
IETF	Internet Engineering Task Force
IFLA	International Federation of Library Associations and Institutions
IIIF	International Image Interoperability Framework
IIIF-C	IIIF Consortium
ILS	Integrated Library System
IRI	Internationalized Resource Identifier
ISAD(G)	General International Standard Archival Description
ITS	Yale's Information Technology Services
JSON	JavaScript Object Notation
JSONL	JSON Lines
JSON-LD	JavaScript Object Notation for Linked Data
KB	Knowledge Base
KG	Knowledge Graph
KOS	Knowledge Organisation Systems
KR	Knowledge Representation

LANL	Los Alamos National Laboratory
LCSH	Library of Congress Subject Headings
LGBTQ+	lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual, and more
LIDO	Lightweight Information Describing Objects
LINCS	Linked Infrastructure for Networked Cultural Scholarship
LLM	Large Language Models
LoA	Levels of Abstraction
LoC	Library of Congress
LOD	Linked Open Data
LOUD	Linked Open Usable Data
LRM	IFLA Library Reference Model
MARC	Machine-Readable Cataloging
METS	Metadata Encoding and Transmission Standard
ML	Machine Learning
MODS	Metadata Object Description Schema
NDL	National Diet Library
NER	Named Entity Recognition
NLP	Natural Language Processing
N2T	Name-To-Thing
OA	Open access
OAC	Open Annotation Collaboration
OAI-ORE	Open Archives Initiative Object Exchange and Reuse
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCR	Optical Character Recognition
OSD	OpenSeadragon
OWL	Web Ontology Language
PB	Petabytes
PI	Philosophy of Information
PIA	Participatory Knowledge Practices in Analogue and Digital Image Archives
PID	Persistent Identifier

PREMIS	PREservation Metadata: Implementation Strategies
R2RML	RDB to RDF Mapping Language
RAIL	Responsible AI Licenses
RDA	Resource Description and Access
RDF	Resource Description Framework
RDFS	RDF Schema
REST	Representational state transfer
RFC	Request for Comments
RiC	Records in Context
RML	RDF Mapping Language
RTI	Reflectance Transformation Imaging
SARI	Swiss Art Research Infrastructure
SDI	standard definition of (semantic) information
SEO	Search Engine Optimization
SHACL	Shapes Constraint Language
SHARED	Sustainable, Harvestable, Available, Reconciled, Enriched, Diverse
SIPI	Simple Image Interface Presentation
SK	Situated Knowledges
SKOS	Simple Knowledge Organization System
SLMS	System-LoA-Model-Structure
SNA	Social Network Analysis
SNSF	Swiss National Science Foundation
SOA	service-oriented architecture
SPARQL	SPARQL Protocol and RDF Query Language
SSFS	Swiss Society for Folklore Studies
STEM	Science, technology, engineering, and mathematics
STS	Science and Technology Studies
TB	Terabytes
TBox	Terminological Box
TLS	Transport Layer Security
TGN	Thesaurus of Geographic Names
TMS	Gallery Systems' The Museum System

TRC	IIIF Technical Review Committee
TSG	Technical Specification Group
ULAN	Union List of Artist Names
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
UV	Universal Viewer
UX	User experience
VIAF	Virtual International Authority File
VQA	Visual Question Answering
VR	virtual reality
VRA	Visual Resource Association
WADM	Web Annotation Data Model
WEMI	Work, Expression, Manifestation, Item
W3C	World Wide Web Consortium
XML	Extensible Markup Language
YCBA	Yale Center for British Art
YPM	Yale Peabody Museum of Natural History
YUAG	Yale University Art Gallery
YUL	Yale University Library

Table of Contents

Acknowledgements	iii
Abstract	v
Acronyms	vii
List of Figures	xvii
List of Tables	xx
List of Code Snippets	xxi
1 Introduction	1
2 Context	4
2.1 PhD Title	4
2.2 The Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) Research Project	5
2.2.1 An Interdisciplinary Endeavour	5
2.2.2 Photographic Collections/Archives as Anchors	6
2.2.3 Project Vision	10
2.3 Contribution to PIA and its Relevance to the Thesis	11
2.4 Involvement within the IIF and Linked Art communities	12
3 Interlinking Cultural Heritage Data	13
3.1 What Makes Cultural Heritage Data Stand Out?	14
3.1.1 Cultural Heritage	14
3.1.1.1 Tangible Heritage	15
3.1.1.2 Intangible Heritage	17
3.1.1.3 Natural Heritage	18
3.1.1.4 Cultural Heritage Data	20
3.1.2 Representation and Embodiment of Cultural Heritage Data	23
3.1.2.1 Materiality	23
3.1.2.2 Digitised Cultural Heritage and Digital Heritage	24
3.1.3 Collectives and Apparatuses	25
3.2 Cultural Heritage Metadata	29
3.2.1 Data about Data	29
3.2.2 Types and Functions	31
3.2.3 Standards	33
3.2.3.1 Library Metadata Standards	34
3.2.3.2 Archival Metadata Standards	37
3.2.3.3 Museum and Gallery Metadata Standards	39
3.2.3.4 Cross-domain Metadata Standards	42

3.2.4	Knowledge Organisation Systems	44
3.2.4.1	Folksonomy	45
3.2.4.2	Taxonomy	45
3.2.4.3	Thesaurus	45
3.2.4.4	Ontology	46
3.3	Trends, Movements, and Principles	48
3.3.1	Current and Emerging Technological Trends in Cultural Heritage	49
3.3.1.1	Linked Data	50
3.3.1.2	Big Data	50
3.3.1.3	Artificial Intelligence	52
3.3.2	Scientific Movements and Guiding Principles	56
3.3.2.1	Towards Open Scholarship	56
3.3.2.2	Citizen Science, Citizen Humanities	59
3.3.2.3	FAIR Data Principles	62
3.3.2.4	CARE Principles for Indigenous Data Governance	65
3.3.2.5	Collections as Data	66
3.4	Open Web Platform and Linked Data	67
3.4.1	Web Architecture	69
3.4.2	The Semantic Web	72
3.4.3	Linked Data Principles	81
3.4.4	Deployment Schemes for Open Data	82
3.4.5	Linked Open Data in the Cultural Heritage Domain	83
3.5	Linked Open Usable Data	86
3.5.1	LOUD Design Principles	87
3.5.2	LOUD: Systematic Review	89
3.5.3	International Image Interoperability Framework (IIIF)	93
3.5.3.1	Inception	93
3.5.3.2	IIIF Community	94
3.5.3.3	IIIF Specifications	96
3.5.3.4	IIIF Ecosystem	106
3.5.4	Web Annotation Data Model	109
3.5.5	Linked Art	111
3.5.5.1	History	113
3.5.5.2	Linked Art Data Model	114
3.5.5.3	Linked Art API and Standards for Linked Art	120
3.5.5.4	Community and Implementations	122
3.5.6	Implications and Opportunities for the Cultural Heritage Domain and the Humanities	123
3.6	Characterising Community Practices and Semantic Interoperability	123
3.6.1	Community Practices	124
3.6.2	Semantic Interoperability	127
3.7	Summary and Preliminary Insights	130
3.7.1	Cultural Heritage Data	131
3.7.2	Cultural Heritage Metadata	131
3.7.3	Trends, Movements, and Principles	132
3.7.4	Linked Open Data	132
3.7.5	LOUD: Design Principles, Communities, Standards	132
3.7.6	Community Practices and Semantic Interoperability	133
3.7.7	The Case for Cultural Heritage Data	133

4	Exploring Relationships through an Actor-Network Theory Lens	134
4.1	Implosion of the Boundaries: Non-humans have Agency	136
4.2	Assembling the Collective	138
4.3	The Translation Process	141
4.4	Epistemological Foundations	144
5	Research Scope and Methodology	147
5.1	Research Questions	147
5.2	Hypotheses	148
5.3	Objectives	148
5.4	Methodology	149
5.4.1	Outline of the Research Methods	149
5.4.1.1	Assessment of Practices in IIIF and Linked Art Communities	151
5.4.1.2	Deployment of LOUD within PIA	153
5.4.1.3	Review of a Large-Scale LOUD Implementation and Data Consistency within LUX	154
5.4.2	Overall Limitations	155
5.4.3	Ethical Considerations	156
5.5	Contribution to the (Digital) Humanities	157
6	The Social Fabrics of IIIF and Linked Art	158
6.1	Exploring the Foundations	159
6.2	Consensus Building and Advocacy Initiatives	160
6.2.1	Types of Meetings and Engagement Opportunities	161
6.2.1.1	IIIF Discovery TSG Meetings	162
6.2.1.2	Linked Art Meetings	164
6.2.2	Collaborative Development through GitHub Documentation	165
6.2.3	Community Guidelines and Best Practice Efforts	168
6.2.3.1	IIIF Cookbook Recipes	168
6.2.3.2	Linked Art Patterns	170
6.3	Community Participation and LOUD Standards: Adoption and Impact	171
6.4	Synthesis and Insights	175
6.5	Perspectives	176
6.5.1	Enhancing Community Engagement	176
6.5.2	Technical Development	176
6.5.3	Integration Practices	177
7	PIA as a Laboratory	178
7.1	Data and Infrastructure Baseline at the Project Onset	178
7.2	LOUD Standards in an Explorative Setting	182
7.2.1	Metadata Extraction and Integration	183
7.2.2	CAS Data Model	184
7.2.3	Implementation of the IIIF APIs and Web Annotation	186
7.2.4	Deployment of APIs and Vocabulary to Support Semantic Interoperability	193
7.3	Synthesis and Insights	197
7.4	Perspectives	200
8	Yale’s LUX and LOUD Consistency	203
8.1	Large-Scale Deployment of LOUD at Yale	204
8.1.1	Background	204

8.1.2	Technology Stack	205
8.1.3	Interviews	209
8.1.3.1	Themes and Questions	209
8.1.3.2	Findings	211
8.2	Linked Art and IIF Consistency	216
8.2.1	Ensuring Compliance: Tools and Validators	216
8.2.2	LUX Data Consistency	218
8.2.2.1	Linked Art Consistency	218
8.2.2.2	IIF Consistency	221
8.3	Synthesis and Insights	225
8.3.1	Cross-Unit Collaboration	225
8.3.2	As Consistent as Required	225
8.4	Perspectives	226
9	Discussion	228
9.1	Empirical Findings	228
9.2	Evaluation: Abstracting LOUD	233
9.3	Retrospective: Trudging like an Ant	235
10	Conclusion	237
	Bibliography	241
	Appendix A AI Tools	277
	Appendix B Supplementary Figures	278

List of Figures

2.1	Map from the SGV_05 Collection Relating to Question 93 Showing Walks and Excursions at Pentecost. The Atlas of Swiss Folklore (ASV). CAS. CC BY-NC 4.0	7
2.2	A photo Album Page from the SGV_10 Collection, Bearing the Following Inscription: <i>Botanische Excursion ins Wallis, Pfingster 1928</i> . SGV_10A_00031.015. Kreis Family. CAS. CC BY-NC 4.0	8
2.3	Picture from the SGV_12 Collection Showing Walkers Looking at the Timetable Train. [Wanderer studieren den Fahrplan in der Bahnhofhalle]. Lucerne, 1938. Ernst Brunner. SGV_12N_00716. CAS. CC BY-NC 4.0	9
3.1	Snapshot from the Swisscovery Platform Showing the Bibliographic Record of Zeng and Qin [2022]	30
3.2	The Functional Requirements for Bibliographic Records (FRBR) Conceptual Framework. Adapted from Zou et al. [2018, p. 36]	35
3.3	Overview of Relationships in IFLA Library Reference Model (LRM) [Riva et al. 2017, p. 86]	36
3.4	Open Archival Information System (OAIS) Functional Model Diagram by Mathieualexhache [2021]	37
3.5	Global Overview of the Core Entities Defined by the Records in Context (RiC) Conceptual Model. Slightly Adapted from https://github.com/ICA-EGAD/RiC-O	38
3.6	CIDOC Conceptual Reference Model (CIDOC-CRM) Top-Level Categories by Bruseker et al. [2017, p. 112]	41
3.7	CIDOC-CRM Family of Models. Diagram Done and Provided by Maria Theodoridou (Institute of Computer Science, FORTH)	43
3.8	Overview of the Structures and Functions of Knowledge Organisation Systems (KOS) [Zeng 2008, p. 161]	44
3.9	Knowledge Base (KB) Representation System Based on Patrón et al. [2011, p. 205]	47
3.10	Open Science Elements, Redrawn Slide from Presentation of Ana Persic [Morrison 2021, citing Persic [2021]]	57
3.11	The Findable, Accessible, Interoperable, Reusable (FAIR)ification Process. Adapted from GO FAIR [2016]	64
3.12	The FAIR Digital Object (FDO) Model [European Commission. Directorate General for Research and Innovation. 2018]	65
3.13	Overlap and Difference between Uniform Resource Identifier (URI), Uniform Resource Locator (URL), and Uniform Resource Name (URN)	70
3.14	Types of Web Architectures: Client-server Model, Three Tier Model, service-oriented architecture (SOA)	71
3.15	The Semiotic Triangle by [Ogden and Richards 1930].	73
3.16	Example of an Resource Description Framework (RDF) Graph	74
3.17	Tweaked Semantic Web Technology Layer Cake by Idehen [2017]	75
3.18	Structure of Web Ontology Language (OWL) 2 [W3C OWL Working Group 2012]	79
3.19	Five-Star Deployment Scheme for Open Data	82
3.20	Example of a LOUD-Driven Infrastructure [Felsing et al. 2023, p. 43]	88
3.21	LOUD References Ordered by Year and α Sub-categories	90

3.22	LOUD References as Venn Diagrams for Each of the Six Intersections	91
3.23	IIIF Image Application Programming Interface (API) Order of Implementation [Appleby et al. 2020a]	98
3.24	IIIF Presentation API Data Model [Appleby et al. 2020b]	99
3.25	The Interaction of the Core IIIF APIs in Mirador	102
3.26	<i>Cologne, Fondation Martin Bodmer, Cod. Bodmer 600a: The Life of Buddha, First Book (Shaka no Honji, jō)</i> Embedded in Jalava, a Customised Client from Durham University Compliant with the IIIF Content State API	105
3.27	Example of a IIIF Architecture Implementing the Six APIs	105
3.28	Example of a IIIF Ecosystem of Software	106
3.29	Screenshot of the Cultural Japan Platform in the IIIF Guides	108
3.30	WADM High-level Overview [Sanderson et al. 2017]	109
3.31	Abstraction and Implementation Level of Linked Art. Adapted from Sanderson [2021]	112
3.32	Linked Art from 50,000 Feet [Raemy et al. 2023a, adapted from Sanderson [2020b]]	115
3.33	Linked Art Digital Integration: Possible Description of Available IIIF resources	119
4.1	Relationships Between Residual Categories, Boundary Objects (BOs), and Standardisation Attempts. Adapted from Star [2010]	140
4.2	Exploring Levels of Abstraction (LoA) in LOUD Ecosystems: Integrating Societal Implications with the Concepts of Immediacy and Hypermediacy	143
4.3	The System-LoA-Model-Structure (SLMS) According to Floridi [2008]	144
4.4	A Synpoiesis of Theories: ANT Entangled with Situated Knowledges (SK), BO, and Philosophy of Information (PI)	145
5.1	Research Outline Visual Template	152
5.2	Research Outline for Assessing the IIIF and Linked Art Practices	153
5.3	Research Outline for Deploying LOUD Standards in the PIA Project	154
5.4	Research Outline for Reviewing LOUD Implementation and Data Consistency within LUX	155
6.1	Overview of the IIIF Community	160
6.2	Stacked Bar Chart Showing IIIF Discovery Technical Specification Group (TSG) Meeting Attendance	163
6.3	Stacked Bar Chart Showing Linked Art Meeting Attendance	164
6.4	Illustrative Diagram of the GitHub-Based Process for Approving and Integrating Use Cases	166
6.5	Detailed Illustration of the GitHub-Enabled Approval Process by the IIIF Technical Review Committee (TRC) for the Image and Presentation APIs 3.0	167
6.6	Exploring IIIF Cookbook Recipes by Support with Gephi Lite	169
6.7	Exploring Linked Art Patterns by Category with Gephi Lite	170
6.8	Year of Involvement in the IIIF Community [Raemy 2023a, p. 20]	172
6.9	Situating IIIF to Movements and Principles [Raemy 2023a, p. 21]	173
6.10	Situating Linked Art to Movements and Principles [Raemy 2023a, p. 22]	174
7.1	Overview of the Legacy Swiss Society for Folklore Studies (SSFS) Data Model	181
7.2	Synoptic View of the PIA Infrastructure: Showcasing its Connection to DaSCH Service Platform (DSP) and the CAS Photo Archive Website	183
7.3	Overview of the Updated CAS Data Model	185
7.4	Overview of the Legacy Automatic Annotation Workflow [Cornut et al. 2023, p. 12]	189
7.5	SGV_12N_08589 Displayed on Mirador with Machine-generated Annotations	191
7.6	Suggested Top-level Entities of the PIA Linked Art API	194
7.7	Boxes Holding Printed Photographs by Ernst Brunner According to his System of Classification	195

7.8	Ernst Brunner Thesaurus: Index Card of HY 59	196
7.9	The Main Actors and their Associations around the PIA Linked Art API	199
7.10	Overview of the IIF Presentation API Workflow Proof of Concept at DaSCH	201
8.1	Snapshot of the LUX Graphical user interface (GUI) (23 May 2024)	205
8.2	LUX Data Pipeline and Architecture	206
8.3	Annotated LUX Data Pipeline and Architecture	208
8.4	LUX Interviews: Key Topics and Their Relationships	215
8.5	Getty Vocabularies Term Presence in the Yale Center for British Art (YCBA) and Yale University Art Gallery (YUAG) Digital Collections	220
8.6	Shared Terminology in YCBA and YUAG via Getty Vocabularies	220
8.7	Validation Results of IIF Presentation API Resources from Yale (Sample)	224
B.1	The FRBR Conceptual Framework (Enlarged View)	279
B.2	Overview of Relationships in LRM (Enlarged View)	280
B.3	The FDO Model (Enlarged View)	281
B.4	Example of a IIF Architecture Implementing the Six APIs (Enlarged View)	282
B.5	Linked Art Digital Integration (Enlarged View)	283
B.6	Overview of the IIF Community (Enlarged)	284
B.7	Stacked Bar Chart Showing IIF Discovery TSG Meeting Attendance (Enlarged View) . . .	285
B.8	Stacked Bar Chart Showing Linked Art Meeting Attendance (Enlarged View)	286
B.9	Exploring IIF Cookbook Recipes by Support with Gephi Lite (Enlarged View)	287
B.10	Exploring Linked Art Patterns by CIDOC-CRM or Linked Art Class with Gephi Lite (En- larged View)	288
B.11	Overview of the Legacy SSFS Data Model (Enlarged View)	289
B.12	Synoptic View of the PIA Infrastructure (Enlarged View)	290
B.13	Overview of the Updated CAS Data Model (Enlarged View)	291
B.14	Overview of the Legacy Automatic Annotation Workflow (Enlarged View)	292
B.15	Overview of the IIF Presentation API Workflow Proof of Concept at DaSCH (Enlarged View)	293

List of Tables

2.1	PIA Team Core Members	6
3.1	Types of Metadata According to Riley [2017, p 7]	32
3.2	Summary of Database Search Results on LOUD	89
3.3	Summary of the LOUD Systematic Review	92
5.1	Empirical Chapters: Research Axes and Scope Overview	150
7.1	Exported Images Along with Their Respective Time-Based Metadata from Salsah (SGV_10 and SGV_12)	184
7.2	Overview of Classes: Legacy and Updated Data Models and its Mapping to the Knora Base Ontology	186
7.3	Relevance of IIF Cookbook recipes to our resources. Adapted from Raemy and Demleitner [2023]	189
8.1	LUX Interview Questions	210
8.2	Validation Summary of Linked Art Example Endpoints and LUX Instances	219
8.3	YCBA - Usage of Getty Controlled Vocabularies	221
8.4	YUAG - Usage of Getty Controlled Vocabularies	221
8.5	Base Yale Manifest URLs and Respective Versions Used for Validation	223
A.1	Artificial Intelligence (AI) Tools used for the thesis	277

List of Code Snippets

3.1	MARC21 Record of Zeng and Qin [2022] in the Swisscovery Platform	34
3.2	Example of a Lightweight Information Describing Objects (LIDO) Object in Extensible Markup Language (XML) from Lindenthal et al. [2023]	39
3.3	Taxonomy Hierarchy	45
3.4	RDF/XML Serialisation of the RDF Graph	74
3.5	Turtle Serialisation of the RDF Graph	77
3.6	SPARQL Protocol and RDF Query Language (SPARQL) Query Against the RDF Graph	77
3.7	Example of using RDF Schema (RDFS) to Determine Class Membership and Property Definitions	78
3.8	Example of a OWL-based Ontology	79
3.9	Example of a Shapes Constraint Language (SHACL) Document in Turtle that can be Used to Validate the RDF Graph	80
3.10	Example of a IIIF Manifest	101
3.11	Example of a Level 2 Activity Compliant with the IIIF Change Discovery API [Appleby et al. 2021]	104
3.12	Example of Basic Annotation According to WADM [Sanderson et al. 2017]	110
3.13	Example of an Annotation According to WADM with a motivation and a purpose [Sanderson et al. 2017]	110
3.14	The Night Watch: Representation in the Linked Art Data Model Including URI; Class; Label; and Classification	116
3.15	The Night Watch: Representation in the Linked Art Data Model Including Names; Identifiers; and Statements	117
3.16	The Night Watch: Representation in the Linked Art Data Model Including Activities and Timespans	118
7.1	SPARQL Query Executed on the DaSCH Fuseki Instance to Retrieve and Construct IIIF Image API URLs	188
7.2	Example of Machine-generated Annotations in a IIIF setting from PIA	190
7.3	IIIF Manifest with Two Images Painted onto the Same Canvas	191
7.4	PIA IIIF Change Discovery API	192
7.5	Modelling in Linked Art of a HumanMadeObject from the Kreis Family Collection	193
7.6	Turtle Snippet of the Ernst Brunner Thesaurus: Brauch	196
8.1	Command to Transcribe the Interviews into Audio Files Using Whisper	210
8.2	Command to Convert Text files into Markdown with New Lines	211
8.3	Validating a Linked Art Resource: Example Concept	218
8.4	Validating a Linked Art Resource: Concept on LUX	218
8.5	CURL Command Against a Local Instance of the IIIF Presentation API Validator . . .	222
8.6	JSON Responses against the Presentation API Validator: V2.1 Resource	222
8.7	JSON Responses against the Presentation API Validator: V3.0 Resource	222

1

Introduction

Since its inception in 2011, the International Image Interoperability Framework (IIIF) has revolutionised³ the accessibility of image-based resources. Initially driven by the needs of manuscript scholars, IIIF focused on two-dimensional images, but has since expanded to encompass a wide range of image-based resources, including audiovisual materials and, in the near future, three-dimensional (3D) images. Similarly, Linked Art, formally established in 2017, initially concentrated on art museum objects but has since broadened its scope to model a variety of Cultural Heritage (CH) entities, leveraging CIDOC Conceptual Reference Model (CIDOC-CRM), a renowned ontology in the museum and Digital Humanities (DH) space. Both initiatives aim to break down silos: IIIF focuses on improving the presentation of digital objects, while both initiatives enhance their dissemination. Together, they make CH data more accessible through IIIF and more meaningful through Linked Art for machines. These efforts have primarily benefited the CH domain.

A key commonality is that the main Application Programming Interfaces (APIs) these communities create align with the LOUD design principles, either intentionally or empirically demonstrated through use cases. These principles enable software developers to develop compliant tools and services without needing to fully understand Resource Description Framework (RDF), a syntax for representing information on the web. Additionally, they may not need to grasp all Linked Open Data (LOD) principles, which promote the interlinking of data from diverse datasets using tools like Knowledge Organisation Systems (KOS) such as thesauri. Web Annotation Data Model (WADM), a World Wide Web Consortium (W3C) standard, is also recognised as a Linked Open Usable Data (LOUD) specification. It provides a framework for creating interoperable annotations on web resources, facilitating the linking and sharing of data across different platforms and applications. These LOUD design principles include the right abstraction for the audience, few barriers to entry, comprehensibility by introspection, documentation with working examples, and the use of many consistent patterns rather than few exceptions. Additionally, both IIIF and Linked Art are driven by vibrant communities, mainly comprising Galleries, Libraries, Archives, and Museums (GLAM) and higher education institutions.

While the standards and principles discussed have broad applications, it is important to clarify the scope of this dissertation. This work does not focus on Knowledge Graphs (KGs) by assessing triplestores – databases specifically designed to store and retrieve triples, which are the fundamental data structures in RDF. Similarly, it does not deal with evaluating SPARQL engines, which are specifically designed to query KGs. Additionally, this dissertation does not address the intersection of Machine Learning (ML) and IIIF, or the ontological reasoning of Linked Art.

³ Throughout this dissertation, British English spelling conventions are predominantly observed. However, there are instances of American English spelling where direct quotations from sources are used as well as referring to names of institutions, standards, or concepts.

Instead, this dissertation concentrates on LOUD, the consistency of its standards, design principles and the vibrant communities behind it. It examines JavaScript Object Notation for Linked Data (JSON-LD) serialisation efforts and the crucial intersection required to establish robust semantic interoperability baselines between presentation and semantic layers. It also presents real-world use case implementations, both on a small scale in a laboratory and flexible space within the Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) research project, and on a large scale at Yale, exemplified by the LUX platform that provides access to (meta)data from Yale University Library (YUL), Yale Center for British Art (YCBA), Yale University Art Gallery (YUAG), and Yale Peabody Museum of Natural History (YPM).

The focus is therefore on digital infrastructures capable of delivering JSON-LD files from the above specifications, which are primarily, though not exclusively, CH resources. It is more about the different actors – both human and non-human – that create and maintain these interconnected systems and the dynamic interactions that sustain them. The deployment of various LOUD specifications addresses the need for semantic interoperability between CH resources and disparate datasets by establishing a standardised approach to representing and linking data, ensuring that information can be seamlessly shared and understood across different platforms and contexts.

This dissertation seeks to carve out a distinct niche by addressing an often-overlooked aspect of IIF and Linked Art. IIF is sometimes perceived and studied merely as a service or an appendix, with the content it delivers taking precedence. However, this PhD thesis positions IIF as a first-class citizen worthy of in-depth study. Similarly, Linked Art, despite its potential and its relatively recent establishment, has been the subject of very few scholarly papers. This gap underscores the significance of LOUD in this context. Furthermore, this thesis elevates Linked Art to a position of primary importance, recognising its significance and advocating for its thorough examination. To thoroughly study LOUD and its adherence to design principles, it is essential to immerse ourselves actively in both communities – an approach I have embraced for years. The thesis also emphasises the importance of participatory efforts and collaboration between research projects, which typically have shorter lifespans, and memory institutions, which need to implement technical standards as a *lingua franca*. In doing so, it reveals the mediating role of LOUD in advancing the heritage sphere. To truly understand IIF, Linked Art, and to a lesser extent WADM, it is crucial to examine the social fabrics and consensus decision-making of each community. Among these considerations are how the specifications can be implemented pragmatically, and how the standards can support the implementation and maintenance of more extensive semantic interoperability efforts.

The significance of this research lies in highlighting the commitment and diligence of the individuals and organisations that make up both the IIF and Linked Art communities. It aims to demonstrate that community-driven practices, such as those exemplified by IIF and Linked Art, have a potential that goes beyond the mere sharing of digital objects and their associated metadata. The more people who embrace these approaches and implement the associated specifications, the more society as a whole will benefit. Furthermore, this research illustrates that IIF is no longer limited to two-dimensional images, that Linked Art is not restricted to artworks, and that WADM is a simple, content-agnostic standard that can be easily integrated into a range of systems. This adaptability is a strength of LOUD standards, which are designed to be simple yet effective. LOUD can serve a variety of purposes, primarily rooted in CH, but with the potential to extend its benefits to other sectors. The true beauty of LOUD lies in its ability to foster networking opportunities and transparent socio-technical practices, demonstrating its value beyond mere technical implementation.

By emphasising these aspects, this dissertation highlights the wider impact of LOUD in promoting semantic interoperability and enhancing collaborative efforts within the heritage field and beyond. In addition, the implementation of standards through PIA underlines the potential for similar participatory or citizen science projects, while the LUX initiative serves as an illustrative example of robust infrastructure and cross-unit engagement. These examples demonstrate the practical applications and far-reaching implications of adopting LOUD standards in different contexts.

This dissertation is structured across ten chapters, each building upon the previous ones up to Chapter 5 to provide a comprehensive understanding of the research. These initial chapters lay the foundation of the study, establishing the context, theoretical framework, and methodological approaches. After this foundational section, Chapters 6, 7, and 8 present empirical studies that, while interconnected, can be read independently if desired. These chapters offer detailed insights into specific aspects of the research and can be appreciated on their own or as part of the broader narrative.

The thesis continues with Chapter 2, which extends this introduction by providing more information about the research setting, specifically PIA. Chapter 3 follows with an extensive literature review, offering a comprehensive overview of methods to interlink CH data. Next, Chapter 4 presents the theoretical framework, conceptualised as a toolbox and firmly rooted in Actor-Network Theory (ANT), guiding the analysis and discussion throughout the dissertation. Following this, Chapter 5 details the research scope and methodology, explaining the approaches and methods employed in the study.

Moving on to the empirical work, Chapter 6 sheds light on the social fabrics of IIIIF and Linked Art, exploring the communities and practices that underpin these initiatives. Chapter 7 then examines the implementation of LOUD standards within PIA, highlighting the practical aspects and challenges encountered. This is followed by Chapter 8, which focuses on the LUX initiative at Yale, examining the underlying governance and interdepartmental ownership of the Yale Collections Discovery platform. The discussion of findings is presented in Chapter 9, where the results from the empirical chapters are synthesised and analysed in relation to the theoretical framework. Finally, Chapter 10 concludes the thesis, summarising the key insights and contributions of the research while outlining potential directions for future study.

2

Context

In this chapter, I will set the stage for my PhD thesis by providing important background information. First, in Section 2.1, I will explain why I chose the title for my thesis. This will give you an understanding of the main focus and the direction of my research. Next, in Section 2.2, I will describe the PIA research project, which is central to my work. This section will cover the project's goals, significance, and overall framework. In Section 2.3, I will detail my specific contributions to the PIA project. I will emphasise how my work fits into the larger project and its importance to my thesis. Finally, in Section 2.4, I will talk about my active participation in the IIF and Linked Art communities. This section will highlight how my involvement in these communities has influenced my research and its broader implications.

2.1 PhD Title

I chose the title 'Linked Open Usable Data for Cultural Heritage: Perspectives on Community Practices and Semantic Interoperability' as it encapsulates the essence of my research focus but I could have indeed chosen other ones.

During the initial stages of my research, multiple working titles were explored to capture the diverse facets of my interests and objectives. If I was quite sure about having 'Linked Open Usable Data' in the title after the third iteration⁴, I was quite unsure of what should follow and if a subtitle was actually needed at all. Amidst this dynamic progression, the underlying theme of my research remained steadfast – to delve into the transformative potential of LOUD for CH.

I also opted to maintain 'Cultural Heritage' in the title of my thesis subsection. While 'Digital Heritage' holds its appeal, my choice reflects a broader narrative that acknowledges the crucial role of Cultural Heritage Institutions (CHIs) and spotlighting the multifaceted nature of heritage preservation, encapsulating both its digital facets and the essential contribution of individuals and institutions in curating, interpreting, and making heritage accessible.

As for the subtitle, while I do explore Communities of Practice (CoP) as defined by [Lave and Wenger \[1991\]](#) and [Wenger \[2011\]](#) through investigating the social fabrics of the IIF and Linked Art communities, my main interest lies in the broader application of LOUD for describing and interlinking CH resources. Thus, I decided to opt for the more generic 'Community Practices' as the first axis or perspective.

For the second perspective, I wanted to see how semantic interoperability can be achieved through standards adhering to the LOUD design principles, as they seem to be key enablers for seamless collaboration and knowledge exchange among practitioners. There was a time in my research when I envisaged

⁴ On <https://phd.julsraemy.ch>, you can see all the working titles. Throughout the course of my research journey, a series of working titles underwent regular revisions until February 2023, when the title finally attained stability.

decoupling ‘semantics’ and ‘interoperability’, perceiving them as two distinct dimensions. However, what really captivates me is the unification of these factors to facilitate collective reasoning for both humans and machines.

In summary, this title reflects my enthusiasm for using web-based and community-driven technologies to transform the way we understand, share and value CH.

2.2 The Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) Research Project

I undertook my doctoral studies within the scope of the Participatory Knowledge Practices in Analogue and Digital Image Archives (PIA) research project financed by the Swiss National Science Foundation (SNSF) under their Sinergia funding scheme from February 2021 to January 2025⁵. The project aimed to analyse the interplay of participants, epistemological orders and the graphical representation of information and knowledge in relation to three photographic collections from Cultural Anthropology Switzerland (CAS). It sought to bring together the world of data and things in an interdisciplinary manner, exploring the phases of the analogue and digital archive from a cultural anthropological, technical and design research perspective [Felsing et al. 2023, p. 42]. As part of this endeavour, interfaces were developed to enable the collaborative indexing and use of photographic archival records [Chiquet et al. 2023, p. 110]. I discuss in more detail the interdisciplinary components and briefly introduce the people involved in the project in Section 2.2.1, then talk about the photographic collections that were the overarching narrative of the research in Section 2.2.2, and lastly in Section 2.2.3, the vision that we had put together.

2.2.1 An Interdisciplinary Endeavour

The project, divided in three interdisciplinary teams, was led by the University of Basel through the Institute for Cultural Anthropology and European Ethnology⁶ (Team A) and the University of Basel’s Digital Humanities Lab (DHLab)⁷ in collaboration with the Databases and Information Systems (DBIS) group (Team B) as well as by the Bern Academy of the Arts (HKB)⁸, an art school and department of the Bern University of Applied Sciences (Team C) [Felsing et al. 2023, p. 43]. Table 2.1 lists the people who contributed to the project, broken down by the three teams and their particular perspectives.

In addition to the core members mentioned in Table 2.1, there were several other individuals who made significant contributions to the CAS photo archive and PIA. Their efforts were invaluable, particularly in digitising, cataloguing, and indexing the collections. Although their names are not listed in the table, I would like to express my gratitude to the following people who played a crucial role: Miriam Kull, Laura Citaku, Andrea Suter, Charlotte Nachtsheim, and Alexandra Tschakert. The project’s success owed a great deal to the collective efforts of these individuals and many others who worked behind the scenes, including Véronique Hilfiker, who provided essential administrative support.

⁵ SNSF Data Portal - Grant number 193788: <https://data.snf.ch/grants/grant/193788>

⁶ *Seminar für Kulturwissenschaft und Europäische Ethnologie*: <https://kulturwissenschaft.philhist.unibas.ch/>

⁷ DHLab: <https://dhlab.philhist.unibas.ch/>

⁸ HKB: <https://www.hkb.bfh.ch/>

Table 2.1: PIA Team Core Members

Perspective	People
A) Anthropological	Prof. Dr. Walter Leimgruber, Team Leader and Dissertation Supervisor Dr. Nicole Peduzzi, Photographic Restoration and Digitisation Supervisor Regula Anklin, Conservation and Restoration Specialist (project partner at Anklin & Assen) Murielle Cornut, PhD Candidate in Cultural Anthropology Birgit Huber, PhD Candidate in Cultural Anthropology Fabienne Lüthi, PhD Candidate in Cultural Anthropology
B) Technical	Prof. Dr. Peter Fornaro, Team Leader and Dissertation Supervisor Prof. Dr. Heiko Schuldt, Dissertation Supervisor (project partner at the University of Basel) Dr. Vera Chiquet, Postdoctoral Researcher Adrian Demleitner, Software Developer (2021-2023) Fabian Frei, Software Developer (2023-2025) Christoph Rohrer, Software Developer (2023-2025) Julien A. Raemy, PhD Candidate in Digital Humanities Florian Spiess, PhD Candidate in Computer Science
C) Communicative	Dr. Ulrike Felsing, Team Leader and Dissertation Supervisor Prof. Dr. Tobias Hodel, Dissertation Supervisor (project partner at the University of Bern) Daniel Schoeneck, Research Fellow Lukas Zimmer, Designer (project partner at A/Z&T) Max Frischknecht, PhD Candidate in Digital Humanities

2.2.2 Photographic Collections/Archives as Anchors

Cultural Anthropology Switzerland (CAS)⁹, established in 1896, has been actively involved in the collection and preservation of photographs since the 1930s [SGV 2021]. These photographs, both amateur and professional, originate from all of the language regions of Switzerland (German, French, Italian, Romansh) and extend to various countries across the globe [Peduzzi 2018]. The CAS archives is listed as a cultural asset of national importance (category A) in the Swiss Cultural Property Inventory¹⁰ and has been undergoing extensive conservation and restoration since 2014, the year in which CAS launched a significant photographic project, focusing on the processing of approximately 105,000 photographic objects from seventeen collections [Graf et al. 2019]. This project encompassed a wide array of essential tasks, including meticulous conservation, expert restoration, digitisation, and meticulous documentation of the photographic objects within the Salsah system¹¹, originally built by the DHLab and now maintained through the Swiss National Data and Service Center for the Humanities (DaSCH) who preserve the photographic archives and their metadata. In 2018, the project reached a milestone when the photographs were made available to the public online¹².

CAS has historically been engaged in active collaborations that bridge the academic research and the public sphere, primarily through traditional analogue methods. The PIA project was created with the intention of exploring the complexities inherent in both analogue and digital approaches, and to encourage and investigate these collaborative endeavours between academia and the wider public. As such, PIA represents a paradigm shift within the scope of projects associated with or supported by CAS, facilitating the seamless integration of digital tools to explore multiple facets of participation

⁹ Cultural Anthropology Switzerland (CAS), formerly Swiss Society for Folklore Studies (SSFS) (English) *Empirische Kulturwissenschaft Schweiz*, formerly *Schweizerische Gesellschaft für Volkskunde* (German) *Anthropologie Culturelle Suisse*, formerly *Société suisse des traditions populaires* (French) *Anthropologia Culturale Svizzera*, formerly *Società svizzera per le tradizioni popolari* (Italian)

¹⁰ Inventory of Cultural Property of National and Regional Importance (see Canton BS): <https://www.babs.admin.ch/en/aufgabenbabs/kgs/inventar.html>

¹¹ Salsah: <https://www.salsah.org/>

¹² CAS photo archive website: <https://archiv.sgv-sstp.ch/>

The newer CAS multimedia archive is also accessible at <https://www.ekws.ch/de/archiv>

and engagement. This transformative endeavour embodies a profound exploration of new intersections where scholarly endeavours intertwine with the active involvement of citizens.

PIA drew on three collections: one focusing on scientific cartography and titled ‘Atlas of Swiss Folklore’ (*Atlas der Schweizerischen Volkskunde*), a second from the estate of the photojournalist Ernst Brunner (1901–1979), and a third collection consisting of vernacular photography which was owned by the Kreis Family (1860–1970).

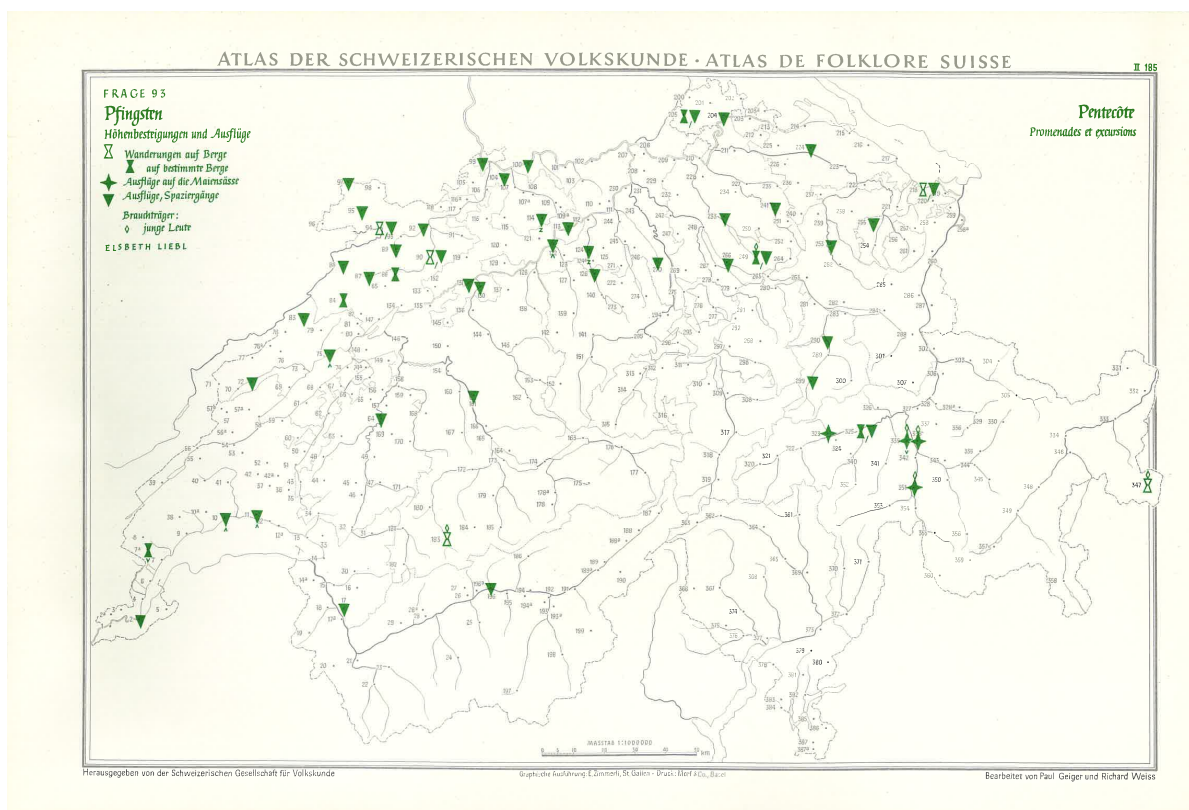


Figure 2.1: Map from the SGV_05 Collection Relating to Question 93 Showing Walks and Excursions at Pentecost. The Atlas of Swiss Folklore (ASV). CAS. CC BY-NC 4.0

SGV_05 The Atlas of Swiss Folklore (ASV) consists of 292 maps and 1000 pages of commentary published from 1950 to 1995 — an example of such a map is shown in Figure 2.1. This collection was commissioned by the CAS to do an extensive survey of the Swiss population in the 1930s and 1940s on many issues pertaining, for instance, to everyday life, local laws, superstitions, celebrations or labour [Weiss 1940]. The contents were compiled by researchers and by people who were described as ‘laymen’¹³. Questions were asked about everyday habits, community rights, work, trade, superstitions, and many other topics [Schmoll 2009a,b]. This collection offers a snapshot of everyday life in Switzerland right before the beginning of a modernisation process that fundamentally changed lifestyles in all areas during the postwar period. A digitised version of the ASV would not only allow the results of that time to be enriched with further findings [Felsing and Frischknecht 2021], but would also make transparent how knowledge was generated in cartographic form through a complex process along different types of media and actors. The restoration, digitisation, cataloguing and indexing efforts took all part through-

¹³ The considerable size of the ASV collection, which includes over 90,000 analogue objects, reflects not just the work of the main authors but also the contributions from numerous explorers and additional material beyond the maps and primary publications.

out PIA under the supervision of Birgit Huber, who extensively based her doctoral research on this particular collection [see [Huber 2023](#)].

SGV_10 Kreis Family comprises approximately 20,000 loose photographic objects, where a quarter of them are organised and kept in 93 photo albums — as illustrated by Figure 2.2, from a wealthy Basel-based family and spanning from the 1850s to the 1980s. This private collection was acquired by CAS in 1991. The collection, which originally arrived in banana cases and was enigmatic due to the lack of clear organisation or accompanying information from the family, posed significant challenges. Despite these initial hurdles, CAS undertook meticulous efforts to catalogue and preserve its contents [[Felsing and Cornut 2024](#), p. 42]. The pictures were taken by studio photographers as well as by family members themselves. The Kreis Family collection represents a typical example of urban bourgeois culture and gives a comprehensive insight into the development of private photography over the course of a century [[Pagenstecher 2009](#)]. The photographic materials and formats are very diverse, ranging from prints to negatives, small, medium or large format photographs, black and white or colour. The collection also encompasses many photographic techniques, from the one-off daguerreotypes and ferrotypes, to the glass-based negatives that could be reproduced *en masse*, to the modern paper prints. While some of the albums and loose images were restored and digitised during the 2014 project, much of this work was completed during PIA and overseen by Murielle Cornut, whose doctoral investigation was centred on the study of photo albums [see [Cornut 2023](#)].



Figure 2.2: A photo Album Page from the SGV_10 Collection, Bearing the Following Inscription: *Botanische Excursion ins Wallis, Pfingsten 1928*. SGV_10A-00031.015. Kreis Family. CAS. CC BY-NC 4.0

SGV_12 Ernst Brunner is a donation of about 48,000 negatives and 20,000 prints to the CAS archives from Ernst Brunner, a self-taught photojournalist, who lived from 1901 to 1979 and who documented mainly in the 1930s and 1940s a wide range of folkloristic themes — as shown by Figure 2.3. He is one of the most important photographers of the era and one of the most outstanding visual

chroniclers of Swiss society [Pfrunder 1995]. His photographs show rural lifestyles, but also urban motifs. In his late work, he led the documentation and research on farmhouses in a specific Swiss district, a project initiated by CAS. Before Ernst Brunner became an independent photojournalist in the mid-1930s, he worked as a carpenter, influenced by the ideas of the *Bauhaus* and *Neues Bauen* movements. This can also be seen in the aesthetics and formal language of his photography. If all the black and white negatives were digitised and recorded between 2014 and 2018, the digitisation of prints, which is a selection done by Ernst Brunner, was conducted at the end of the PIA research project. The latter was supervised by Fabienne Lüthi, whose PhD was about organisational systems and knowledge practices in the Ernst Brunner Collection.



Figure 2.3: Picture from the SGV_12 Collection Showing Walkers Looking at the Timetable Train. [Wanderer studieren den Fahrplan in der Bahnhofhalle]. Lucerne, 1938. Ernst Brunner. SGV_12N_00716. CAS. CC BY-NC 4.0

Whereas for each of the PhD Candidates in Cultural Anthropology, a particular collection was assigned to them and its content was to varying degrees part of their subject of study, this was not exactly the same for the PhD Candidates in DH, including myself, and in Computer Science. Put differently, we had relative leeway in terms of what interested us in each or all of these three photographic collections. In my case, I briefly explain my contribution to the project more in Section 2.3 and then in Chapter 7 as part of the empirical portion of my thesis focusing on the deployment of LOUD specifications using

the three CAS photographic collections.

Florian Spiess focused on the use of virtual reality (VR) through *vitivr*, a multimedia retrieval system developed by the DBIS research group at the Department of Mathematics and Computer Science [Spiess and Schuldt 2022, Spiess and Stauffiger 2023, Spiess et al. 2024]. His work included experiments with PIA-related collections, such as the creation of virtual galleries clustered according to content-based similarity [see Peterhans et al. 2022]. In the case of Max Frischknecht, his doctoral research centred on generative design¹⁴, a methodology to visualise dynamic cultural archives. He mostly worked on the ASV collection and on a mapping tool which is a cartographic visualisation designed to explore the CAS photographic archives [see Frischknecht 2022, Huber and Frischknecht 2024a].

It should also be mentioned that not only did we use the three collections of the CAS photographic archives within the project, but that both formal and informal meetings took place most commonly within the photographic archives at the Spalenvorstadt premises in the old *Gewerbemuseum* and later either at the ‘*Provisorium*’ on *Allschwilerstrasse*, though less frequently, or at *Rheinsprung* where the *Institute for Cultural Anthropology and European Ethnology* is located. This meant that there was a strong and sometimes blurred entanglement between those involved in the archives and the PIA core team members.

2.2.3 Project Vision

Between December 2021 and March 2022, we worked together to develop and finalise a vision for the project¹⁵. It includes seven key priorities, or pillars, which were meant to strengthen the interdisciplinary perspectives of PIA. Although ambitious, these elements were of paramount importance to us and served as a guiding blueprint for all PIA activities. Hereafter is a modified version of the vision¹⁶ taken from Cornut et al. [2023, p. 4].

1. **Accessibility** by developing open interfaces and offering the possibility of expanding the archive and turning it into an instrument of current research that collects and evaluates knowledge with the participation of other users (Citizen Science).
2. **Heterogeneity** by making visible where, why and under what circumstances the objects were created, how they were handled and what path they have taken to get to and in the archive. We work on visualisations that take into account the heterogeneous character of archival materials and make their respective biographies visible.
3. **Materiality** by conveying the material properties of the objects: they have front and back sides, inscriptions, traces, development errors, they are transparent, multi-layered or fabric-covered. They tell of their origin, use, and peculiarities. We want to make this knowledge accessible and understandable in digital form. To this end, we also consider the necessary infrastructure involved in the creation as part of their narrative: the restoration, the relocation, the indexing, the storage devices, the research tools, the display medium, as well as the process of repro-photography.
4. **Interoperability** as a crucial component and which will be done by supporting digital means that allow different stakeholders to freely access and interact with the project’s data. Both humans and machines can use, contribute to, correct and annotate the existing data in an open and interoperable manner, thus encouraging exchange and the creation of new knowledge. To do this, we use web-based standards that are widely adopted in the cultural heritage field.

¹⁴ Max Frischknecht’s PhD: <https://phd.maxfrischknecht.ch/>

¹⁵ PIA project website: <https://about.participatory-archives.ch/>

¹⁶ The vision of the PIA project was first written in German and then translated into English and French.

5. **Affinities** by leveraging data models and pattern recognition which can uncover semantic relationships between entities that were previously incomplete or difficult for users to access. Using specific interfaces and visualisations, we make it possible to explore digital assets and discover forms of relationships and similarities between images.
6. **Artificial Intelligence (AI)** that facilitates automated searches for simple image attributes such as colour, shapes, and localisation of image components. It should also become possible to recognise texts and object types for extracting metadata.
7. **Bias Management** by taking into account that associated metadata was human-made¹⁷ and thus is never objective. Collections and their metadata reflect biases or focus narrowly on selected areas and perceptions. Machines working on the basis of such data automatically reproduce the implicit biases in decision-making due to so-called biased algorithms. Therefore, understanding the data used for training and the algorithms applied for decision making is crucial to ensure the integrity of the application of these technologies in archives. We take ethical issues into account when using AI and visualisations, because the higher the awareness of a possible bias, the faster it can be detected or brought up for consideration with users.

As my thesis is notably concerned with semantic interoperability, **Interoperability** and **Affinities** are of particular importance to my PhD thesis, although I recognise the importance of all pillars. Each of these resonated with me and my fellow PhD Candidates. As we immersed ourselves in the vision of the PIA research project, it became a unifying thread that brought us together in our research ambitions. We found that all these priorities within the project spoke to us at different points and provided a strong point of communication and practice in the development of processes, prototypes or interfaces.

2.3 Contribution to PIA and its Relevance to the Thesis

To develop a participatory platform, an open and sustainable technological foundation for facilitating the reuse of CH resources was needed [Raemy 2021]. Throughout the PIA project, I was mainly involved in the extension of the data infrastructure, the uptake of IIF as well as designing the data model, leveraging Linked Art and WADM [Raemy 2024a]. As a member of Team B, I undertook this PhD as a bridge between the different teams, mostly participating in discussions with the three doctoral candidates from Team A to further develop and agree on the CAS data model and with the software developers from my team to discuss the impact of the data model on our evolving — yet transitory — infrastructure as well as helping in implementing the APIs adhering to the LOUD design principles.

It was necessary to redesign the data model within the context of a database migration, from Salsah to the DSP, that happened between November 2021 and March 2024. This updated version, based on the Knora Base Ontology¹⁸, corresponded to the needs of the CAS archives and to some extent to those of PIA, in particular to enable the PhD Candidates in Cultural Anthropology to make more precise assertions, whether in terms of descriptive metadata, or in the ability to link one object to another or to provide comments on these objects in several narrative forms.

Moreover, an assessment of the appropriate technical standards for improved usability of the objects by both humans and machines was carried out, as a basis for extending the capabilities provided by DaSCH, such as helping the software developers to implement Simple Image Interface Presentation

¹⁷ In our joint paper, we wrote ‘man-made’, corrected here, which makes me think of the transition within the CIDOC-CRM for the Entity E22 **Human-Made Object** from version 6.2.7 onward.

¹⁸ Knora Base Ontology: <https://docs.dasch.swiss/2023.07.01/DSP-API/02-dsp-ontologies/knora-base/>

(SIPI)¹⁹, a C++ image server compatible with the IIIF Image API and build services that create IIIF Presentation API 3.0 resources.

While the theoretical framework of the thesis extends across the scope of PIA, the empirical part focuses on a specific set of findings derived from the research project outlined in Chapter 7, under the title ‘PIA as a Laboratory’. In this chapter, I discuss the data model and its refinement as well as the generation of custom IIIF Manifests during the specific digitisation, cataloguing and indexing efforts that took place throughout the project for the three CAS collections (SGV_05, SGV_10 and SGV_12) under investigation, the implementation of LOUD standards, and the overall design of the technological underpinnings.

2.4 Involvement within the IIIF and Linked Art communities

I must acknowledge the invaluable role that my involvement within the IIIF and Linked Art communities has played in shaping my journey as a trained information specialist and an aspiring DH practitioner. Being an active participant in both communities has not only broadened my understanding of the latest developments in the field but has also profoundly influenced the trajectory of this dissertation.

I have been involved within the IIIF community since October 2016 and the Working Groups Meeting that happened in The Hague²⁰. This significant journey was, in fact, initiated by a recommendation from my first supervisor, Peter Fornaro, during my time as an undergraduate doing an internship at the DHLab. Little did I know that this recommendation would lead me to be carrying out a PhD and looking at IIIF not only as community-driven standards but as an object of study. Engaging with the IIIF community exposed me to cutting-edge advances in image interoperability and standards, and fostered a deeper appreciation for the importance of digital representations of cultural heritage. Through collaborative discussions with experts from diverse backgrounds, I gained new perspectives on the potential of technology to advance humanities research and preserve our collective cultural memory.

Similarly, my involvement in the Linked Art community introduced me to the opportunities offered by LOUD and its transformative impact on research discourse. Exposure to Linked Data methodologies and the CIDOC-CRM has significantly influenced the way I have structured and interpreted the data in this dissertation, thereby enriching its scholarly breadth and rigour. I started to be actively involved in Linked Art at the beginning of my PhD in 2021, but I was already a ‘fly on the wall’ by 2020, driven by the efforts of Rob Sanderson, my third supervisor. By mid-2023, I had become a member of the Editorial Board.

The individuals I have met and the knowledge shared in these vibrant communities have deeply informed my approach as a scholar. The invaluable connections and collaborations I have made have expanded my network of fellow researchers, educators, and experts, leading to fruitful discussions that have significantly shaped the research questions addressed in this thesis. The events and workshops organised by these communities have also provided immersive learning experiences, giving me first-hand insights into the tools, technologies and methodologies used in the context of describing and disseminating CH data. The dynamic ecosystem of these communities has served as an inspiring backdrop, fostering innovative thinking and encouraging a more holistic approach to my research.

¹⁹ SIPI documentation: <https://sipi.io/>

²⁰ IIIF Working Groups Meeting, The Hague, 2016: <https://iiif.io/event/2016/thehague/>

3

Interlinking Cultural Heritage Data

Interlinking CH data is an important aspect of publishing heritage collections over the web, in particular by using Linked Open Data (LOD) technologies to make assertions more easily readable and meaningful to machines [Marcondes 2021]. Due to the complexity of CH data and their intrinsic inter-relationships, it is necessary to define its nature and introduce controlled vocabularies and ontologies that can be integrated with existing web standards and interoperable with relevant platforms [Bruseker et al. 2017, Hyvönen 2020].

Efforts to interlink CH data have brought about significant advancements, but challenges remain. One such challenge is finding a balance between completeness and precision of expression to ensure that the that CH data remain accessible and usable to a wider audience. Addressing this challenge, the LOUD design principles and the specifications that adhere to those, such as the IIF Presentation API 3.0 and Linked Art, offer a promising approach [Raemy et al. 2023b]. By focusing on usability aspects from the perspective of software developers and data scientists involved in designing visualisation tools and data aggregation approaches, LOUD strives to enhance the overall user experience [Sanderson 2019b].

Finding this equilibrium becomes crucial as CH data continues to grow in complexity and size, necessitating the seamless integration of native web technologies. The LOUD concept cultivates an environment that encourages the formation of vibrant CoP and the seamless integration of native web technologies, wherein an essential principle is the availability of comprehensive documentation supplemented with practical examples [Raemy 2022a]. Moreover, the emphasis on leveraging widely adopted technologies enhances the interoperability of data and promotes its wider dissemination. With LOUD principles guiding the linking of CH data, the resulting web of knowledge becomes more than just a machine-readable resource; it transforms into a user-centric ecosystem where both accessibility of Linked Data and usability intersect to enable scholars and a wider audience to engage in the exploration and appreciation of CH [Newbury 2018]. Finally, by fostering a collaborative, knowledge-sharing mindset, LOUD empowers software developers to implement data in a robust way, drawing insights from shared experiences [see Page et al. 2020].

In this chapter, which serves as the literature review of the PhD thesis, I attempt to draw on this brief introduction by dividing the insights into seven sections in order to provide an overview of the key concepts related to interlinking data in the CH domain. The literature review primarily encompasses works published up until December 2023, providing a comprehensive snapshot of the field's current state and its evolution. Section 3.1 discusses what makes CH data stand out and Section 3.2 is about CH metadata standards, while Section 3.3 explores the technological trends, scientific movements and guiding principles that have shaped the field. Section 3.4 provides an overview of the web as an open platform, which are essential to understanding the current landscape of interlinking CH data. Section 3.5 focuses on LOUD, while Section 3.6 looks at characterising the community practices and semantic interoperability dimensions for CH. Finally, in Section 3.7, I summarise key elements from

each section and within each of these I give some initial thoughts with respect to LOUD, and then conclude the chapter with some considerations on why we as a society need to care about CH data.

3.1 What Makes Cultural Heritage Data Stand Out?

Here, I aim to establish the indirect territory of my study, as I am situated on a distinct plane that focuses on web technologies and standards — as well as software and services that enable them — as the subjects of investigation. However, it is crucial to acknowledge that LOUD specifications owe their existence to the available data that have served as case studies. Thus, their significance can be best understood through the lens of data and I recognise here the pivotal role played by CH practitioners — encompassing individuals from research and memory institutions — who have had a significant impact on specifying a series of web-based standards and who have helped to move forward the discovery of CH data and beyond, in particular those belonging to the public domain, in an open manner.

In Section 3.1.1, I provide an introduction to CH as recognised by the United Nations Educational, Scientific and Cultural Organization (UNESCO). I explore the tangible, intangible, and natural dimensions of CH, laying the foundation for further discussions on its representation and preservation, notably by giving a first definition of CH data. Next in Section 3.1.2, I look at the challenges of representation and embodiment of CH data. This subsection examines the challenges in describing and preserving its materiality or embodied aspects. Understanding the significance of collective efforts, communities, and the interplay of technologies. Thirdly, I discuss what I called ‘Collectives and Apparatuses’ in Section 3.1.3 where I highlight how actors in terms of collaborative actions and apparatuses play a pivotal role in CH.

3.1.1 Cultural Heritage

The legacy of CH encompasses physical artefacts and intangible aspects inherited from past generations, reflecting the history and traditions of societies. Meanwhile, CH constantly evolves due to complex historical processes, necessitating preservation and protection efforts to prevent its loss over time [Loulanski 2006]. The dynamic nature of CH demands collaborative actions, including documentation and the use of a range of technologies.

The concept of CH is also characterised by perpetual evolution, mirroring the historical processes that shape societies over time. Social, political, economic, and technological shifts invariably influence the definition and perception of CH, prompting continuous reinterpretations and reevaluations of its significance. Over the years, the enthusiasm for the protection of cultural property has enriched the term with new shades of meaning. As societies undergo transformations, new layers of meaning and relevance are superimposed on existing CH, perpetually enriching its essence. As articulated by [Ferrazzi 2021, p. 765]:

“Cultural heritage”, as an abstract legacy or as a merge of tangible and intangible values, is able to encompass the totality of culture(s); in so, assuming a symbolic value that brings a clear break with all other terminologies. In conclusion, ‘cultural heritage’ as a legal term has demonstrated more than any others to be a real ensemble of historical stratification and cultural diversity.’

The advent of globalisation and rapid advancements in technology have further accelerated the evolution of CH. Increased interconnectedness and cross-cultural interactions have led to the fusion of traditions and the emergence of novel cultural expressions. Moreover, the digital era has facilitated the dissemination of CH resources on a global scale, transcending geographical barriers and preserving cultural knowledge for future generations as ‘Information and Communication Technologies (ICTs) provide

researchers with powerful tools to deal with the digital acquisition, storage, conservation, recreation, reconstruction, and representation of CH assets' [Portalés et al. 2018].

Thus, the intriguing nature of CH resources can be attributed to their multifaceted and diverse characteristics. The conservation and promotion of these resources demand a nuanced comprehension of the various types of heritage resources, culminating in effective preservation and promotion strategies that can account for their heterogeneity [Windhager et al. 2019].

According to UNESCO Institute for Statistics [2009], CH includes tangible and intangible heritage. Tangible CH refers to physical objects such as artworks, artefacts, monuments, and buildings, while intangible CH comprises practices, knowledge, folklore and traditions that hold cultural significance [Munjeri 2004]. The concept of heritage has evolved through a process of extension to include objects that were not traditionally considered part of the heritage. The criteria for selecting heritage have also changed, taking into account cultural value, identity, and the ability of the object to evoke memory. This shift has led to the recognition and protection of intangible CH, challenging a Eurocentric perspective and embracing cultural diversity as a valuable asset for humanity [Vecco 2010].

Conservation guidelines have broadened the concept of heritage to include not only individual buildings and sites but also groups of buildings, historical areas, towns, environments, social factors, and intangible heritage [Ahmad 2006]. In 2019, another instance of UNESCO defines CH in an even more comprehensive manner, taking into account natural heritage:

‘Cultural heritage is, in its broadest sense, both a product and a process, which provides societies with a wealth of resources that are inherited from the past, created in the present and bestowed for the benefit of future generations. Most importantly, it includes not only tangible, but also natural and intangible heritage.’ [UNESCO. Culture for Development Indicators 2014, p. 130]

In thinking about the concept of CH, I find this last definition particularly resonant. This broader perspective is motivated by my interest with LOUD specifications as a research area, particularly because of their notable data agnosticism and as it resonated with Hyvönen [2012, pp. 1-3]’s subdivision of CH as well. These services have the adaptability to process and use different types of data, transcending the boundaries of specific domains or disciplines. Although grounded in concrete CH cases, their potential to extend to any type of data, including those from Science, technology, engineering, and mathematics (STEM), is a compelling prospect that warrants further exploration, a point that I will explore later.

The following sub-subsections aim to briefly discuss tangible, intangible, and natural heritage, as well as providing a definition of CH data which can serve as a foundational reference for this thesis.

3.1.1.1 Tangible Heritage

Tangible CH encompasses physical artefacts and sites of immense cultural significance that are passed through generations in a society [Vecco 2010]. These objects are tangible manifestations of human creativity, representing artistic creations, architectural achievements, archaeological remains as well as collections held by CHIs.

One aspect of tangible CH is artistic creations such as paintings, sculptures and traditional handicrafts. These artefacts embody cultural values and artistic expressions and serve as essential reflections

of a society's collective ethos. For example, artworks such as 'Irisés' from Vincent van Gogh²¹ and Alberto Giacometti's 'L'Homme qui marche I'²² are revered works of art that have deep cultural significance in Europe and all over the world.

The built heritage, including monuments, temples and historic buildings, is another important component of the tangible CH. These architectural marvels not only represent past civilisations, but also convey the social values and aspirations of their time. The Taj Mahal, an exemplary white marble structure in India, stands as a poignant testament to Mughal architecture. Closer to where I write this dissertation one can mention the Abbey of St Gall, a convent from the 8th century which is inscribed on the UNESCO World Heritage List. In the context of urban heritage, conventional definitions of built heritage often focus narrowly on the architectural and historical value of individual buildings and monuments, which are well protected by existing legislation. However, the challenge is to preserve urban fragments - areas within towns and cities that may not qualify as designated conservation areas, but are of significant cultural and morphological importance [Tweed and Sutherland 2007]. For instance, Rautenberg [1998] proposes two categories of built CH: heritage by designation and heritage by appropriation. Heritage by designation involves experts conferring heritage status on sites, buildings, and cultural objects through a top-down approach, often without public participation. This method can be predictable and uncontroversial, but can be criticised for being elitist and neglecting unconventional heritage. On the other hand, heritage by appropriation emphasises community and public involvement in identifying and preserving cultural expressions, leading to a more inclusive and dynamic understanding of heritage.

Archaeological sites are also an integral part of the tangible CH, offering invaluable insights into past societies and ways of life. As per May 2024, UNESCO's long list of World Heritage Sites includes 1,199 cultural and natural sites in 168 different state parties — including 48 sites in transboundary regions²³. Sites such as Machu Picchu, an impressive Inca citadel in the Peruvian Andes, bear witness to the architectural achievements and cultural practices of ancient civilisations. If archaeological sites are invaluable, they face significant threats such as looting, destruction, exploitation, and extreme weather phenomena [Bowman 2008, Micle 2014]. To safeguard them, conservation efforts must be case-specific and include documentation and assessment of experiences gained [Aslan 1997].

The preservation of tangible CH extends beyond physical objects to include libraries, archives and museums that house collections of books, manuscripts, historical documents and artefacts.

Incidentally, the term 'cultural property' is also employed as a related concept to tangible CH, encompassing both movable and immovable properties as opposed to less tangible cultural expressions [Ahmad 2006]. Cultural property is protected by a number of international conventions and national laws. For instance, the Blue Shield²⁴ — an international organisation established in 1996 by four non-governmental organisations²⁵ — aims to protect and preserve heritage in times of armed conflict and natural disasters [Van der Auwera 2013]. Its mission has been revised in 2016:

²¹ Van Gogh, Vincent. (1889). *Irisés* [Oil on canvas]. Getty Museum, Los Angeles, CA, USA. <https://www.getty.edu/art/collection/object/103JNH>

²² Giacometti, Alberto. (1956). *L'homme qui marche I* [Sculpture]. Carnegie Museum of Art, Pittsburgh, PA, USA. <https://www.wikidata.org/entity/Q706964>

²³ UNESCO World Heritage List: <https://whc.unesco.org/en/list/>

²⁴ Blue Shield International: <https://theblueshield.org/>

²⁵ The International Committee of the Blue Shield (ICBS) was founded by the International Council on Archives (ICA), International Council on Museums (ICOM), International Council on Monuments and Sites (ICOMOS), and International Federation of Library Associations and Institutions (IFLA).

The Blue Shield is ‘committed to the protection of the world’s cultural property, and is concerned with the protection of cultural and natural heritage, tangible and intangible, in the event of armed conflict, natural- or human-made disaster.’ [Blue Shield 2016, art. 2.1]

Overall, tangible CH is a testament to human ingenuity and cultural diversity, and serves as a bridge between the past and the present. Its preservation is a collective responsibility, ensuring that the legacy of past generations endures and the wealth of cultural diversity continues to enrich the fabric of society.

3.1.1.2 Intangible Heritage

The concept of intangible heritage emerged in the 1970s and was coined at the UNESCO Mexico Conference in 1982 [Leimgruber 2010] with the aim of protecting cultural expressions that were previously excluded from preservation efforts [Hertz et al. 2018]. UNESCO’s previous focus had been on material objects, primarily from wealthier regions of the global North, leaving the intangible cultural heritage of the South overlooked. Attempts to protect intangible heritage through legal measures like copyright and patents were ineffective due to the collective nature of these cultural expressions and the anonymity of creators. The Convention acknowledges that intangible CH is essential for cultural diversity and sustainable development.

Below is the definition given by the ‘Convention for the Safeguarding of the Intangible Cultural Heritage’:

“The Intangible Cultural Heritage” means the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their cultural heritage. This intangible cultural heritage, transmitted from generation to generation, is constantly recreated by communities and groups in response to their environment, their interaction with nature and their history, and provides them with a sense of identity and continuity, thus promoting respect for cultural diversity and human creativity.’ [UNESCO 2022]

According to UNESCO, intangible CH can be manifested in the following domains:

- oral traditions and expressions, including language as a vehicle of the intangible CH;
- performing arts;
- social practices, rituals and festive events;
- knowledge and practices concerning nature and the universe;
- traditional craftsmanship.

Overall, intangible CH is a multifaceted concept that encompasses both traditional practices inherited from the past and contemporary expressions in which diverse cultural groups actively participate [Leimgruber 2008, Munjeri 2004]. It includes inclusive elements shared by different communities, whether they are neighbouring villages, distant cities around the world, or practices adapted by migrant populations in new regions. These expressions have been passed down from generation to generation, evolving in response to their environment, and play a crucial role in shaping our collective identity and continuity. Intangible CH promotes social cohesion, strengthens a sense of belonging and responsibility, and enables individuals to connect with different communities and society at large.

Central to the nature of intangible CH is its representation within communities. Its value goes beyond mere exclusivity or exceptional importance; rather, it thrives on its association with the people who preserve and transmit their knowledge of traditions, skills and customs to others within the community and across generations. The recognition and preservation of intangible CH depends on the communities, groups or individuals directly involved in its creation, maintenance and transmission. Without their recognition, no external entity can decide on their behalf whether a particular practice or expression constitutes their heritage. The community-based approach ensures that intangible CH remains authentic and deeply rooted in the living fabric of society, protected by those who care for and perpetuate it.

In Switzerland, the Winegrower’s Festival in Vevey (*La Fête des Vignerons*), a plurisecular event celebrating the world of wine making [Vinck 2019] and the Carnival of Basel (*Basler Fasnacht*) [Chiquet 2023] are examples of traditions that are listed among UNESCO’s intangible CH.

(In)tangibility is not always a straightforward concept and can indeed be blurred, i.e. it goes beyond the mere idea of materialisation. Many artefacts and elements of CH possess both tangible and intangible qualities that intertwine and complement each other, making the distinction less clear-cut.

For instance, this Male Face Mask, held at the Art Institute Chicago²⁶, also known as ‘Zamble’, from the Guro people in the Ivory Coast holds dual significance as both a tangible and intangible CH. As a tangible object, the mask is a physical artefact made from wood and pigment, fabric, and various adornments, that combines animal and human features representing the Guro people’s artistic skills. On the other hand, as an intangible cultural object, the Zamble mask carries profound spiritual and cultural meaning. It plays a significant role in commemorating the deceased during a man’s second funeral. These second funerals are organised months or even years after the actual burial as a way to honour and remember the departed [see Haxaire 2009]. Thus, the preservation and appreciation of both the tangible and intangible aspects of the mask are essential to its cultural relevance.

Another example of the blurred line between tangible and intangible heritage is emphasised by De Muynke et al. [2022] in recreating reported perceptions of the acoustics of Notre-Dame de Paris through a collaboration between sciences of acoustics and anthropology. The authors highlight the heritage value of how people subjectively perceive sound in a space, particularly in places of worship where sound and music are integral to the religious experience. The authors advocate integrating the study of both material and non-material aspects to understand the changing sonic environments of heritage buildings [De Muynke et al. 2022, pp. 1-2]. Katz [2023] articulates that ‘*acoustics is an intangible product of a tangible building*’. This integrated perspective could lead to a more holistic understanding of the dynamics between physical spaces and the perceptual and experiential dimensions attached to them.

3.1.1.3 Natural Heritage

Natural heritage, encompassing geological formations, biodiversity, and ecosystems of cultural, scientific, and aesthetic value, shares a significant overlap with CH. Many natural sites hold spiritual and symbolic importance for communities, becoming repositories of cultural memory and identity [Lowenthal 2005]. Traditional ecological knowledge developed by various cultures also underscores the interconnectedness of cultural and natural heritage, as indigenous communities have accumulated wisdom on sustainable resource use and ecological balance [Azzopardi et al. 2023]. Moreover, the conservation and

²⁶ Guro. (1900-1950). *Male Face Mask (Zamble)* [Wood and pigment]. Art Institute of Chicago, Chicago, IL, USA. <https://www.artic.edu/artworks/239464>

sustainable management of natural heritage is often intertwined with efforts to protect CH, fostering a collective commitment to preserve these entangled legacies for future generations.

The link between natural and CH goes beyond their shared values; spatial overlaps further accentuate their interdependence. Natural sites may have cultural significance, while CH sites may be situated within natural landscapes. For example, a national park may include archaeological sites or culturally revered landscapes, thus intertwining the cultural and natural dimensions. This spatial intermingling highlights the inextricable relationship between human societies and the natural environment, as cultural practices and beliefs become intertwined with the landscapes they inhabit. In this way, the preservation of both natural and cultural heritage becomes essential not only for their intrinsic worth but also for sustaining the narrative of our shared human and environmental history.

Additionally, the distinction between nature and culture is not only subjective and dependent on human appreciation [Vandenhende and Van Hoorick 2017]. Rather, it is a concept intrinsically linked with the overarching framework of modernism, a perspective that has been critically examined and deconstructed by the influential sociologist and philosopher, Bruno Latour, that have argued that ‘*we have never been modern*’ [Latour 1993]. Latour’s deconstruction of the modernist perspective extends to the recognition that the ‘*the proliferation of hybrids has saturated the constitutional framework of the moderns*’ [Latour 1993, p. 51]. This assertion underscores the fundamental challenge posed by hybrid entities – those that blur the boundaries between nature and culture – to the traditional categories upon which modernist thinking has been predicated. In essence, the concept of hybrids disrupts the neat divisions between the natural and social worlds that have been a hallmark of modernist discourse and provide us an opportunity to situate ourselves as ‘amodern’ as opposed to postmodern [Latour 1990].

In addition to Latour’s critique of the modernistic distinction between nature and culture, the concept of the ‘parasite’, as expounded by Michel Serres, one of the influential thinkers who significantly influenced Latour’s intellectual development [Berressem 2015]. It offers a valuable lens through which to examine the intricacies of interconnectedness and interdependence within our world. In his view, everything is enmeshed in a complex web of relationships that negates the existence of self-contained entities. Rather than seeing discrete and isolated entities, Serres invites us to see everything as an integral part of a larger system in which each component is inextricably dependent on the others [Serres 2014]. Together, these complementary perspectives invite us to reevaluate our understanding of the intricate tapestry of existence, emphasising the complexities of our relationship with the world.

Thus, the appreciation of nature and culture is not mutually exclusive, but rather forms a continuous and evolving relationship. The modern perspective has historically separated these realms, treating them as distinct and disconnected. However, a more inclusive approach dissolves this artificial boundary and recognises the interconnectedness of nature and culture [Haraway 2008, 2016b]. This paradigm shift challenges the traditional modern understanding and invites a more holistic view in which natural and cultural heritage are mutually constructed within a complex network of relationships.

Recognition of this relationship is essential in the context of heritage conservation and understanding. The dynamic interplay between nature and culture is recognised, and the acknowledgement of their coexistence promotes a more holistic approach to heritage conservation, where cultural practices, traditions and ecological systems are seen as interdependent aspects of the wider heritage tapestry. This recognition encourages us to see heritage sites not as isolated entities, but as part of a larger web of interconnectedness, and urges us to conserve and value both cultural and natural heritage with a shared responsibility. Adopting this interconnected perspective enables us to appreciate the profound

connections between human societies and the natural world, and inspires a collective commitment to safeguarding these precious legacies for future generations.

3.1.1.4 Cultural Heritage Data

As I embark on the exploration of CH data, it is first necessary to establish a basic understanding of data in this context. At its core, data represents more than mere numbers and facts; it constitutes a collection of discrete or continuous values that are assembled for reference or in-depth analysis. In essence, data are the rich tapestry upon which the narratives of CH are woven, making its comprehension a critical prerequisite for our expedition into this domain.

Luciano Floridi — a prominent philosopher in the field of information and digital ethics — provides a thorough perspective on the term ‘data’ and offers valuable insights into its fundamental nature in its Philosophy of Information (PI). He perceives ‘*data at its most basic level as the absence of uniformity, whether in the real world or in some symbolic system. Only once such data have some recognisable structure and are given some meaning can they be considered information*’ [Floridi 2010]. This initial definition sets the stage for a deeper exploration of Floridi’s understanding of data, as he further focuses on its transformative journey into a more meaningful and structured form, which we will explore next.

Building upon Floridi’s foundational concept of data as the absence of uniformity, his subsequent definition provides a more comprehensive perspective. In a previous work, Floridi [2005, p. 357] argues that ‘*data are definable as constraining affordances, exploitable by a system as input of adequate queries that correctly semanticise them to produce information as output*’. This definition highlights the dynamic role of data, not only as raw entities awaiting structure and meaning but also as elements imbued with the potential to constrain and guide systems towards the generation of meaningful information.

Transitioning from Floridi’s concept of data, we progress to the view that data can be notably seen as interpretable texts within the Digital Humanities (DH) perspective. According to Owens [2011]: there are four main perspectives on how Humanists can engage with data:

- **Data as constructed artefacts:** data are a product of human creation, not something inherently raw or neutral;
- **Data as interpretable texts:** Humanists can interpret data as authored works, considering the intentions of the creators and how different audiences understand and use the data;
- **Data as processable information:** data can be processed by computers, allowing various forms of visualisation, manipulation and analysis, which can lead to further perspectives and insights;
- **Data can hold evidentiary value:** data, as a form of human artefact and cultural object, can provide evidence to support claims and arguments.

These considerations highlight the multifaceted nature of data within the field of DH. It is in this complex landscape that we recognise that data transcends its traditional role as a passive entity. As Rodighiero [2021, p. 26, citing Akrich et al. [2006]] suggests ‘*there is no doubt that data are full-fledged actors that take part in the social network the actor-network theory describes, in which both human and non-human intertwine and overlap.*’ This notion – rooted and borrowed from Science and Technology Studies (STS) – reinforces the idea that data, as an active and dynamic entity, plays a significant role in shaping the interactions between human and non-human actors in any digital spheres.

From these angles, I can look at the characteristics of CH data. Bruseker et al. [2017, p. 94] articulate that ‘*Data coming from the CH community comes in many shapes and sizes. Born from different disciplines, techniques, traditions, positions, and technologies, the data generated by the many different specializations that fall under this rubric come in an impressive array of forms*’.

In exploring CH data, it is important to recognise the inherent diversity stemming from diverse disciplines, techniques, and traditions. Bruseker et al. [2017, p. 94] aptly emphasise this, highlighting the extensive array of forms in which data manifests. This heterogeneity raises fundamental questions about the unity and identity of CH data — a crucial aspect deserving acknowledgement within this context. As the authors astutely ponder:

‘It could be a natural problem to pose from the beginning: if the data of this community indeed presents itself in such a state of heterogeneity, does it not beg the question if there is truly an identity and unity to cultural heritage data in the first place? It could be argued that Cultural Heritage, as a term, offers a fairly useful means to describe the fuzzy and approximate togetherness of a wide array of disciplines and traditions that concern themselves with the human past.’

Expanding on these insights, CH data refer to digital or data-driven affordances of CH²⁷, embodying a rich and varied compilation of insights originating from a variety of disciplines, techniques, traditions, positions and technologies. It encompasses both tangible and intangible aspects of a society’s culture as well as natural heritage. These data, derived from a wide range of disciplines, offer a latent capacity to support the generation of knowledge relating to historical time periods, geospatial areas, as well as current and past human and non-human activities. They are collected, curated and maintained by various entities such as libraries, archives, museums, higher education institutions, non-governmental organisations, indigenous communities and local groups as well as by the wider public.

Building further on the mosaic of CH data, three primary dimensions come to the fore: heterogeneity, knowledge latency, and custodianship.

Heterogeneity As a fundamental characteristic, signifies the diverse forms and origins that shape this invaluable reservoir of human heritage. Different techniques and varying viewpoints in treating modelling also contribute to this heterogeneity [Guillem et al. 2023a].

Knowledge latency It highlights the temporal dimension, presenting CH data as a repository of latent knowledge awaiting discovery and interpretation. Notably, not all artefacts are – or should be – digitised, and even among those that are, (mis)representation and challenges in interconnecting them persist [Rossenova and Di Franco 2022]. Besides, the issue of structured data – or the lack of it – reinforces the aspect of knowledge latency [Hacıgüzeller et al. 2021].

Custodianship This dimension reinforces the essential role played by a variety of entities, predominantly CHIs, in safeguarding and managing resources, ensuring their preservation and accessibility for present and future generations. However, it is very important to acknowledge the great divide in terms of resources, with indigenous and local communities often facing challenges in custodianship responsibilities.

Taken together, these dimensions contribute to a comprehensive understanding of the nuanced fabric of CH data. They reveal the diversity of forms and origins, the temporal aspects and the responsible stewardship that are crucial to the sustainability of such data.

²⁷ I have opted for the term ‘affordance’ and not ‘representation’ as my intention is to maintain a comprehensive scope that encompasses various modalities such as modelling endeavours.

By shifting our focus to the sphere of humanities data, we broaden our scope to extend beyond the peculiarities of CH data. Drawing parallels between these areas allows us to grasp the interconnectedness of our heritage.

CH data usually refers to information about cultural artefacts, sites, and practices that hold historical or cultural significance. Humanities data encompasses information about human culture, history, and society, including literature, philosophy, art, and language [Tasovac et al. 2020]. Both often involve ethical considerations, such as ownership, access, and preservation, and require a comprehensive understanding of their various meanings and values [Ioannides and Davies 2019]. Moreover, Schöch [2013] explains that data in the humanities, such as text and visual elements, have unique qualities. While these analogue forms could be considered data, they lack the ability to be analysed computationally as they are non-discrete. The semiotic nature of language, text and art introduces dimensions tied to meaning and context, making the term ‘data’ problematic. Critics question its use because it conflicts with humanistic principles such as contextual interpretation and the subjective position of the scholar.

Schöch [2013] distinguish data in the humanities further into two core types: smart and big data. The former tends to be small in volume, carefully curated, but harder to scale such as digital editions. As for the latter, it describes voluminous and varied data and it loosely relies on the three *Vs* by Laney [2001]: volume, velocity and variety (see Section 3.3.1.2). Yet, big data in the humanities differs significantly from other fields as it rarely requires rapid real-time analysis, is less focused on handling massive volumes, and instead deals with diverse, unstructured data sources. Schöch [2013] concludes by arguing that *‘I believe the most interesting challenge for the next years when it comes to dealing with data in the humanities will be to actually transgress this opposition of smart and big data. What we need is bigger smart data or smarter big data, and to create and use it, we need to make use of new methods.’*

Data processing offers great potential for humanities research as Owens [2011] argues: *‘In the end, the kinds of questions humanists ask about texts and artifacts are just as relevant to ask of data. While the new and exciting prospects of processing data offer humanists a range of exciting possibilities for research, humanistic approaches to the textual and artifactual qualities of data also have a considerable amount to offer to the interpretation of data’.*

While the term ‘data’ in the context of the humanities may raise questions due to its semiotic and contextual complexities, it serves as a foundation for understanding both CH data and broader humanities data. The data originating from CH and the humanities are inherently intertwined, as they often share a similar nature and purpose for scholars. This strong interconnection leads to a collaborative relationship between the Galleries, Libraries, Archives, and Museums (GLAM) sector and the humanities or DH. Scholars in the humanities frequently rely on digitised cultural artefacts, historical records, linguistic resources, and literary works provided by GLAM institutions to gain valuable insights into human history, culture, and traditions. The digitisation efforts and research collaborations between these entities play a pivotal role in preserving CH data and advancing our understanding of diverse societies, fostering a deeper appreciation of our shared human heritage. CH data and humanities data are distinct from other scientific data due to their qualitative and subjective nature, which requires different methods of analysis than quantitative scientific data. They include archival and special collections, rare books, manuscripts, photographs, recordings, artefacts, and other primary sources that reflect the cultural beliefs, identity, and memory of a people [see Izu 2022, Sabharwal 2015].

In summary, while CH data and humanities data share some commonalities, they differ in terms of scope and subject matter. CH data focuses specifically on the preservation and documentation of

physical artefacts and intangible attributes, while humanities data encompasses a broader range of disciplines within the humanities [Münster et al. 2019]. However, it is important to note that the distinction between CH data and humanities data can be blurred, as (meta)data should ideally be co-created and integrated across both domains.

3.1.2 Representation and Embodiment of Cultural Heritage Data

Digital representation of CH data, while preserving their context and complexity, remain a significant challenge. Those representations, sometimes referred to as digital surrogates or digital twins [Conway 2015, Semeraro et al. 2021, Shao and Kibira 2018], of CH data can potentially lead to a loss of context and a reduction in the richness of the CH represented. For instance, a digital image of a cultural artefact may not capture its materiality, such as its texture, weight, and feel, which are essential aspects of the artefact's cultural significance [Force and Smith 2021]. Furthermore, digital representations may also exclude vital social, cultural, and historical contexts surrounding the object, which is crucial to understanding its full cultural value [Cameron 2007].

This subsection is structured around two key dimensions. Firstly, it explores materiality, highlighting how digital representations may fail to capture important aspects that are integral to understanding the significance of CH resources. Secondly, it navigates the convergence and divergence between digitised CH and digital heritage.

3.1.2.1 Materiality

Briefly, materiality refers to the physical qualities of an object or artefact, such as its colour, texture, and composition. As part of built heritage, the emphasis for materiality relates primarily to architecture, its associated techniques and the range of materials used in the construction or renovation of a building. More specifically, materiality acts as a pivotal factor in the transformation of disparate fragments of material culture into heritage, providing a vital link to the intangible facets of heritage. It contributes significantly to an individual's social position and ability to navigate specific social milieus, thereby determining their ability to transmit cultural knowledge and values to future generations. The transformative potential of materiality in this regard underscores its fundamental role in perpetuating heritage and the transmission of cultural legacies [Carman 2009]. The physical attributes of objects, including texture, colour and shape, can evoke different emotions and associations, shaping people's perceptions and memories of these events. Beyond retrospective influences, the potential of materiality extends to the creation of new memories and meanings, as exemplified by the use of materials such as glass in contemporary art. In such cases, materials evoke not only their inherent properties but also symbolic connotations, adding new layers of meaning and memory to the artistic narrative [Fiorentino and Chinni 2023].

Edwards and Hart [2004, p. 3] argue that materiality is not just concerned with physical objects in a positivist sense, but also involves complex and fluid relationships between people, images, and things. This relationship is influenced by social, cultural, and historical contexts, and plays a crucial role in shaping our perceptions and experiences of the world. Moreover, materiality is central to giving meaning to non-human entities [see Haraway 2003, Latour 1996, Star and Griesemer 1989], which emphasises the role of both humans and non-humans in shaping social and cultural phenomena. For CH data, diversity is at its core, as it allows for the exploration of different ways of knowing, experiencing, and expressing the world. Therefore, it is important to approach materiality not as a static and fixed concept, but as a dynamic and evolving phenomenon that is shaped by multiple forces [Müller 2018, pp. 62-63]. When discussing materiality, there is also its negation, i.e. the notion of space or emptiness, such as how people interact with it through built heritage, which is regarded as a primordial medium of material culture, as expounded by Guillem et al. [2023b, p. 2]:

‘The most intuitive and foundational definition of architecture is the built thing, that is the architecture qua building or built work. Human beings continuously interact with the built materiality through the non-materiality of space. Space as emptiness is formed and defined by the materiality that affects its existence. That relation between fullness and emptiness is what makes possible architecture as lived and experienced space.’

Materiality also offers a means of challenging dominant narratives and power structures, particularly the Western-centric perspective on CH. It gives greater recognition to the importance of intangible CH, which often takes a back seat to tangible objects in dominant narratives [Lenzerini 2011]. By highlighting the materiality of marginalised or forgotten elements, individuals can reclaim their heritage and challenge dominant narratives that marginalise certain groups, contributing to a more inclusive and accurate representation of CH.

The primary focus in terms of digitisation is also on preserving material-based knowledge, often overlooking the dynamic and living nature of intangibility. Hou et al. [2022] stress the crucial role of computational heritage and information technologies advances in preserving and improving access to intangible CH. Effectively documenting the ephemeral aspects of intangible heritage and communicating the knowledge that is deeply linked to individuals are pressing challenges. Recent initiatives seek to capture the dynamic facets of cultural practices, using visualisation, augmentation, participation and immersive experiences to enhance experiential narratives. There is a strong call for a strategic re-evaluation of the intangible CH digitisation process, emphasising the human body as a vessel for traditions and memories, such as capturing traditional Southern Chinese martial arts, who has been passed down colloquially from generations and needs a methodological approach to capture such embodied knowledge [see Adamou et al. 2023, Hou and Kenderdine 2024].

Even in cases where considerable efforts have been devoted to digitisation of physical objects such as medieval manuscripts and rare books over the past few decades [Nielsen 2008], a lingering concern persists regarding the authentic encounter with the original artefact, despite its enhanced accessibility through digital surrogates [van Lit 2020]. Material attributes present a persistent challenge to achieving full replication. Despite advances facilitated by techniques such as Reflectance Transformation Imaging (RTI), 3D digitisation, or VR and Augmented reality (AR), which offer better experiential immersion and are more effective than two-dimensional representations in addressing certain materiality concerns, the ability to replicate the multifaceted sensory experience associated with the original object, including the palpable emotions and spatial sensation, remains an ongoing endeavour, presenting a complex and multifaceted dimension of a challenge that is quite unlikely and may never be fully feasible [see Endres 2019].

3.1.2.2 Digitised Cultural Heritage and Digital Heritage

The concepts of digitised CH and digital heritage intersect through the use of digital technology for the preservation, access, and dissemination of CH resources. Digitised CH focuses on converting physical artefacts into digital forms, ensuring their long-term preservation and accessibility through digital means. Conversely, digital heritage includes a broader range of digital tools and resources ‘to preserve, research and communicate CH’ [Münster et al. 2021, p. 2, citing Georgopoulos [2018]].

Digitised CH acts as a critical bridge, facilitating a transition from traditional or analogue GLAM practices to a digital environment. This shift is pivotal in unlocking the potential of digitised CH. These values extend beyond scholarly pursuits, despite the majority of digitisation efforts being driven by research funding. In doing so, it becomes evident that the creative reuse and data-driven innovation stemming from digitised CH necessitate substantial and sustained investment in the GLAM sector. This

investment is fundamental, especially amidst reduced funding due to years of austerity. Terras et al. [2021] underscore this need, shedding light on the delicate balance required with commercial outcomes. They emphasised that leveraging CH datasets offers vast opportunities for technological innovation and economic benefits, urging professionals from various domains to collaborate and experiment in a low-risk environment.

Digital heritage²⁸ encompasses a wide range of human knowledge and expression in cultural, educational, scientific and various other domains. In today's rapidly evolving technological landscape, an increasing amount of this knowledge is either digitally created or in the process of being converted from analogue to digital formats [He et al. 2017]. These digital resources cover a wide range, including text, multimedia, software and more, and require deliberate and strategic management to ensure their long-term preservation. This valuable heritage, spread across the globe and expressed in multiple languages [UNESCO 2009].

In summary, digitised CH not only forges the path to digital heritage but also embodies an ever-evolving cultural landscape. Recognising the transformative potency with digital heritage is essential to enriching our understanding and engagement with our cultural roots. Both concepts are intimately embedded in CH and play a vital role as conduits.

3.1.3 Collectives and Apparatuses

The collaborative efforts of collectives and the operation of various apparatuses play a fundamental part in shaping the preservation, interpretation and dissemination of cultural artefacts and practices. This subsection is concerned with the central contributions of human and non-human actors engaged in cooperative action and the *modus operandi* of various apparatuses, such as building (digital) infrastructures. Some of these considerations are drawn from STS, which are more fully captured in Chapter 4, serving as the theoretical framework for the thesis.

Bruno Latour's concept of the importance of collectives and apparatuses [see Latour 2022, p. 15] can be extrapolated to CHIs. Every institution's or project's ultimate success hinges on the collaboration and support of individuals, as well as the tools, systems and technologies they use. Indeed, paralleling CHIs with wider contexts suggests that collective efforts and apparatuses play a critical role in shaping the effectiveness of any institution. This highlights the importance of recognising the influence of both human and non-human entities in institutional functioning and underlines the need for a more comprehensive understanding of the dynamics involved therein.

ANT can be a useful lens to analyse the creation, use, and dissemination of CH data. ANT posits that actors are not independent entities but are instead part of a network that consists of both human and non-human entities. According to ANT, every actor, be it a person or a technology, is a node in the network and contributes to the overall functioning of the network [Callon 2001, Latour 2005]. When we apply this framework to CHIs, we can identify the different actors involved in the creation, use, and dissemination of CH data. These actors can include individuals, such as curators, conservators, and historians, as well as non-human entities, such as databases, digitisation equipment, and software. Moreover, this approach can help us understand the interactions between these actors and how they shape the overall functioning of CHIs. For instance, digitisation equipment can enable the creation of high-quality digital images of artefacts, which can then be disseminated globally through online platforms. Examining the Notre-Dame de Paris, one can discern the keystones at the summit

²⁸ To some degree, parallels can be drawn between the distinctions of cultural and digital heritage with those drawn between the humanities and DH.

of its arches as indispensable actors within its architectural narrative. These keystones, imbued with historical narratives and a non-human facet, played a central role in the (digital) rescue and subsequent restoration efforts following the tragic roof fire in April 2019. Guillem et al. [2023a]’s study further elucidates this restoration journey, emphasising how the keystones, with their individual narratives and structural significance, contributed to the (digital) reassembly.

Building on this perspective, we can explore the importance of community involvement in the preservation and management of CH data, thereby increasing the potential for sustainable practices and inclusive engagement.

Local communities have an integral part to play in the management and preservation CH data, especially in the digital age where resources are often scarce for GLAM institutions. Community involvement has several benefits, including increased engagement and participation, access to local knowledge and expertise, and more sustainable and inclusive management and preservation practices [Ridge et al. 2021a]. For instance, geophysical technologies such as ground-penetrating radar have been used with great success in identifying and evaluating the depth, extent, and composition of CH resources for research and management purposes, easing tensions when working with sensitive ancestral places [Nelson 2021]. Collaborative environments can also help with CH information sharing and communication tasks because of the way in which they provide a visual context to users, making it easier to find and relate CH content [Respaldiza Hidalgo et al. 2011].

Embarking on Brown et al. [2023, pp. 6-7]’s insightful analysis, a prominent illustration of exemplary community practice can be found in the sphere of community museums in Latin America: *Inicio - Museos Comunitarios de América*²⁹. The author highlights the role of community engagement and leadership in the creation and operation of these museums. Such engagement ensures that these museums are not imposed from outside, but rather emerge organically as museums ‘of’ the community, resonating with its unique CH and identity. This approach is consistent with the ethos of ‘telling a story, building a future’, which embodies a deep commitment to community empowerment and cultural preservation. This community-centric approach amplifies the museum’s resonance with the community’s lived experiences and historical narratives.

At the same time, institutions can also benefit from collaborating with peer communities like IIF to promote greater access to their collections. IIF provides a set of open standards for delivering high-quality digital objects online at scale, which can help memory and academic institutions share their collections with each other and with the wider public [Snydman et al. 2015, Weinthal and Childress 2019]. By adopting IIF standards, organisations can make their collections more discoverable and accessible to researchers, developers, and other CH professionals [Padfield et al. 2022]. Involvement in communities such as IIF also helps to mitigate costs as they develop shared or adaptable resources and services [Raemy 2017].

Participation of communities in the management and preservation of CH resources is essential to ensure that CH is protected and accessible for future generations. By involving and participating in communities, GLAMs can tap into local as well as peer knowledge and expertise, making management and preservation practices more sustainable and inclusive. This approach also increases engagement and participation, ensuring that CH is valued and appreciated by the wider community. Thus, memory institutions need to collaborate closely with communities to ensure that CH data, and their underlying infrastructures and services, is being effectively curated [Delmas-Glass and Sanderson 2020].

²⁹ Inicio - Museos Comunitarios de América: <https://www.museoscomunitarios.org/>

Closely related to this context, [Star \[1999\]](#) points out the often unacknowledged role of infrastructure within society. She argues that infrastructures are necessary but often invisible and taken for granted:

‘People commonly envision infrastructure as a system of substrates – railroad, lines, pipes and plumbing, electrical power plants, and wires. It is by definition invisible, part of the background for other kinds of work. It is ready-to-hand. This image holds up well enough for many purposes – turn on the faucet for a drink of water and you use a vast infrastructure of plumbing and water regulation without usually thinking much about it.’ [[Star 1999](#), p. 380]

[Star \[1999, pp. 381-382, citing Star and Ruhleder \[1994\]\]](#) identifies nine dimensions to define infrastructure. They provide a comprehensive framework to comprehend the nuanced nature of infrastructure and its pervasive impact on diverse societal facets. The following dimensions are vital for analysing the often imperceptible, yet deeply embedded structures that constitute the foundational framework of both daily life and broader societal operations³⁰:

Embeddedness Infrastructure is sunk into and inside of other structures, social arrangements, and technologies. People do not necessarily distinguish the several coordinated aspects of infrastructure.

Transparency Infrastructure is transparent to use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly supports those tasks.

Reach or scope This may be either spatial or temporal – infrastructure has reach beyond a single event or one-site practice.

Learned as part of membership Strangers and outsiders encounter infrastructure as a target object to be learned about. New participants acquire a naturalised familiarity with its objects, as they become members.

Links with conventions of practice Infrastructure both shapes and is shaped by the conventions of a community of practice.

Embodiment of standards Modified by scope and often by conflicting conventions, infrastructure takes on transparency by plugging into other infrastructures and tools in a standardised fashion.

Built on an installed base Infrastructure does not grow *de novo*; it wrestles with the inertia of the installed base and inherits strengths and limitations from that base.

Becomes visible upon breakdown The normally invisible quality of working infrastructure becomes visible when it breaks: the server is down, the bridge washes out, there is a power blackout.

Is fixed in modular increments, not all at once or globally Because infrastructure is big, layered, and complex, and because it means different things locally, it is never changed from above. Changes take time and negotiations, and adjustment with other aspects of the systems are involved.

An appreciation of these dimensions is crucial to the analysis of the network of infrastructural systems that underpin contemporary society, and is necessary for the analysis of any digital infrastructure that manages CH data.

³⁰ The descriptions of each of these nine dimensions are selected excerpts from [Star \[1999\]](#).

Digital infrastructures – also known as e-infrastructures or cyberinfrastructures – are forms of infrastructure that are essential for the functioning of today’s society [see Jackson et al. 2007, Ribes and Lee 2010]. These kinds of infrastructure need to be understood as socio-technical systems, showcasing the interplay between technological components (such as hardware, software, and networks) and the social and organisational contexts in which they operate [Star and Ruhleder 1994]. According to Fresa [2013, p. 33], digital CH infrastructures should be able to serve the research needs of humanities scholars as well as having dedicated services for education, learning, and general public access. In terms of requirements, Fresa [2013, pp. 36-39] identifies three different layers of services: for content providers, for managing and adding value to the content, and for the research communities. For the latter, several sub-services tailored to research communities are listed. These encompass long-term preservation, Persistent Identifiers (PIDs)³¹, interoperability and aggregation, advanced search, data resource set-up, user authentication and access control, as well as rights management.

Overall, (digital) infrastructures are imperative apparatuses in preserving and sharing CH data. First, they support preservation by archiving digital artefacts and their metadata, protecting them from deterioration and loss. Secondly, these infrastructures facilitate accessibility, allowing a global audience to explore and appreciate cultural heritage online. Finally, they encourage interpretation and engagement, promoting cross-cultural understanding and knowledge sharing.

Moreover, infrastructure is a fundamental component that demands extensive investment, particularly in the creation of streamlined integration layers capable of interacting seamlessly with different systems. This can be exemplified by such institutions as the Rijksmuseum³², where a well-constructed infrastructure allows for efficient integration and interaction with various technological and organisational systems [Dijkshoorn 2023]. This investment serves as the foundation for an institution’s functionality, allowing for the smooth flow of data, the coordination of processes and the optimal use of resources. In a similar vein, Canning et al. [2022] argue that the often invisible structures of metadata, particularly in Linked Data ontologies, play a crucial role in shaping the interpretation of data. These structures, while not immediately apparent, are imbued with value judgements and ideological implications, extending the impact of metadata beyond mere technicalities to encompass diverse and inter-sectional perspectives. This multidimensional ontological approach addresses the complexity and diversity of data sources, paralleling the need for sophisticated infrastructures in institutions like the Rijksmuseum. It underscores the importance of integrating inter-sectional feminist principles in information systems, reflecting a commitment to diverse ways of knowing and nuanced storytelling.

Furthermore, as all (meta)data requires storage, it raises an important concern in terms of the entrenched power dynamics governing knowledge representation within information systems, as pointed out by Canning [2023]. This perspective, initially centred around museum objects, holds broader implications for all CH resources [see Simandiraki-Grimshaw 2023]. Canning strongly advocates for the essential adaptation of databases to embrace a diverse array of epistemological approaches by introducing new types of affordances. Databases, despite their role in information preservation, wield significant influence that can inadvertently stifle diverse modes of knowledge interpretation and ‘*can constrain ways of knowing*’. Furthermore, she compellingly argues that modifications to databases extend beyond technical adjustments; they are inextricably linked to shifts in institutional power dynamics and the

³¹ ‘A PID is a long-lasting reference to a digital resource. It usually has two components: a unique identifier and a service that locates the resource over time, even if its location changes. The first helps to ensure the provenance of a digital resource (that it is what it purports to be), whilst the second will ensure that the identifier resolves to the correct current location.’ [Digital Preservation Coalition 2017]

³² Rijksmuseum: <https://www.rijksmuseum.nl/>

enduring, often inequitable, power dynamics governing the world of museums – *or any CHIs* – and their curation.

In understanding the interplay of collectives and apparatuses, it is clear that key actors, including individuals, institutions, local and global communities, as well as the sophisticated fabric of (digital) infrastructures and their components, are deeply entangled and interconnected. These entities, both human and non-human, collectively shape and navigate the rich networks of human interactions and technologies that underpin the foundations of contemporary society.

3.2 Cultural Heritage Metadata

This subsection offers insights into the importance of metadata in CH, underlining its role in enhancing the understanding and accessibility of cultural artefacts. It is structured into three essential parts. I start with an introductory segment in subsection 3.2.1, then I explore the types and functions of metadata in subsection 3.2.2, thirdly in subsection 3.2.3, I outline some of the most important CH metadata standards, and finally in subsection 3.2.4, I explore the use of KOS, such as generic classification systems and controlled vocabularies.

3.2.1 Data about Data

For curating CH resources, metadata³³, ‘data about data’, is probably one of the key concept that needs to be introduced here. Metadata permeate our digital and physical landscapes, playing a vital role in organising, describing and managing a vast array of information. Rather than being confined to a specific domain, they are ubiquitous and pervade many aspects of our everyday lives [Riley 2017, pp. 2-3]. From websites and databases to social media platforms and online marketplaces, metadata adds meaning to data, enabling users to understand their context, relevance and provenance. As an example, Figure 3.1 shows the metadata of a book³⁴.

Metadata are central to the management and preservation of CH data, providing essential information to ensure that data can be properly organised, discovered and retrieved. For example, they facilitate the understanding and interpretation of data, enabling scholars and the public to access and use them effectively [Constantopoulos and Dallas 2008]. Metadata also help to ensure the long-term preservation and accessibility of CH data [Zeng and Qin 2022, pp. 490-491]. Providing metadata in a structured manner facilitates forms of aggregation, i.e. individuals and institutions being able to harvest and organise metadata from multiple sources or repositories into a centralised location [see Freire et al. 2017, 2021]. In addition, the importance of metadata as a gateway to information is particularly compelling when the primary embodiment of a record is either unavailable or lost. In cases where resources, time constraints, sensitive content or strategic decisions prevent the digitisation of an item, metadata becomes the principal means of representation and access. If a physical record is lost or damaged, the metadata associated with that record acts as a proxy for the record.

Riley [2017, p. 5] discusses the transformation of libraries over time. Initially, libraries moved from search terminals to the modern web-based resource discovery systems we use today. This shift was driven by advances in computerisation. Libraries’ basic approach to metadata is ‘bibliographic’, deeply

³³ Zeng and Qin [2022, p. 11] articulate that ‘as with “data”, metadata can be either singular or plural. It is used as singular in the sense of a kind of data; however, in plural form, the term refers to things one can count’. In the context of this thesis, I have chosen to favour the plural form of (meta)data. However, I acknowledge that I may occasionally use the singular form when referring to the overarching concepts or when quoting references verbatim.

³⁴ The snapshot of this bibliographic record was taken from https://swisscovery.slsp.ch/permalink/41SLSP_UBS/11jfr6m/alma991170746542405501.

Details	
Title	Metadata
Attribution	Marcia Lei Zeng and Jian Qin
Publication	Chicago : ALA Neal-Schuman
Edition, date	Third edition 2022
Format	xxvi, 613 Seiten : Illustrationen
Language	English
Notes	Includes bibliographical references and index
Subject	Metadata > Metadaten >
Classifications	025.3 (Dewey) 020 (Dewey)
Identifier	ISBN : 9780838948750 (Broschur) ISBN : 0838948758
Author	Qin, Jian 1956- (author) > Zeng, Marcia Lei 1956- (author) >
Display in the GND	Zeng, Marcia Lei 1956- Metadaten Qin, Jian 1956-

Figure 3.1: Snapshot from the Swisscovery Platform Showing the Bibliographic Record of [Zeng and Qin \[2022\]](#)

rooted in their traditional expertise in describing books. This approach involves providing detailed descriptions of individual items so that users can easily locate them within the library's collection.

On the other hand, archives use 'finding aids', which are descriptive inventories of their collections, coupled with historical context. These aids are essential for users to understand the material and to find groups of related items within the archive. The metadata used in archives allows for the contextualisation of materials, particularly papers of individuals or records of organisations, providing a richer understanding of the content.

Similarly, museums actively manage and track their acquisitions, exhibitions and loans through metadata. Museum curators use metadata to interpret collections for visitors, explaining the historical and social significance of artefacts and describing the relationships and connections between different objects. This helps to enhance the overall visitor experience and understanding of the artefacts on display or the digital resources on a particular website.

3.2.2 Types and Functions

CHIs share common objectives and concerns related to information management, as highlighted by [Lim and Li Liew \[2011\]](#), pp. 484-485]. These goals typically include facilitating access to knowledge and ensuring the integrity of CH data. However, it is important to note that CHIs also differ widely in how they deal with metadata. Different domains have unique approaches and standards for describing the materials they collect, preserve and disseminate, and even within a single domain there are significant differences.

There have been different attempts to categorise the metadata landscape. For instance, [Gilliland \[2016\]](#) identified the following five categories of metadata and their respective functions:

Administrative Metadata used in managing and administering collections and information resources, such as acquisition and appraisal information or documentation related to repatriation.

Descriptive Metadata used to identify, authenticate, and describe collections and related trusted information resources. Finding aids, cataloguing records, annotations by practitioners and end users, as well as metadata generated by or through a given Digital asset management (DAM) system can often be classified as descriptive metadata.

Preservation Metadata related to the preservation management of collections and information resources. Common examples of preservation metadata are documentation of physical condition of resources or of any actions taken to preserve resources, whether physical restoration or data migration.

Technical Metadata related to how a system functions or metadata behaves. Examples include software documentation and digitisation information.

Use Metadata related to the level and type of use of collections and information resources, such as circulation records, search logs, or rights metadata.

Meanwhile [Riley \[2009\]](#), as illustrated in a comprehensive visualisation graph³⁵, suggested seven functions, i.e. the role a standard play in the creation and storage and metadata, and seven purposes referring to the general type of metadata.

Functions Conceptual Model, Content Standard, Controlled Vocabulary, Framework/Technology, Markup Language, Record Format, and Structure Standard.

Purposes Data, Descriptive Metadata, Metadata Wrappers, Preservation Metadata, Rights Metadata, Structural Metadata, and Technical Metadata.

³⁵ Seeing Standards: A Visualization of the Metadata Universe. 2009-2010. Jenn Riley.
<https://jennriley.com/metadatamap/seeingstandards.pdf>

Almost a decade later, [Riley \[2017, pp. 6-7\]](#) summarised metadata types into four groupings instead of the seven purposes previously mentioned. ‘Data’ is removed from the list and technical, preservation and rights metadata are now grouped into a newly created administrative metadata category.

1. **Descriptive metadata** For finding or understanding a resource.
2. **Administrative metadata** Umbrella term referring to the information needed to manage a resource or that relates to its creation.
 - 2.1 **Technical metadata** For decoding and rendering files.
 - 2.2 **Preservation metadata** Long-term management of files.
 - 2.3 **Rights metadata** Intellectual property rights attached to content.
3. **Structural Metadata** Relationships of parts of resources to one another.
4. **Markup Language** Integrates metadata and flags for other structural or semantic features within content³⁶.

This classification of metadata types and function differs to the categories identified by [Gilliland \[2016\]](#) mostly due to the addition of structural metadata and markup language as their own categories [[Zeng and Qin 2022, p. 19](#)]. Table 3.1 lists the major types of metadata according to [Riley \[2017, p 7\]](#) and include example properties and their primary uses.

Table 3.1: Types of Metadata According to [Riley \[2017, p 7\]](#)

Metadata (Sub)type	Example properties	Primary uses
1. Descriptive metadata	Title, Author, Subject, Genre, Publication date	Discovery, Display, Interoperability
2.1 Technical metadata	File type, File size, Creation date, Compression scheme	Interoperability, Digital object management, Preservation
2.2 Preservation metadata	Checksum, Preservation event	Interoperability, Digital object management, Preservation
2.3 Rights metadata	Copyright status, Licence terms, Rights holder	Interoperability, Digital object management
3. Structural metadata	Sequence, Place in hierarchy	Navigation
4. Markup languages	Paragraph, Heading, List, Name, Date	Navigation, Interoperability

Ultimately, metadata can also be leveraged to create more inclusive and diverse representations of CH. For instance, metadata can be used to document and promote underrepresented communities and their heritage, providing greater visibility and recognition. This approach aligns with the principles of decolonising CH, promoting equity and social justice by recognising and valuing diverse cultural perspectives, especially in the prevailing anglophone and Western-centric standpoint in DH [[Mahony 2018, Philip 2021](#)].

Moreover, the distinction between data and metadata, as discussed in the work of [Alter et al. \[2023\]](#), is not always distinct, leading to the concept of ‘semantic transposition’. This complexity reflects

³⁶ A widespread example in the CH domain is the serialisation of metadata in Extensible Markup Language (XML), a World Wide Web Consortium (W3C) standard.

in CH where what is considered metadata in one context might be primary data in another, underscoring the necessity for adaptable frameworks in data management. This understanding is crucial for fostering inclusive and diverse representations in CH, ensuring that all cultural narratives are appropriately documented and acknowledged.

3.2.3 Standards

Metadata standards play a crucial role in ensuring that data are organised and consistent, facilitating mutual understanding between different stakeholders [Raemy 2020]. CHIs such as GLAMs typically follow established conventions or standards when organising their resources. Current methods of cataloguing have historical roots dating back to the 19th century, particularly with the development of cataloguing systems such as Antonio Panizzi’s ‘Ninety-One Cataloguing Rules’ at the British Museum and Charles Coffin Jewett’s efforts to mechanically duplicate entries at the library of the Smithsonian Institution [Zeng and Qin 2022, pp. 14-15].

Unique metadata standards, rules and models have been established and maintained within specific sub-fields. In addition, certain standards for information resources have been endorsed by authoritative bodies [Greenberg 2005], and some are used exclusively within specific domain communities [Hillmann et al. 2008]. Riley [2017, p. 5] underscores the predilection of CH metadata – whether these standards emanate from libraries, archives, or museums – toward accentuating descriptive attributes. The foundational CH metadata standards, primarily conceived to ‘describe what resources are and what they are about, and organising them according to controllable criteria’ [Zeng and Qin 2022, p. 11], manifest this thematic focus. Within the CH domain, metadata standards vary widely in scope, and a number of different standards have been developed to meet different needs and priorities³⁷ [Freire et al. 2018a].

The following quoted passage sheds some light on the different approaches and levels of collaboration in metadata standardisation, namely among the library and museum sectors.

‘Despite the striving for homogeneity, in practice, the production of metadata among information specialists and the use of metadata standards is already marked by considerable diversity. This has come about for very pragmatic reasons. Different types of objects and collections require different types of metadata. The curatorial interest for particular information differs for example between images held in an art gallery and a library, as does the information specialists’ domain expertise. Accordingly, diversity in metadata practice seems to be greatest in museums as they are the institutions that govern the most diverse collections. While the library sector has ‘systematically and cooperatively created and shared’ metadata standards since the 1960s, the museum sector, mostly handling images and objects, has been slower to establish such collaboration and consensus.’ [Dahlgren and Hansson 2020, p. 244]

In this context, I want to focus on some metadata standards that have proved vital across libraries, archives, museums and galleries. These standards, which I will briefly describe, serve as the foundation for organising, describing, and enabling efficient access to vast and diverse collections. Of particular interest I will be taking a closer look at CIDOC Conceptual Reference Model (CIDOC-CRM) as it serves as the cornerstone of Linked Art, a fundamental LOUD standard.

³⁷ It is noteworthy that the diversity of metadata standards in the heritage domain, characterised primarily by a common emphasis on descriptive attributes, is not counter-intuitive. This variation reflects the diverse nature of CH resources and the nuanced needs of GLAMs.

3.2.3.1 Library Metadata Standards

In libraries, several metadata standards have played crucial roles in organising and accessing collections over the years. The most prevalent historical standard, Machine-Readable Cataloging (MARC)³⁸, was a pilot project from the 1960s funded by the Council on Library and Information Resources (CLIR) and led by the Library of Congress (LoC) to structure cataloguing data and distribute them through magnetic tapes [Avram 1968, p. 3]. The standard evolved into MARC21 in 1999 [Zeng and Qin 2022, p. 418] – as exemplified by Code Snippet 3.1, providing a structured format for bibliographic records and related information in machine-readable form. It uses codes, fields, and sub-fields to structure data. Another significant historical standard is the Anglo-American Cataloguing Rules (AACR), published in 1967 and revised in 1978 that provides sets of rules for descriptive cataloguing of various types of information resources.

Code Snippet 3.1: MARC21 Record of Zeng and Qin [2022] in the Swisscovery Platform

```

1 leader 01424nam a2200397 c 4500
2 001 991170746542405501
3 005 20220427104002.0
4 008 210818s2022 xxu b 001 0 eng
5 010 ##$a 2021031231
6 020 ##$a9780838948750 $qBroschur
7 020 ##$a0838948758
8 035 ##$a(0CoLC)1264724191
9 040 ##$aDLC $bger $erda $cDLC $dCH-ZuSLS UZB ZB
10 042 ##$apcc
11 050 00$aZ666.7 $b.Z46 2022
12 082 00$a025.3 $223
13 082 74$a020 $223sdb
14 100 1##$aZeng, Marcia Lei $d1956- $4aut $0(DE-588)136417035
15 245 10$aMetadata $cMarcia Lei Zeng and Jian Qin
16 250 ##$aThird edition
17 264 #1$aChicago $bALA Neal-Schuman $c2022
18 300 ##$axxvi, 613 Seiten $bIllustrationen
19 336 ##$btxt $2rdacontent
20 337 ##$bn $2rdamedia
21 338 ##$bnc $2rdacarrier
22 504 ##$aIncludes bibliographical references and index
23 650 #0$aMetadata
24 650 #7$aMetadata $2fast $0(0CoLC)fst01017519
25 650 #7$aMetadaten $2gnd $0(DE-588)4410512-5
26 776 08$iErscheint auch als $nOnline-Ausgabe $tMetadata $z9780838937969
27 776 08$iErscheint auch als $nOnline-Ausgabe $tMetadata $z9780838937952
28 700 1##$aQin, Jian $d1956- $4aut $0(DE-588)1056085541
29 856 42$3Inhaltsverzeichnis $qPDF $uhttps://urn.ub.unibe.ch/urn:ch:slsp:0838948758:ihv:pdf
30 900 ##$aOK_GND $xUZB/Z01/202203/klei
31 900 ##$aStoppsignal FRED $xUZB/Z01/202203
32 949 ##$ahttps://urn.ub.unibe.ch/urn:ch:slsp:0838948758:ihv:pdf

```

AACR is no longer maintained and was replaced by Resource Description and Access (RDA)³⁹ around 2010 to be a more adaptive standard to contemporary needs. RDA, while not a markup language like MARC, serves as a content standard that guides the description and discovery of resources, focusing on user needs and facilitating improved navigation of library collections. Its goal is to provide a flexible and extensible framework for the description of all types of resources, ensuring discoverability, accessibility, and relevance for users⁴⁰ [Sprochi 2016, p. 130].

³⁸ MARC Standards: <https://www.loc.gov/marc/>

³⁹ RDA: <https://www.loc.gov/aba/rda/>

⁴⁰ If RDA was initially envisioned as the third edition of AACR, it faces the challenge of maintaining a delicate balance between preserving the AACR tradition while embracing the necessary shifts required for a successful and relevant future for library catalogues that can easily be interconnected with standards from archives, museums, and other communities [see Coyle and Hillmann 2007].

Libraries often leverage other standards to enrich their metadata practices. Metadata Object Description Schema (MODS)⁴¹, introduced in 2002, offers a more flexible XML-based schema for bibliographic description, allowing for better integration with other standards and systems. It was initially developed to carry ‘existing MARC21 records as well as to enable the creation of original resource description records’ [Zeng and Qin 2022, p. 423]. MODS provides a balance between human readability and machine processing, making it suitable for a wide range of resources and use cases [Guenther 2003, p. 139]. Metadata Encoding and Transmission Standard (METS)⁴², on the other hand, is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library. METS, developed as an initiative of the Digital Library Federation (DLF), provides a flexible and extensible framework for structuring metadata, allowing for the packaging of complex digital objects [Cantara 2005, pp. 238-239]. While MODS is primarily concerned with bibliographic information, METS focuses on structuring metadata for digital objects, making it particularly useful for digital libraries and repositories.

A further important standard is FRBR, a conceptual framework for understanding and structuring bibliographic data and access points. Originally developed by IFLA in 1997 as part of its functional requirements family of models, FRBR describes three main groups of entities, relationships, and attributes as illustrated by Figure 3.2⁴³. The first group of entities are the foundation of the model which characterises four levels of abstraction: Work, Expression, Manifestation, Item (WEMI) [Denton 2006, p. 231]. FRBR has had a significant impact on the development of RDA, which is loosely aligned with the principles and structures defined by the conceptual framework, but as it isn’t a data model *per se*; it does not inform how to record bibliographic information in day-to-day practice and focus heavily on textual resources⁴⁴ [Sprochi 2016, pp. 130-131]. Furthermore, Cossham [2017, p. 11] asserts that FRBR and RDA, ‘[don’t] align well with the ways that users use, understand, and experience library catalogues nor with the ways that they understand and experience the wider information environment.’

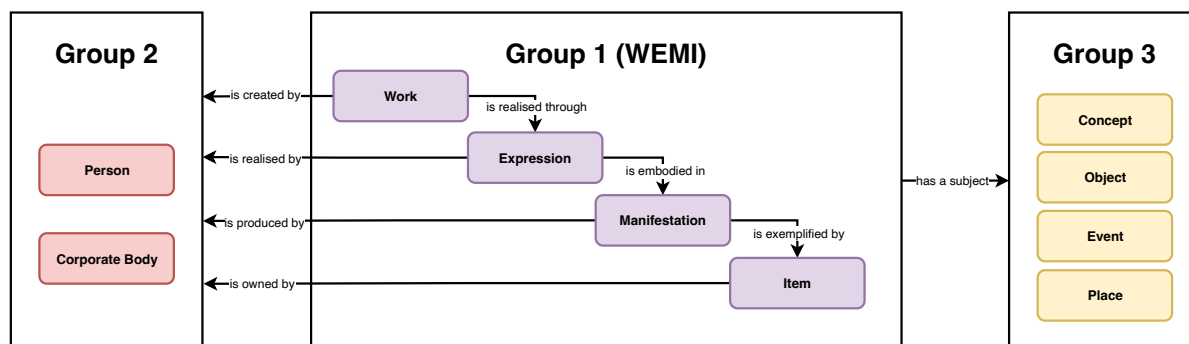


Figure 3.2: The FRBR Conceptual Framework. Adapted from Zou et al. [2018, p. 36]

A further important standard in the field of library science is the IFLA Library Reference Model (LRM), which was introduced as a comprehensive conceptual framework. It provides a broad understanding of bibliographic data and user-centric design principles, aligning with FRBR. LRM defines key entities, attributes, and relationships important for bibliographic searches, interpretation, and navigation – as shown in Figure 3.3⁴⁵. It operates at the conceptual level and does not dictate data storage methods. Attributes in LRM can be represented as literals or URIs. The model is presented in a struc-

⁴¹ MODS: <https://www.loc.gov/standards/mods/>

⁴² METS: <https://www.loc.gov/standards/mets/>

⁴³ This figure is also available in an enlarged format in Figure B.1.

⁴⁴ People might even argue that FRBR is only interesting as an ‘intellectual exercise’ [Zumer 2007, p. 27].

⁴⁵ This figure is also available in an enlarged format in Figure B.2.

tured document format to support LOD applications and reduce ambiguity. During its development, a parallel process created FRBRoo (see subsection 3.2.3.3), a model that extends the original FRBR model by incorporating it into CIDOC-CRM. FRBRoo focuses on CH data and is more detailed than LRM, which is designed specifically for library data and follows a high-level, user-centric approach [Riva et al. 2017, pp. 9-13]. The LRM model, known as LRMer⁴⁶, was released in 2020 by IFLA [Zeng and Qin 2022, p. 163].

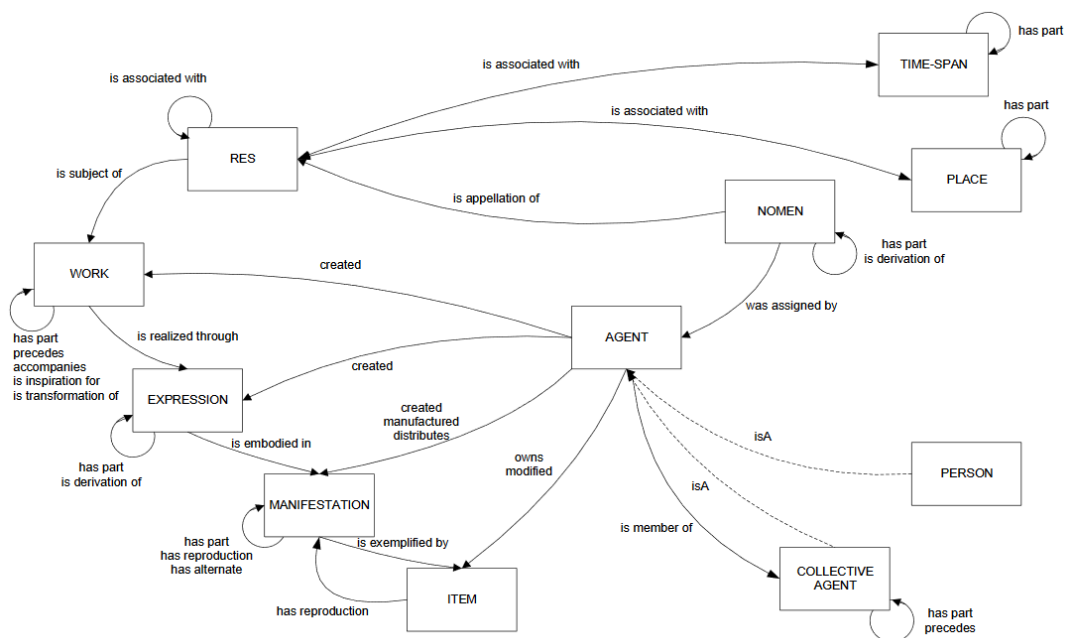


Figure 3.3: Overview of Relationships in LRM [Riva et al. 2017, p. 86]

Bibliographic Framework (BibFrame)⁴⁷ is another metadata standard in the library domain. It was initiated around 2011 by the LoC to be a successor of MARC, which had become obsolete [see Tennant 2002] as well as being invisible to web crawlers and search engines preventing adequate discoverability of bibliographic resources [Sprochi 2016, p. 132]. BibFrame is a loosely RDF-based model [Sanderson 2015], intending to improve the interoperability and discoverability of library resources. While the BibFrame model may not perfectly correspond with the WEMI entities outlined in FRBR, it is possible to effectively link BibFrame resources to FRBR entities, ensuring their compatibility [Sprochi 2016, p. 133]. BibFrame aims to transition from MARC by providing a more web-friendly framework, focusing on the relationships between entities, improving data sharing, and accommodating the digital environment. Conversely, Edmunds [2023] argues that BibFrame is unaffordable and leads to elitism within libraries, with the main beneficiaries being well-funded institutions, particularly in North America, while placing a financial burden on others. This approach, endorsed by bodies such as the LoC, is criticised for its high cost, impracticality, inequity and limited benefits for cataloguers, libraries, vendors and the public they serve. In addition, the author highlights BibFrame's lack of user friendliness, regardless of the intended users, and criticises the notion of adopting Linked Data for its own sake without substantial practical benefits.

⁴⁶ LRMer: <https://www.iflstandards.info/lrm/lrmer>

⁴⁷ BibFrame: <https://www.loc.gov/bibframe/>

3.2.3.2 Archival Metadata Standards

For archives, metadata standards like Encoded Archival Description (EAD)⁴⁸ and General International Standard Archival Description (ISAD(G))⁴⁹ have been pivotal. EAD, introduced in the mid-1990s – it originated in 1993 and the first version of EAD was released in 1998, provides a hierarchical structure for representing information about archival collections, offering comprehensive descriptions that aid researchers, archivists, and institutions in managing and providing access to archival records. Its goal is to create a standard for encoding finding aids to improve accessibility and understanding of archival collections [Pitti 1999, pp. 61-62]. On the other hand, ISAD(G), released in its first version in 1994 by ICA, offers a more general international standard for archival description, providing a framework for describing all types of archival materials, including fonds, sub-fonds, series, files, and items [Shepherd and Smith 2000, p. 57]. ISAD(G) aims to establish consistent and standardised archival description practices on a global scale, facilitating the sharing and exchange of archival information.

PREservation Metadata: Implementation Strategies (PREMIS)⁵⁰, is another metadata standard that was initially released in 2005 – version 3.0 is the latest specification, published in 2016 – and focuses on the preservation of digital objects, consisting of four interrelated entities: **Object**, **Event**, **Agent**, and **Rights** [Caplan and Guenther 2005, p. 111]. The main objective of PREMIS is to help institutions ensure the long-term accessibility of data by capturing key details about their creation, format, provenance, and preservation events. It is seen as an elaboration of OAIS, which categorises information required for preservation in several functional entities and types of information package [see Lee 2009, pp. 425-426] – as illustrated by Figure 3.4, expressed through the mapping of preservation metadata onto the conceptual model [Zeng and Qin 2022, pp. 493-494].

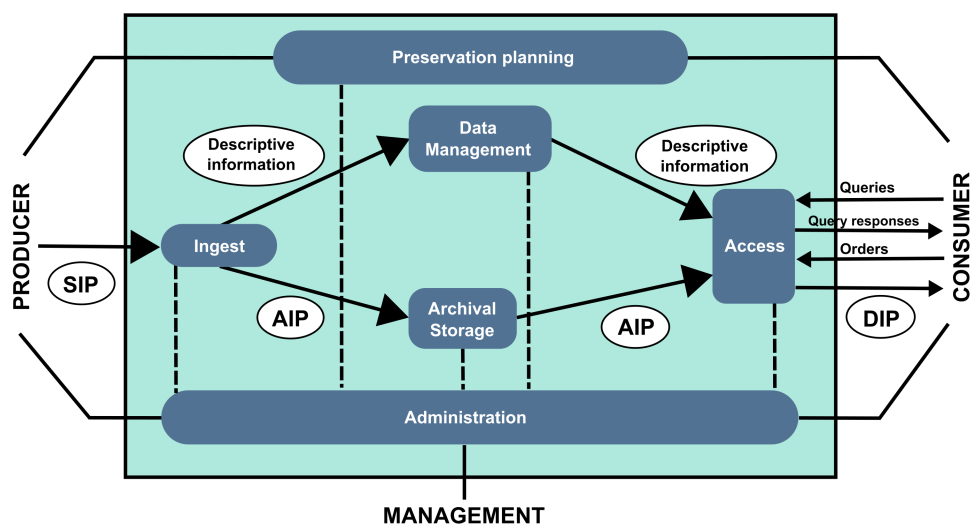


Figure 3.4: OAIS Functional Model Diagram by Mathieualexhache [2021]

The latest development in metadata standards for archives is the creation of RiC, which has been developed since 2012 by ICA [Clavaud and Wildi 2021, pp. 79-80]. RiC is structured into four complementary parts [ICA Expert Group on Archival Description 2023, p. 1] intended to cover and replace existing archival standards such as ISAD(G):

⁴⁸ EAD: <https://www.loc.gov/ead/>

⁴⁹ ISAD(G): General International Standard Archival Description - Second edition
<https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>

⁵⁰ PREMIS: <https://www.loc.gov/standards/premis/>

RiC Foundations of Archival Description A brief description of the foundational principles and purposes of archival description.

RiC Conceptual Model A high-level framework for archival description⁵¹, as shown in Figure 3.5.

RiC-O The ontology⁵², which embodies a specific implementation of the conceptual model. It is formally expressed in OWL to make archival description available using LOD techniques – which facilitating extensions [see [Mikhaylova and Metilli 2023](#)] – and adheres to a conceptual vocabulary specific to archival description. It provides the ability to navigate and interpret complex archival holdings and foster meaningful research and discovery. The ontology includes seven main groups of entities: Record, Agent, Rule, Event, Date, Place, and Instantiation.

RiC Application Guidelines A part in development at the time of writing which will provide practitioners and software developers with guidance and examples for implementing the conceptual model and the ontology in records and archival management systems.

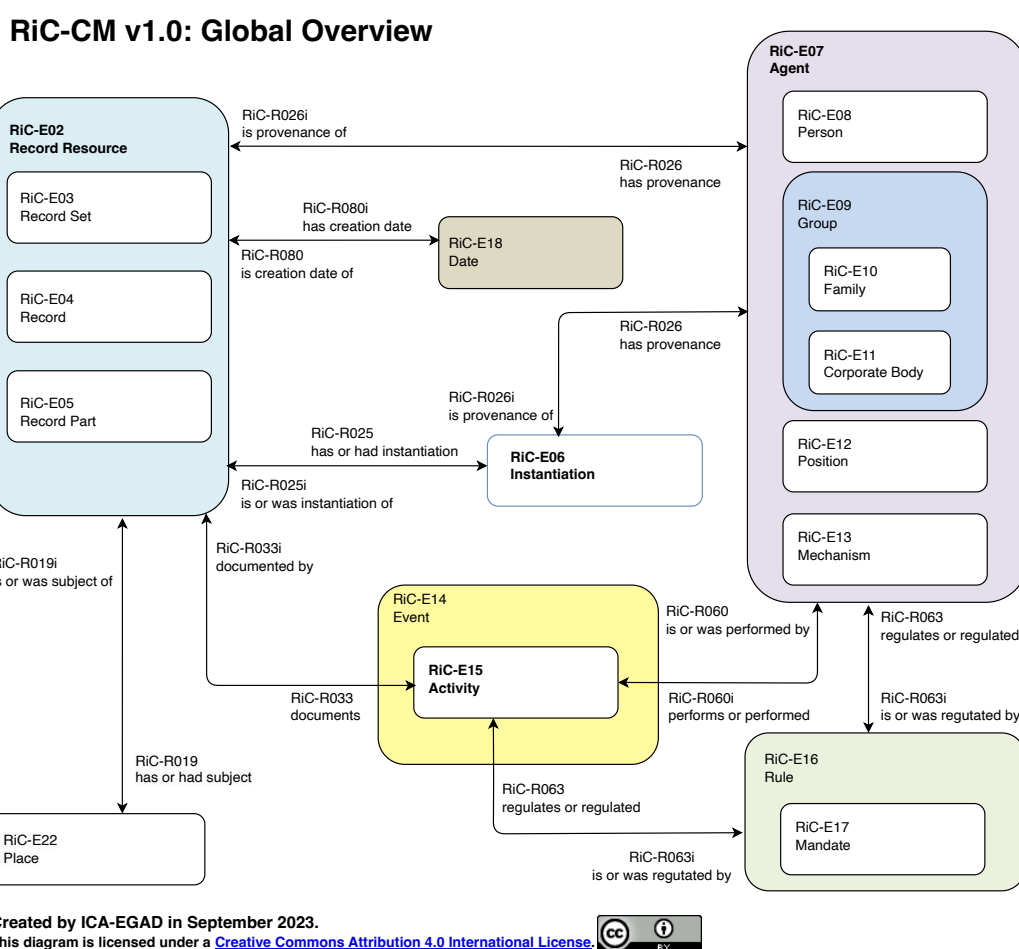


Figure 3.5: Global Overview of the Core Entities Defined by the RiC Conceptual Model. Slightly Adapted from <https://github.com/ICA-EGAD/RiC-O>

⁵¹ RiC Conceptual Model: <https://www.ica.org/en/records-in-contexts-conceptual-model>

⁵² RiC-O: <https://www.ica.org/standards/RiC/ontology>

3.2.3.3 Museum and Gallery Metadata Standards

In the museum and gallery domain, various metadata standards and conceptual models have significantly contributed to the management, organisation, and accessibility of CH objects and artworks. Notable among these are Categories for the Description of Works of Art (CDWA), Cataloging Cultural Objects (CCO), LIDO, CIDOC-CRM, as well as Linked Art.

CDWA⁵³, developed in the mid-1990s and maintained by the Getty Vocabulary Program, and CCO⁵⁴ created by the Visual Resource Association (VRA)⁵⁵, introduced in the early 2000s, primarily focus on describing art and cultural artefacts, providing a framework for recording essential details like artist, title, medium, date, and provenance. CDWA is a comprehensive set of guidelines for cataloguing and describing various cultural objects, including artworks, architectural elements, material culture items, collections of works, and associated images. While not a data model itself, it offers a conceptual framework for designing data models and databases, as well as for information retrieval. It then evolved into CDWA Lite, an XML schema for data harvesting purposes [Baca and Harpring 2017, pp. 1-2].

CCO comprises of both rules and examples of the CDWA categories and the VRA Core 4.0 for describing, documenting, and cataloguing cultural works and their visual surrogates⁵⁶ [Coburn et al. 2010a, pp. 17-18]. Both CCO and CDWA are standards that the ICOM International Committee for Documentation (CIDOC)⁵⁷ recommends and supports for museum documentation.

LIDO⁵⁸ is a CIDOC standard introduced in the early 2000s which offers a lightweight XML-based serialisation used for describing museum-related information – as shown in Code Snippet 3.2. It provides a format for the interchange of data about art and CH objects, complementing CDWA and CCO as it integrates and extends CDWA Lite with elements of CIDOC-CRM [Stein and Balandi 2019, p. 1025]. Ultimately, LIDO's goal is to enhance interoperability, accessibility, and the sharing of collection information, enabling institutions to connect and showcase their collections in diverse contexts [Coburn et al. 2010b, p. 3]. LIDO is also a CIDOC Working Group, which are created to tackle particular issues or areas of interest⁵⁹.

Code Snippet 3.2: Example of a LIDO Object in XML from Lindenthal et al. [2023]

```

1 <lido:lido>
2   <lido:lidoRecID
3     lido:source="ld.zdb-services.de/resource/organisations/DE-Mb112"
4     lido:type="http://terminology.lido-schema.org/lido00099">
5     ld.zdb-services.de/resource/organisations/DE-Mb112/lido/obj/00076417
6   </lido:lidoRecID>
7   <lido:descriptiveMetadata xml:lang="en">
8     <lido:objectClassificationWrap>
9       <lido:objectWorkTypeWrap>
10        <lido:objectWorkType>
11          <skos:Concept
12            rdf:about="http://vocab.getty.edu/aat/300033799">
13            <skos:prefLabel
14              xml:lang="en">
15              oil paintings (visual works)
16            </skos:prefLabel>
17          </skos:Concept>

```

⁵³ CDWA: https://www.getty.edu/research/publications/electronic_publications/cdwa/

⁵⁴ CCO: <https://www.vraweb.org/cco>

⁵⁵ VRA: <https://www.vraweb.org/>

⁵⁶ VRA Core 4.0 and CCO have a symbiotic relationship, with CCO providing data content guidelines and incorporating the VRA Core 4.0 methodology. The latter also been leveraged in other contexts to form the basis for more granular Linked Data vocabularies [see Mixter 2014].

⁵⁷ In French, the original language used for this acronym, CIDOC stands for *Comité international pour la documentation du Conseil international des musées*.

⁵⁸ LIDO: <https://cidoc.mini.icom.museum/working-groups/lido/>

⁵⁹ CIDOC Working Groups: <https://cidoc.mini.icom.museum/working-groups/>

```

18     </lido:objectWorkType>
19   </lido:objectWorkTypeWrap>
20 </lido:objectClassificationWrap>
21 <lido:objectIdentificationWrap>
22   <lido:titleWrap>
23     <lido:titleSet>
24       <lido:appellationValue
25         lido:pref="http://terminology.lido-schema.org/lido00169"
26         xml:lang="en">
27         Mona Lisa
28       </lido:appellationValue>
29     </lido:titleSet>
30   </lido:titleWrap>
31 </lido:objectIdentificationWrap>
32 </lido:descriptiveMetadata>
33 </lido:lido>

```

CIDOC-CRM⁶⁰, developed since 1996 by the CIDOC and more specifically maintained by the CIDOC-CRM Special Interest Group (CRM-SIG) — which convenes quarterly⁶¹, is a formal and top-level ontology that offers a comprehensive conceptual framework for describing CH resources, allowing for a deep understanding of relationships between different entities, events, and concepts for museums [Doerr 2003, pp. 75-76]. It aims to provide a common semantic framework for information integration, supporting robust knowledge representation and fostering collaboration and interoperability within the CH sector as it can also mediate different resources from libraries and archives. The latest stable version of the conceptual model is version 7.1.2⁶², published in June 2022, and comprises of 81 classes and 160 properties⁶³ [see Bekiari et al. 2021].

Within the base ontology of CIDOC-CRM – or CRMBase – and despite the emergence of new developments and gradual changes, there is a fundamental and stable core that can be succinctly outlined. This fundamental structure acts as a basic orientation for understanding the way in which data is structured within CIDOC-CRM. Examining the hierarchical structure of CIDOC-CRM, one can identify the main top-level branches, namely:

- E18 Physical Thing** This class comprises all persistent physical items with a relatively stable form, human-made or natural.
- E28 Conceptual Object** This class comprises non-material products of our minds and other human produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication. The production of such information may have been supported by the use of technical devices such as cameras or computers.
- E39 Actor** This class comprises people, either individually or in groups, who have the potential to perform intentional actions of kinds for which someone may be held responsible.
- E53 Place** This class comprises extents in the natural space we live in, in particular on the surface of the Earth, in the pure sense of physics: independent from temporal phenomena and matter. They may serve describing the physical location of things or phenomena or other areas of interest.
- E2 Temporal Entity** This class comprises all phenomena, such as the instances of **E4 Periods** and **E5 Events**, which happen over a limited extent in time.

⁶⁰ CIDOC-CRM: <https://cidoc-crm.org/>

⁶¹ CRM-SIG Meetings: https://www.cidoc-crm.org/meetings_all

⁶² CIDOC-CRM V7.1.2: <https://www.cidoc-crm.org/html/cidoc-crm-v7.1.2.html>

⁶³ For a quick overview of the classes and properties of CIDOC-CRM, I recommend visiting the dynamic periodic table created by Remo Grillo (Digital Humanities Research Associate at *I Tatti*, Harvard University Center for Italian Renaissance Studies): https://remogrillo.github.io/cidoc-crm_periodic_table/

Complemented by entities tailored for the documentation of E41 Appellation and E55 Type, the structure – as shown in Figure 3.6 – provides a potent set of means to capture a broad range of general-level CH reasoning in a holistic manner [Bruseker et al. 2017, pp. 111-112].

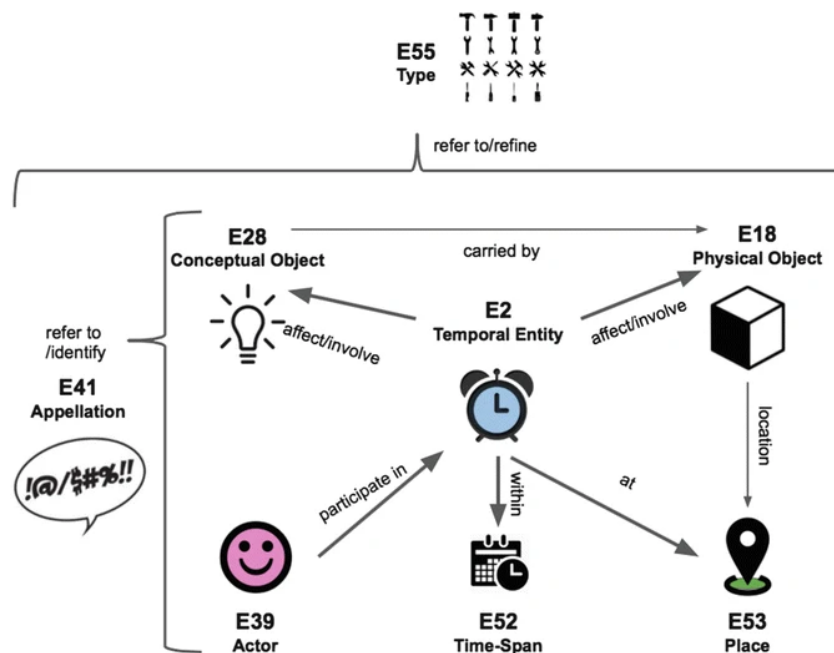


Figure 3.6: CIDOC-CRM Top-Level Categories by Bruseker et al. [2017, p. 112]

CRMBase, is supplemented by a series of extensions – sometimes referred to as the CIDOC-CRM family of models – intended to support various types of specialised research questions and documentation, such as bibliographic records or geographical data. These compatible models⁶⁴, ordered alphabetically, include both works in progress and models to be reviewed by CRM-SIG⁶⁵. They comprise as follows:

CRMact⁶⁶ An extension that defines classes and properties for integrating documentation records about plans for future activities and future events.

CRMarchaeo⁶⁷ An extension of CIDOC-CRM created to support the archaeological excavation process and all the various entities and activities related to it.

CRMba⁶⁸ An ontology for documenting archaeological buildings. Its primary purpose is to facilitate the recording of evidence and material changes in archaeological structures.

CRMdig⁶⁹ An ontology to encode metadata about the steps and methods of production (‘provenance’) of digitisation products and synthetic digital representations such as 2D, 3D or even animated models created by various technologies.

⁶⁴ CIDOC-CRM compatible models and collaborations: <https://www.cidoc-crm.org/collaborations>

⁶⁵ At the time of writing none of these CIDOC-CRM extensions have been formally approved by CRM-SIG. It is also worth mentioning that other extensions based on CIDOC-CRM have been developed by the wider community, such as Bio CRM, a data model for representing biographical data for prosopographical research [see Tuominen et al. 2017] or ArchOnto, which is a model created for archives [see Koch et al. 2020].

⁶⁶ CRMact: <https://www.cidoc-crm.org/crmact/>

⁶⁷ CRMarchaeo: <https://www.cidoc-crm.org/crmarchaeo/home-3> — this is a curiously encoded URL that may not be permanently accessible.

⁶⁸ CRMba: <https://www.cidoc-crm.org/crmba/>

⁶⁹ CRMdig: <https://www.cidoc-crm.org/crmdig/>

- CRMgeo**⁷⁰ An ontology intended to be used as a global schema for integrating spatio-temporal properties of temporal entities and persistent items. Its primary purpose is to provide a schema consistent with the CIDOC-CRM to integrate geoinformation using the conceptualisations, formal definitions, encoding standards and topological relations.
- CRMinf**⁷¹ An extension of CIDOC-CRM that facilitates argumentation and inference in descriptive and historical fields. It serves as a universal schema for merging metadata related to argumentation and inference, primarily focusing on these disciplines.
- CRMsci**⁷² The Scientific Observation Model is an ontology that extends CIDOC-CRM for scientific observation, distinguishing the process from results and providing a formal ontology for scientific data integration and research modelling.
- CRMsoc**⁷³ An ontology for integrating data about social phenomena and constructs that are of interest in the humanities and social sciences based on analysis of documentary evidence.
- CRMtex**⁷⁴ An extension of CIDOC-CRM created to support the study of ancient documents by identifying relevant textual entities and by modelling the scientific process related with the investigation of ancient texts and their features.
- FRBRoo**⁷⁵ An ontology intended to capture and represent the underlying semantics of bibliographic information which interprets the conceptualisations of the FRBR framework.
- PRESSoo**⁷⁶ An ontology intended to capture and represent the underlying semantics of bibliographic information about continuing resources, and more specifically about periodicals (journals, newspapers, magazines, etc.). PRESSoo is also an extension of FRBRoo.

Figure 3.7 shows CRMbase and eight of the extensions previously outlined in a pyramid shape, where the lower you go in the pyramid, the more specialised the concepts.

Linked Art⁷⁷, a recent addition to this landscape, is a community-driven initiative and a metadata application profile that has been in existence since the end of 2016 [Raemy 2022a, pp. 136-137]. This community – recognised as a CIDOC Working Group – has created a common Linked Data model based on CIDOC-CRM for describing artworks, their relationships, and the activities around them (see subsection 3.5.5).

3.2.3.4 Cross-domain Metadata Standards

There are a few cross-domain standards that have been used to describe CH resources. For instance, the Dublin Core Elements, containing the original core sets of fifteen basic elements, and Dublin Core Metadata Terms⁷⁸, its extension, are widely used metadata standards for describing CH resources. It provides metadata properties and classes that are applicable to a wide range of resources [Weibel and Koch 2000]. Another good example is the Europeana Data Model (EDM) that has been specified so that national, regional and thematic aggregators in Europe can deliver resources of content providers to Europeana [see Charles and Isaac 2015, Freire and Isaac 2019].

⁷⁰ CRMgeo: <https://www.cidoc-crm.org/crmgeo/>

⁷¹ CRMinf: <https://www.cidoc-crm.org/crminf/>

⁷² CRMsci: <https://www.cidoc-crm.org/crmsci/>

⁷³ CRMsoc: <https://www.cidoc-crm.org/crmsoc/>

⁷⁴ CRMtex: <https://www.cidoc-crm.org/crmtex/>

⁷⁵ FRBRoo: <https://www.cidoc-crm.org/frbroo/>

⁷⁶ PRESSoo: <https://www.cidoc-crm.org/pressoo/>

⁷⁷ Linked Art: <https://linked.art>

⁷⁸ DCMI Metadata Terms: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

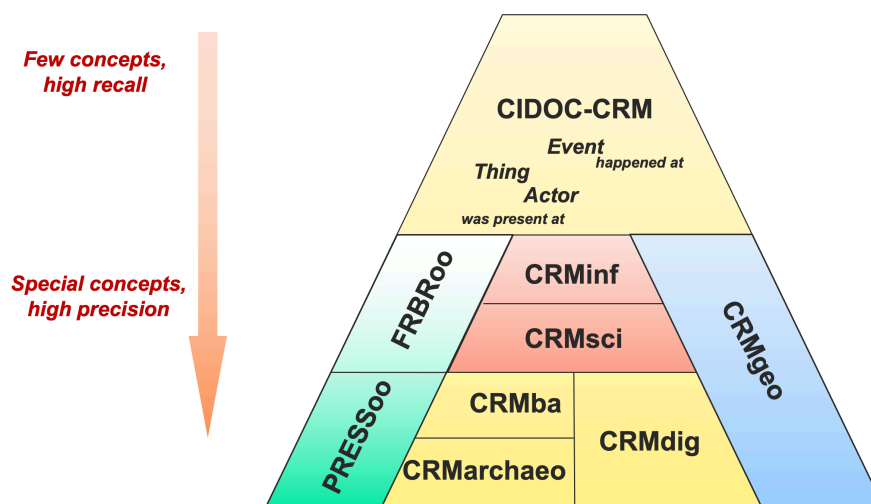


Figure 3.7: CIDOC-CRM Family of Models. Diagram Done and Provided by Maria Theodoridou (Institute of Computer Science, FORTH)

Despite the presence of cross-domain standards and efforts to map between standards, whether from one version to another or across different domains, reconciling metadata from various sources remains a significant challenge in the CH sector. Institutions may collect metadata in different ways, using different standards and schemas, making it difficult to merge and compare metadata from different sources. Additionally, metadata may be incomplete, inconsistent, or contain errors, further complicating data reconciliation. To address these challenges, standardised, interoperable metadata are necessary to enable data sharing and reuse. While the use of different metadata standards can present challenges for data reconciliation, the adoption of standardised, interoperable metadata can facilitate data sharing and reuse, promoting the long-term preservation and accessibility of CH resources. Controlled vocabularies – included in what Zeng and Qin [2022, pp. 24-25] called ‘*standards for data values*’ – such as those maintained by the Getty Research Institute⁷⁹: the Art & Architecture Thesaurus (AAT), the Thesaurus of Geographic Names (TGN), and the Union List of Artist Names (ULAN), as well as various kinds of KOS (see subsection 3.2.4). These vocabularies provide a common language for describing CH objects and can improve the interoperability of metadata across different institutions and communities.

Alongside metadata reconciliation comes also the question of aggregation. Apart from LIDO in museums, the general and current operating model for aggregating CH (meta)data is still the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [see Raemy 2020], which is an XML-based standard that was initially specified in 1999 and updated in 2002 [Lagoze et al. 2002]. Alas, OAI-PMH does not align to contemporary needs [Van de Sompel and Nelson 2015], and there are now some alternative and web-based technologies for harvesting resources that are slowly being leveraged such as Activity Streams (AS) [Snell and Prodromou 2017], a W3C syntax and vocabulary for representing activities and events in social media and other web application. It can also be easily extended and used in different contexts, such as it is the case with the IIIF Change Discovery API (see subsection 3.5.3.3) or with ActivityPub [Lemmer-Webber and Tallon 2018], a decentralised W3C protocol being leveraged by Mastodon⁸⁰, a federated and open-source social network.

⁷⁹ Getty Vocabularies: <https://www.getty.edu/research/tools/vocabularies/>

⁸⁰ Mastodon: <https://joinmastodon.org/>

Overall, the evolution of metadata standards in the CH domain paves the way for a more interconnected and accessible digital environment, thereby providing better access to disparate collections and facilitating cross-domain reconciliation. This transformation is complemented by a growing emphasis on web-based metadata aggregation technologies that are more suited to today’s needs.

3.2.4 Knowledge Organisation Systems

Knowledge Organisation Systems (KOS), also known as concept systems or concept schemes, encompass a wide range of instruments in the area of knowledge organisation. They are distinguished by their specific structures and functions [Mazzocchi 2018, p. 54]. KOS include authority files, classification schemes, thesauri, topic maps, ontologies, and other related structures. Despite their differences in nature, scope and application, all share a common goal: to facilitate the structured organisation of knowledge and classification of information. According to Zeng and Qin [2022, p. 284], ‘KOS have a more important function: to model the underlying semantic structure of a domain and to provide semantics, navigation, and translation through labels, definitions, typing relationships, and properties for concepts’. This overarching intent underpins the practice of information management and retrieval.

The term KOS ‘became even more popular after the encoding standard Simple Knowledge Organization System (SKOS) was recommended by W3C’, although the use of such systems can be traced back over 100 years, whereas others have been created in the advent of the web [Zeng and Qin 2022, p. 188]. According to Hill et al. [2002, pp. 46-47, citing Hodge [2000]], KOS can be divided into four main groups: term lists, metadata-like models, classification and categorisation, as well as relationship models.

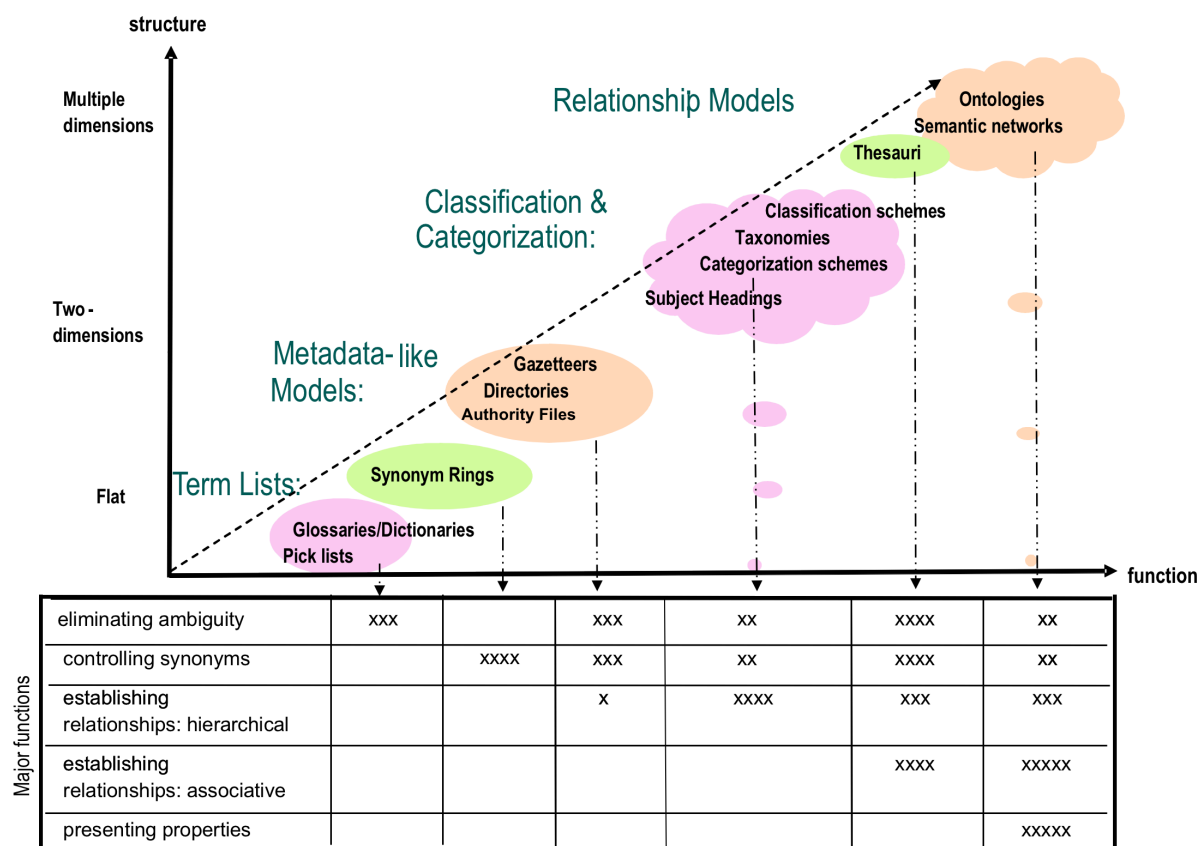


Figure 3.8: Overview of the Structures and Functions of KOS [Zeng 2008, p. 161]

Term lists encompass authority files, dictionaries, and glossaries, serving as controlled sources for managing terms, definitions, and variant names within a knowledge organisation framework. Metadata-like models encompass directories and gazetteers, offering lists of names and associated contact information as well as geospatial dictionaries for named places, with can be extended for representing events and time periods. In the classification and categorisation domain, you find categorisation schemes and classification schemes that organise content, subject headings that represent controlled terms for collection items, and taxonomies that group items based on specific characteristics. Finally, relationship models feature ontologies, semantic networks, and thesauri, each capturing complex relationships between concepts and terms [Hill et al. 2002, Zeng 2008]. Figure 3.8 represents an overview of the structure and functions of these four main groups, showcasing as well the subcategories of KOS previously mentioned. In this figure, the x characters indicate the extent to which each type of KOS embodies five key functions identified by Zeng [2008], such as eliminating ambiguity or controlling synonyms.

In this subsection, I will explore four subcategories of KOS, each representing a continuum from a more linear to a more structured network. These include folksonomy, taxonomy, thesaurus, and ontology. These KOS have been selected due to their significant impact on the organisation and interlinking of data within the contexts of CHI practices and LOD. Furthermore, the intent of these systems is to help bridge the gap between human understanding and machine processing.

3.2.4.1 Folksonomy

Positioned at one end of the organisational spectrum, folksonomies, also known as community tagging or social bookmarking, are characterised by their user-generated nature. These systems rely on individual users' tagging of content with keywords or tags that reflect their personal perspectives and preferences. Folksonomies '*raise a substantial challenge to the authority and effectiveness of controlled semantics*' as integration or reconciliation is often hard to achieve [Zeng and Qin 2022, p. 401]. However, they do provide a wealth of source material for studying social semantics [Zeng and Qin 2022, p. 403] and can be done in parallel to more structured KOS.

3.2.4.2 Taxonomy

Moving towards the centre of the spectrum, taxonomies present a more structured approach to knowledge organisation. '*A taxonomy is a type of KOS which consists of preferred terms, all of which are connected in a hierarchy or polyhierarchy*' [Zeng 2008, p. 169].

Taxonomies employ hierarchical classifications to systematically categorise information into distinct classes and sub-classes, or in a parent/child relationship [SAA Dictionary 2023] - as shown by Code Snippet 3.3 [NISO 2010, p. 18]. Taxonomy, in this context, extends beyond mere categorisation; it also establishes relationships.

Code Snippet 3.3: Taxonomy Hierarchy

```
1 Chemistry
2   Physical Chemistry
3     Electrochemistry
4       Magnetohydrodynamics
```

3.2.4.3 Thesaurus

Moving further along the spectrum, thesauri offer a more detailed and formalised method of organisation. They include not only hierarchical relationships but also explicit semantic connections between terms, making them valuable tools for information retrieval. As defined by NISO [2010, p. 9]:

‘A thesaurus is a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators.’

For instance, consider a thesaurus related to photography, which encompasses categories for various aspects of photography, including photographic techniques, equipment, and materials. Within this taxonomy, ‘Kodachrome’ could be categorised not only as a specific type of colour film but also as a distinct photographic process. As a type, it could fall under the sub-category of ‘colour film photography’, and as a process, it would fit within the broader framework of ‘photographic techniques’.

The AAT, commonly employed in the CH domain, stands as a significant example of a thesaurus [Harpring 2010, p. 67]. Homosaurus⁸¹ is another example of a thesaurus with a distinct focus on enhancing the accessibility and discoverability of lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual, and more (LGBTQ+) resources and related information. Leveraging Homosaurus in metadata can effectively contribute to diminishing biases present in such data, an essential step in promoting inclusivity and equity within information systems [see Hardesty and Nolan 2021].

3.2.4.4 Ontology

At the structured end of the spectrum, ontologies define complex relationships and attributes between concepts, whereby a series of concepts have been chosen to express what we understand, so that a computer can start making sense of our world. Ontologies are formalised KOS, enabling advanced data integration and Knowledge Representation (KR) for more sophisticated applications. The term is drawn from philosophy, where an ontology is a discipline concerned with studying the nature of existence, as articulated by Gruber [1993, pp. 199-200]:

‘An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what “exists” is exactly that which can be represented.’

There are different kinds of ontologies, including axiomatic formal ontologies, foundational ontologies, and domain-specific ontologies [Beretta 2022]. These different types of ontologies cater to various knowledge representation needs. Foundational ontologies, such as Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)⁸², provide a high-level framework for modelling knowledge and offers a comprehensive system for representing entities, qualities, and relationships [see Borgo et al. 2022, Masolo et al. 2003].

Description logics (DLs), a family of formal KR languages, play also a key role in developing ontologies and serve as the foundation for OWL (see subsection 3.4.2), notably by providing a logical formalism. DLs are characterised by their ability to provide substantial expressive power that goes well beyond propositional logic, while maintaining decidable reasoning [Chang et al. 2014].

In computer science, the concepts of Assertion Box (ABox) and Terminological Box (TBox), both statements in KBs, are relevant to the structuring and enrichment of KGs [Giacomo and Lenzerini 1996]⁸³. The ABox, representing the ‘assertion’ or ‘instance’ level, encapsulates concrete data instances and their relationships, contributing to the factual knowledge of a given system. Conversely, the TBox,

⁸¹ Homosaurus: <https://homosaurus.org/>

⁸² DOLCE: www.loa.istc.cnr.it/dolce/overview.html

⁸³ It must be noted though that the use of DLs in KR predates the emergence of ontological modelling in the context of the Web, with its origins going back to the creation of the first DL modelling languages in the mid-1980s [Krötzsch et al. 2013].

representing the ‘terminology’ or ‘schema’ level, defines the conceptual framework and hierarchies that govern the relationships and attributes of the instances. These two complementary components work in harmony to improve data interoperability, reasoning and knowledge sharing. Figure 3.9 depicts a high-level overview of a KB representation system.

Consider a scenario around artwork provenance held in a museum. The ABox strives to encapsulate the rich narratives of individual artworks, tracing their journey through time, ownership transitions and exhibition travels. At the same time, the TBox creates a conceptual scaffolding, imbued with classes such as *Artwork*, *Creator*, and *Exhibition*, painting an abstract portrait that contextualises each artefact within a broader cultural tapestry. It is here that the DL comes in, harmonising the symphony with its logical relationships and axioms, i.e. a rule or principle widely accepted as obviously true [Baader and Lutz 2007]. The DL is represented as $\mathcal{K} = (\mathcal{T}, \mathcal{R}, \mathcal{A})$, where:

\mathcal{T} represents the TBox, defining the conceptual framework, which encompasses the hierarchical relationships, classes, and concepts within the KB.

\mathcal{R} represents the set of binary roles, delineating the relationships and connections between individuals or instances in the domain. These roles facilitate the understanding of how entities relate to one another within the KB.

\mathcal{A} represents the ABox, encompassing the specific assertions or instances in the KB.

This symbiotic interplay ensures that the provenance of each artwork is not just a static account, but a dynamic, interconnected narrative. The ABox-TBox relationship thrives in the realm of reasoning. Imagine an axiom embedded in the TBox: ‘A work of art presented in an exhibition curated by a distinguished patron is of heightened cultural significance’, or here phrased in DL terms: $\exists \text{curates.Artwork.CulturalSignificance} \sqsubseteq \text{true}$. This axiom serves as a beacon to guide the system’s reasoning. When an ABox instance of an artwork is woven into an exhibition curated by a prominent authority, the DL-informed engine responds by inferring an enriched cultural value that

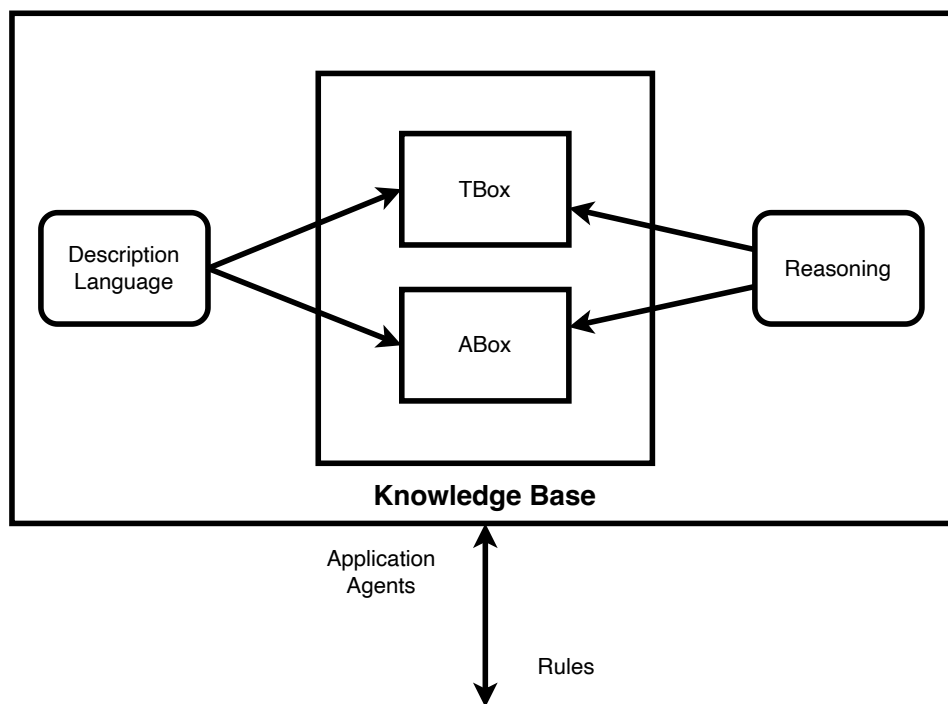


Figure 3.9: KB Representation System Based on Patrón et al. [2011, p. 205]

resonates beyond the artefact itself. This is where the TBox takes data and gives it life, producing insights that transcend the boundaries of individual instances. The KB, \mathcal{K} , captures this orchestration, encapsulating the logical relationships for meaningful interpretation and knowledge discovery.

Overall, the relationship between ABox and TBox in DL is vital for achieving semantic clarity, enabling meaningful data integration, and facilitating advanced reasoning mechanisms. The museum provenance scenario showcases a precisely orchestrated convergence of assertion, terminology, and rigorous logical reasoning. This engenders a computational landscape where historical artefacts intricately mesh within the complex network of human history's data structures, seamlessly aligning with the underlying framework of algorithmic representation. These components enable software developers to harmonise disparate datasets, extract insightful knowledge, and support decision-making processes across a wide range of domains. In essence, the use of DL, ABox, and TBox in ontological KR enhances interoperability between different systems and allows for sophisticated reasoning and decision support.

Moving beyond these foundational concepts, it is noteworthy to consider the work of [Ehrlinger and Wöß \[2016\]](#), who address the need for a clear and standardised definition of KGs. They highlight the term's varied interpretations since its popularisation by Google in 2012 and propose a definitive, unambiguous definition to foster a common understanding and wider adoption in both academic and commercial realms. They define a Knowledge Graph (KG) as follows: '*A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge*'. This definition crystallises the essence of KGs as dynamic and integrative systems that not only store but also process and enrich data through advanced reasoning. This conceptualisation underlines the transformative potential of KGs in various domains, bridging the gap between raw data and actionable insights.

Finally, it is important to recognise that the importance of ontologies extends beyond individual systems. Shared ontologies are a cornerstone of semantic interoperability, thus facilitating a paradigm shift in the way systems and applications communicate. As [Sanderson \[2013\]](#) argues: '*shared ontologies increases semantic interoperability*' and '*shared identity makes it possible for graph to merge serendipitously*'. This shared understanding ensures that various entities can seamlessly connect and engage in meaningful interactions.

3.3 Trends, Movements, and Principles

Technological trends, scientific movements, and guiding principles have played a crucial role in shaping the landscape of contemporary research. In recent years, there has been an increased emphasis on the need for academic and CH practices to be more transparent, inclusive, and accountable. This shift reflects a broader trend towards integrating advanced technological solutions and open-science principles in heritage management. As such, understanding the evolution of CH becomes imperative to comprehend how these practices have adapted and transformed in response to these guiding trends.

The evolution of CH has been characterised by a series of technological and methodological shifts. Initially, the primary focus was on digitising physical artefacts to preserve information from degrading originals. This phase was crucial for transitioning tangible CH into a digital format, mitigating the risk of loss due to physical degradation. Following this, efforts shifted towards ensuring the persistence of digitised resources. This stage involved addressing challenges related to digital preservation, including data degradation and format obsolescence, to ensure the longevity of digital cultural assets.

The advent of open data principles marked the next phase in CH development. This approach facilitated broader access to information, aligning with contemporary values of transparency and inclu-

sivity in, governmental, academic, and cultural contexts. Subsequently, the focus expanded to enhancing the utility of this data. This stage involved contextualising and enriching CH data, thereby increasing their applicability and relevance across various domains.

The current frontier in CH involves developing applications that leverage rich CH data. These applications serve not only as tools for engagement and education but also as justifications for the ongoing costs associated with data storage and archival. They illustrate the tangible benefits derived from preserving heritage resources, encompassing both cultural and economic returns. In summary, the trajectory of CH development mirrors broader technological and societal trends, transitioning from preservation to active utilisation. This progression underscores the dynamic nature of research and CH processes, highlighting the evolving requirements for transparency, inclusivity, and accountability in CH management.

While automation has significantly enhanced the efficiency of digitisation processes in CH, cataloguing and indexing remain complex challenges. The intricacies involved in accurately understanding and categorising resources necessitate more than just technological solutions; they require context-aware and culturally sensitive approaches. Here, ML offers promising perspectives. ML, particularly in its advanced forms like deep learning, can assist in cataloguing and indexing by analysing large datasets to identify patterns, categorise content, and even suggest metadata. This can be particularly useful in handling large volumes of CH data, where manual processing is time-consuming and prone to human error. Typical applications of ML in this field include image recognition for identifying and classifying visual elements in artefacts, Natural Language Processing (NLP) for analysing textual content, and pattern recognition for sorting and organising data based on specific characteristics. Furthermore, prospective developments may entail the refinement of metadata mapping and the enhancement of quality control mechanisms. Moreover, ML algorithms can be trained to recognise stylistic elements, historical contexts, and other nuances that are essential for accurate cataloguing in CH. However, it is crucial to note that the effectiveness of ML depends heavily on the quality and diversity of the training data. Biases in this data can lead to inaccuracies in cataloguing and indexing. Thus, a collaborative approach, where ML is supplemented by expert human oversight, is often the most effective strategy.

Overall, this section provides a comprehensive overview of six technological trends as well as five key scientific movements and guiding principles that are shaping research and how universities and GLAMs should provide environments, services, and tools with a view to collecting and disseminating content. By exploring each of these trends, movements, and principles, we can gain a deeper understanding of how research and CH processes are permeated by dynamic movements and how resources can be made more transparent, inclusive and accountable, as well as how data can be made available to human and non-human users.

3.3.1 Current and Emerging Technological Trends in Cultural Heritage

I will explore some current and emerging technological trends in CH, organised into three components: Linked Data, big data, and AI. Each represents a critical driver shaping the landscape and practices of heritage data. The three trends have been around for a few decades, with the ‘Linked Data’ principles and underlying standards coming from the late 1990s, ‘big data’ being coined in 1990 and AI in 1956.

Before considering the trends discussed hereafter, note that current technological developments do not exist in isolation, but tend to intertwine and act synergistically. A vivid example of this interplay can be seen in AI and its latent impact on the semantic web, particularly in facilitating more efficient

querying and crawling processes such as the LinkedDataGPT proof-of-concept service⁸⁴ from Liip on the City of Zurich that combines ChatGPT — a generative AI solution — on top of a Linked Data portal to facilitate querying open datasets [Stocker 2023]. Inversely AI can be fed by data on the web to learn and reason, as outlined by Gandon [2019a].

3.3.1.1 Linked Data

Linked Data, and most precisely LOD, is a set of design principles adhering to Resource Description Framework (RDF) which is a significant approach to interconnect data on the web in order to make semantic queries more useful [Berners-Lee et al. 2001]. In other words, this standardisation allows data to be not only linked, but also openly accessible and reusable. As noted by Gandon [2019a, p. 115, citing Gandon [2017]]:

‘The Web was initially perceived and used as a globally distributed hypertext space for humans. But from its inception, the Web has always been more: its hypermedia architecture is in fact linking programs world-wide through remote procedure calls.’

This deeper understanding of the web’s architecture as a conduit for linking programs on a global scale holds profound implications. It signifies that the web is not merely a medium for accessing information but a dynamic environment where data-driven programs interact, exchange data, and collaborate across geographical boundaries. In this context, Linked Data emerges as a powerful enabler, providing a structured and standardised approach for these programs to communicate and share meaningful data [Bizer et al. 2008].

In the context of CH, institutions such as museums, libraries and archives can publish their collections using Linked Data principles, enabling a web of linked information that is accessible to all. As this dissertation’s main topic revolves around Linked (Open) (Usable) Data, two dedicated sections have been written within this literature review in Section 3.4 and Section 3.5.

Beyond formal LOD, CHIs may also link their databases or collections in more informal ways. This interconnection may take the form of shared metadata, common identifiers, or simply hyperlinks. These links can enhance the user experience by supporting a more seamless navigation between related items or pieces of information. For instance, a parallel strategy is the use of graph-based data representation, i.e. property graph which consists of a set of objects or vertices, and a set of arrows or edges connecting the objects, that are most likely not RDF-compliant [see Bermès 2023]. Graph databases, such as Neo4j⁸⁵ which is quite prevalent in DH [see Darmont et al. 2020, Drakopoulos et al. 2019, Webber 2012], allow for efficient storage and retrieval of interconnected data through nodes representing entities and relationships linking them.

3.3.1.2 Big Data

Big Data refers to extremely large and complex datasets that exceed the capabilities of traditional data processing methods and tools. It encompasses a massive volume of structured, semi-structured and unstructured data that is currently flooding across a variety of sectors, companies and organisations [see Emmanuel and Stanier 2016]. The characteristics of big data are often described by the three *Vs* model [Laney 2001]:

Volume Big data refers to a massive amount of data. This can encompass a spectrum of data sizes,

⁸⁴ LinkedDataGPT: <https://ld.gpt.liip.ch/>

⁸⁵ Neo4j: <https://neo4j.com/>

extending from Gigabytes (GB) and Terabytes (TB), to Petabytes (PB)⁸⁶ and beyond. The sheer size of the data is a key aspect of big data, making traditional database systems inadequate for storage and analysis.

Velocity Data is being generated and collected at an unprecedented rate. Social media posts, sensor data, online transactions and more are constantly being generated, requiring real-time or near real-time processing and analysis.

Variety Big data comes in a variety of formats, including structured data (e.g. databases), semi-structured data (e.g. XML, JavaScript Object Notation (JSON)) and unstructured data (e.g. text, images, video). The variety of data types requires flexible processing methods.

In addition to the three *Vs* model, two more characteristics are often included [Saha and Srivastava 2014, p. 1294]:

Veracity It refers to the quality of the data, including its accuracy, reliability and trustworthiness. Big data sources can be inherently uncertain or inaccurate, and addressing data quality is a critical challenge.

Value Extracting value and actionable insights from big data is the ultimate goal. Analysing and interpreting Big data should lead to better decision-making, improved business strategies, as well as enhanced User experience (UX)⁸⁷.

Regarding the two latter dimensions, Debattista et al. [2015] argue that that Linked Data is the most suitable technology to increase the value of data over conventional formats, thus contributing towards the value challenge in Big Data. As for veracity, they describe a semantic pipeline with eight key metrics to address the veracity dimension. Building on this technological foundation, the integration of Linked Data and Big Data analytics takes centre stage.

Big data analytics can be employed on CH content to uncover insights and correlations that can be used in decision-making. Barrile and Bernardo [2022, p. 2708] highlight the transformative potential of using big data by investigating how analytical approach can enhance conservation strategies, aid resource allocation and optimise the management of CH resources. Pouloupoulos and Wallace [2022, pp. 188-189] emphasise that emerging technology trends, including big data, have a significant impact on related research areas such as CH. Big data primarily originates from sources such as social media, online gaming, data lakes⁸⁸, logs and frameworks that generate or use significant amounts of data. They stress that the incorporation of multi-faceted analytics in the CH domain is an area of active research, and present a data lake that provides essential user and data/knowledge management functionalities. However, they emphasise a crucial consideration - the need to bridge the theoretical foundations of disciplines such as cultural sociology with the technological advances of big data.

⁸⁶ GB and PB are units of digital information storage capacity. 1 GB is equal to 1,000,000,000 (10^9) bytes, 1 TB is equal to 1,000,000,000,000 (10^{12}) bytes, and 1 PB, is equal to 1,000,000,000,000,000 (10^{15}) bytes. If a standard high-definition movie is around 4-5 GB, then 1 PB could store tens of thousands of movies. In 2011, Gomes et al. [2011, p. 414] reported that the Internet Archive held 150,000 million contents of archived websites – crawled through the Wayback Machine – or approximately 5.5 PB. As of December 2021, it was about 57 PB of archived websites and a total used storage of 212 PB, see <https://archive.org/web/petabox.php>.

⁸⁷ In this context, UX is understood as an umbrella term encompassing both user and/or customer service, emphasising that the focus is on individuals who need or use a given service, regardless of their categorisation as users or customers.

⁸⁸ According to Nargesian et al. [2019, p. 1986], a data lake is a vast collection of datasets that has four characteristics. It can be stored in different storage systems, exhibit varying formats, may lack useful metadata or use differing metadata formats, and can change autonomously over time.

3.3.1.3 Artificial Intelligence

Artificial Intelligence (AI) has been coined for the first time by John McCarthy, an American computer scientist and cognitive scientist, during the 1956 Dartmouth Conference, which is often considered the birth of AI as an academic field [Andresen 2002, p. 84]. According to the [Oxford English Dictionary \[2023b\]](#), AI is described as follows:

‘The capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this. In later use also: software used to perform tasks or produce output previously thought to require human intelligence, esp. by using machine learning to extrapolate from large collections of data.’

While AI is not the central focus of my PhD thesis, I acknowledge its impact in several instances. As a rapidly developing technology, AI has the potential to significantly transform various aspects of society, including the way we describe, analyse, and disseminate CH resources. It is worth mentioning that I endeavour to engage in a broader discourse concerning the domain of AI. In this context, I use the acronyms AI to talk about the overarching domain or its ethics, and ML to discuss the specifics of methodologies and algorithmic approaches, while refraining from delving into the intricacies of Deep Learning, which is a distinct subdomain within ML.

AI and ML offer great potential for digitising, curating and analysing CH, leveraging the vast digital datasets from CHIs. Some of the examples include text recognition mechanisms using Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR), NLP and Named Entity Recognition (NER) for enriching unstructured text, as well as object detection methods for finding patterns within still and moving images [Neudecker 2022, Sporleder 2010]. Textual works can also be analysed, for instance for sentiment analysis [see Susnjak 2023], and generated using Large Language Models (LLM) – a variety of NLP, such as BERT or ChatGPT, which predicts the likelihood of a word given the previous words present in recorded texts. However, challenges such as data quality and biases in AI persist [Neudecker 2022].

In addition, there are still uncertainties regarding the licensing and reuse of CH datasets by ML algorithms⁸⁹. Neudecker [2022] emphasises the importance of well-curated digitised CH resources that are openly licensed, accompanied by relevant metadata, and accessible through APIs or download dumps in various formats. These curated resources have the potential to address the existing gap in this domain.

Building on the theme of enhancing CH through digital technologies, McGillivray et al. [2020] explore the synergies and challenges found at the intersection of DH and NLP. DH is aptly described as ‘*a nexus of fields within which scholars use computing technologies to investigate the kinds of questions that are traditional to the humanities [...] or who ask traditional kinds of humanities-oriented questions about computing technologies*’ [Fitzpatrick 2010]. This broad characterisation encapsulates the transformative potential of digital tools, including ML techniques, in enriching humanities research.

McGillivray et al. [2020] highlight the critical need for bridging the communication gap between DH and NLP to drive progress in both fields. They propose increased interdisciplinary collaboration, encouraging DH researchers to actively utilise NLP tools to refine their research methodologies. A primary challenge in this convergence is the application of NLP to the complex, historical, or noisy texts often encountered in DH research. They conclude by advocating for stronger cooperation between

⁸⁹ An interesting initiative in this area is the use of Responsible AI Licenses (RAIL), which empower developers to restrict the use of AI on the software they develop to prevent irresponsible and harmful applications: <https://www.licenses.ai/>

practitioners in these fields. This collaborative effort is vital for harnessing the full potential of ML in analysing and interpreting CH.

The use of ML scripts in the context of CH — and beyond — is inherently limited by their applicability, namely when dealing with historical photographs. In such cases, the use of algorithms that are mostly trained and grounded in contemporary image data becomes quite incongruous due to the dissimilarity in temporal contexts. This dilemma is exemplified by datasets such as Microsoft’s Common Object in Context (COCO)⁹⁰ [Lin et al. 2014], where the available data are predominantly contemporary photographic content, which is misaligned with the historical nuances inherent in most of the digitised CH images. Coleman [2020] corroborates that a sound approach would be for ML practitioners to collaborate with libraries as they can draw practical lessons from critical data studies and the thoughtful integration of AI into their collections, using guidelines from DH. She also advocates that as handing over datasets would be a disservice to library patrons and that ‘*Librarians need to master the instruments of AI and employ them both to learn more about their own resources—to see and analyze them in new ways—and to help shape applications of AI with the expertise and ethos of libraries.*’

Ethical concerns, particularly regarding social biases and racism, are prevalent in technologies like ImageNet, where facial recognition may yield AI statements with strong negative connotations [Neudecker 2022]. Addressing this, Gandon [2019a] suggest the production of AI services that are ‘*benevolent-by-design for the good of the Web and society*’. Furthermore, Floridi [2023] introduces the double-charge thesis, asserting that all technology design is a moral act, challenging the neutrality thesis. He emphasises that technologies are not neutral and can be influenced by a dynamic equilibrium of values, predisposing them towards morally good or evil directions.

As mentioned previously, the ML training datasets are often not enough representative to be properly leveraged in the CH sector [Strien et al. 2022]. Fine-tuning is now a topic though and new ground truth datasets have been created and tailored for the needs of CH, such as Viscount⁹¹, a large-scale Visual Question Answering (VQA) dataset — i.e a dataset containing open-ended questions about images which requires an understanding of vision, language and commonsense knowledge to answer [Goyal et al. 2017] — for CH in English and Italian [see Becattini et al. 2023].

Jaillant and Caputo [2022] argue that the governance of AI ought to be carried out in partnership with GLAM institutions. However, while this collaboration has been proposed as a promising way forward, it still requires further exploration and evaluation, particularly with regards to the specific challenges and opportunities that it presents. On the one hand, the involvement of GLAMs in AI governance could enhance the development of digital CH projects that promote social justice and equity. However, on the other hand, this collaboration raises several challenges, such as the need to address issues of privacy, data protection, and intellectual property rights, and to ensure that the values and perspectives of GLAM professionals are adequately represented in the development of AI algorithms and systems. Therefore, it is crucial to examine the specific challenges and opportunities of this collaboration and to develop appropriate frameworks and guidelines that enable effective and ethical governance of AI in the GLAM sector.

⁹⁰ Common Objects in Context: <https://cocodataset.org/>

⁹¹ Viscount – A Large Dataset for Visual Question Answering for Cultural Heritage: <https://github.com/misaelmongiovi/IDEHADataset>

One of these platforms that address these issues is AI4LAM, which is an international and participatory community focused on advancing the use of AI in, for and by libraries, archives, and museums⁹². The initiative was launched by the National Library of Norway and Stanford University Libraries in 2018 inspired by the success of the IIF community. Another agency is the AEOLIAN Network⁹³, AI for Cultural Organisations, which investigates the role that AI can play to make born-digital and digitised cultural records more accessible to users [Jaillant and Rees 2023, p. 582].

As an illustrative case, the LoC's exploration into ML technologies, as highlighted by Allen [2023], demonstrates a strategic commitment to enhancing the accessibility and utility of its diverse collections. This initiative reflects the LoC's acknowledgement of the transformative potential of ML, balanced with a cautious approach due to the necessity for accurate and responsible information stewardship. The LoC faces several challenges in applying ML, particularly the limitations of commercial AI systems in handling its varied materials and the requirement for substantial human intervention. This cautious exploration into ML is indicative of a broader trend in CHIs, where maintaining a balance between embracing technological advancements and preserving authenticity and integrity is crucial.

The specific experiments and projects undertaken by the LoC in the realm of ML are diverse and illustrative of the institution's comprehensive approach to innovation. For instance, image recognition systems have been tested for identifying and classifying visual elements in artefacts, a task that requires a nuanced understanding of historical and cultural contexts. In another initiative, speech-to-text technology was employed to transcribe spoken word collections, confronting challenges such as accent recognition and audio quality variation. Additionally, the LoC explored the potential of ML in enhancing search and discovery capabilities through projects like Newspaper Navigator⁹⁴, which aimed to identify and extract images from digitised newspaper pages.

These experiments not only highlight the potential of ML in transforming the way LoC manages and disseminates its collections but also reveal the complexities and limitations inherent in these technologies. As Allen [2023] notes, the ongoing research and experimentation in ML at the LoC are critical in revolutionising access and discovery in the cultural heritage sector. These efforts, while facing challenges, represent a diligent integration of advanced technologies, upholding principles of responsible custodianship and setting a precedent for similar institutions globally in the adoption and adaption of ML and AI in CHIs.

The integration of LLM and KG presents a groundbreaking opportunity, particularly within the realm of CHIs, where there is already considerable expertise. This is aptly demonstrated in the work of Pan et al. [2023], which elucidates the harmonisation between explicit knowledge and parametric knowledge, i.e. knowledge derived from patterns in data, as learned by models such as LLMs. The authors highlight three key areas for the advancement of KR and processing:

1. **Knowledge Extraction**, where LLMs improves the extraction of knowledge from diverse sources for applications such as information retrieval and KG construction;
2. **Knowledge Graph Construction**, which involves LLMs in tasks such as link prediction and triple extraction from data, albeit with challenges in precision and management of long tail entities;
3. **Training LLMs Using KGs**, where KGs provides structured knowledge for LLMs, helping to build retrieval-augmented models on the fly, enriching LLMs with world knowledge and increasing its adaptability.

⁹² Artificial Intelligence for Libraries, Archives & Museums: <https://sites.google.com/view/ai4lam>

⁹³ AEOLIAN Network: <https://www.aeolian-network.net/>

⁹⁴ Newspaper Navigator: <https://news-navigator.labs.loc.gov/>

In a report for the University of Leeds in the UK, [Pirgova-Morgan \[2023\]](#) explores the potential and practical implications of AI in libraries. The project, forming part of the university's ambitious vision for digital transformation, aims to understand how AI can be effectively integrated into library services. This research looks at both the use of general AI for long term strategic planning and specific AI applications for improving UX, process optimisation and enhancing the discoverability of collections. The methodology used in this study involves a multi-faceted approach including desk-based assessments, a university-wide survey and expert interviews. Specifically, the study highlights the following key findings:

AI for UX and Process Optimisation The integration of AI technologies offers substantial opportunities for improving user experiences in libraries. This includes optimising library processes, enhancing collections descriptions, and improving their discoverability.

Challenges and Opportunities of AI Application While AI presents exciting possibilities, its practical application in library settings faces challenges. These include evaluating specific AI technologies in the unique context of the University of Leeds, ensuring they align with the institution's needs and goals.

Perceptions of AI in Libraries The report reveals varying perceptions among librarians and users regarding AI. This includes views on how AI can contribute to resilience, awareness of climate change, and practices promoting equality, diversity, and inclusion.

Role of AI in Strategic Library Development General AI technologies are seen as instrumental in shaping long-term strategies for libraries, highlighting the need for ongoing adaptation and development in response to evolving AI capabilities.

Expert Perspectives on AI in Libraries Interviews with experts from around the world underscore the importance of understanding both general and specific applications of AI. These insights help in identifying priority areas where AI can significantly enhance library operations and services.

These insights from the University of Leeds report illustrate the complex impact of AI on library services, from enhancing user interaction to influencing strategic decision-making, while also emphasising the importance of adapting AI applications to specific institutional needs.

It must be also stated that AI lacks inherent intelligence and consciousness, and have been ultimately built by people. An important concern, namely with LLM, is the perceptual illusion of cognitive interaction, where the machine appears to be engaging in dialogue and reasoning, when in fact it is generating content through predictive algorithms [see [Ridge 2023](#)]. Furthermore, regarding the topic of data colonialism, poor people in underprivileged nations are often burdened with the responsibility of cleaning up the toxic repercussions of AI, shielding affluent individuals and prosperous countries from direct exposure to its harmful effects⁹⁵.

Concluding this segment, it is essential to perceive Machine Learning (ML) algorithms as uncertain '*socio-material configurations*', which can be seen as both powerful and inscrutable, demanding an axiomatic and problem-oriented approach in their understanding and application. [Jaton \[2017\]](#) elaborates on this by examining how these algorithms, while technologically complex, are firmly rooted in and shaped by the social, material, and human contexts in which they are developed. Beyond their computational complexity, these algorithms are deeply embedded in the process of constructing 'ground

⁹⁵ [Perrigo \[2023\]](#) investigated that Kenyan workers made less than USD 2 an hour to identify and filter out harmful content for ChatGPT.

truths'. These ground truths are not inherent or fixed; instead, they emerge from collaborative efforts that reflect the varied inputs of actors. This process underscores the algorithms as socio-material constructs, influenced by the characteristics and contexts of their creators. Understanding algorithms in this light highlights their deep integration with human actions and societal norms, offering a more nuanced view of their design and implementation [see [Jaton 2021, 2023](#)].

3.3.2 Scientific Movements and Guiding Principles

First, subsection 3.3.2.1 examines the movement towards more open and transparent forms of research. Open scholarship is a broad concept that encompasses practices such as open access publishing, open data, open source software, and open educational resources. The subsection explores the benefits and challenges of open scholarship, and how it can help to increase the accessibility and impact of research data.

Then, subsection 3.3.2.2 explores the growing trend of involving members of the public in scientific research. Citizen science and citizen humanities involve collaborations between scientists and non-expert individuals, with the aim of generating new knowledge or solving complex problems. The subsection examines the benefits and challenges of citizen science and citizen humanities, and how they can help to democratise research.

Subsubsection 3.3.2.3 examines the set of guiding principles designed to ensure that research outputs are Findable, Accessible, Interoperable, Reusable (FAIR). It explores the importance of each data principle for research integrity, reproducibility, and collaboration, and provides examples of how they can be implemented in practice.

Subsubsection 3.3.2.4 explores the importance of ethical and culturally sensitive data governance practices for indigenous communities that are materialised through Collective Benefit, Authority to Control, Responsibility, Ethics (CARE). These principles provide a framework for managing data in a way that is consistent with the values and cultural traditions of indigenous communities. This part explores as well the challenges and opportunities of implementing the CARE Principles for Indigenous Data Governance.

Finally, subsection 3.3.2.5 explores the concept of 'Collections as Data', a perspective that has emerged from the practical need and desire to improve decades of digital collecting practice. This approach re-conceptualises collections as ordered digital information that is inherently amenable to computational processing.

3.3.2.1 Towards Open Scholarship

According to the FOSTER⁹⁶, Open Science can be described as '*[...] the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods*' [[FOSTER 2019](#)].

In recent years, the principles of Open Science, that historically include Open methodology, Open source, Open data, Open access (OA), Open peer review, as well as open educational resources, have become increasingly important as they emphasise transparency, collaboration and accessibility in sci-

⁹⁶ FOSTER Plus (Fostering the practical implementation of Open Science in Horizon 2020 and beyond) was a 2-year EU-funded project initiated in 2017 with 11 partners across 6 countries. Its main goal was to promote a lasting shift in European researchers' behaviour towards Open Science becoming the norm.

entific research [Bezjak et al. 2019]. Open methodology refers to the sharing of research processes and methods, allowing other researchers to reproduce and build on existing work [see Vicente-Saez and Martinez-Fuentes 2018]. Open source software and tools enable researchers to collaborate, while open data practices promote the sharing of research data in ways that are accessible, discoverable and reusable by others⁹⁷. Open access seeks to remove financial and other barriers to accessing scientific knowledge, while open peer review provides greater transparency and accountability in the publication process. Finally, open educational resources encourage the sharing of teaching and learning materials, thereby facilitating the dissemination of knowledge and skills.

UNESCO [2019] conducted a preliminary study of the technical, financial and considerations related to the promotion of Open Science. This research underscored the necessity for a holistic approach to Open Science and stressed the significance of tackling international legal matters, as well as the existing challenges stemming from unequal access to justice, which can hinder global scientific collaboration. This study laid the groundwork for a recommendation on making ‘[...] multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation’ [UNESCO 2021, p. 7]. UNESCO identified five types of access related to Open Science: infrastructures, societal actors, as well as associated and diverse knowledge systems where dialogue is needed. This includes acknowledging the rights of indigenous peoples and local communities to govern and make decisions on the custodianship, ownership, and administration of data on traditional knowledge and on their lands and resources. Figure 3.10 provides a visual summary of this.

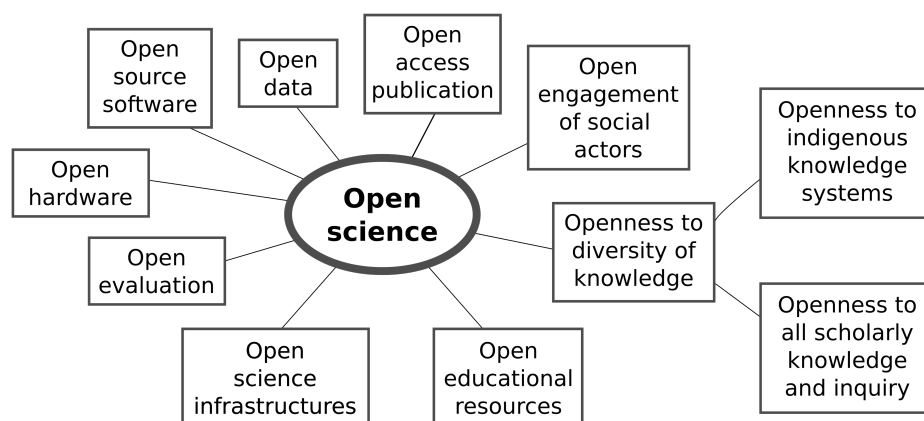


Figure 3.10: Open Science Elements, Redrawn Slide from Presentation of Ana Persic [Morrison 2021, citing Persic [2021]]

While Open Science offers numerous benefits, it also presents challenges and potential drawbacks that warrant careful consideration. One major concern is the risk of exacerbating inequities between researchers from well-resourced institutions and those from less privileged backgrounds. Open access publishing often entails significant costs in the form of article processing charges, which can disproportionately burden researchers without adequate funding support [Burchardt 2014]. Additionally, Open Science practices relying on open protocols may be vulnerable to misuse, such as automated bots exces-

⁹⁷ According to the Open Knowledge Foundation, a non-profit network established in 2004 in the U.K., which aims to promote the idea of open knowledge, sets out some principles around the concept of openness and defines it as follows: ‘Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)’. See <https://opendefinition.org/>

sively crawling open repositories or datasets. This can lead to overloading systems, unauthorised data extraction, or unintended uses of research outputs [see [Irish and Saba 2023](#), [Li et al. 2021](#)]. These risks underscore the importance of balancing openness with safeguards that ensure equitable participation and secure, sustainable access to research materials.

These challenges are particularly relevant in the context of DH, a field that harnesses the promise and impact of digital technologies and methodologies for the study and understanding of cultural phenomena. The adoption of Open Science principles has contributed to greater collaboration, transparency and accessibility in research practices in this field. Open data practices are particularly relevant, as they allow scholars to work with large and complex datasets, including digitised archives and social media data. Open educational resources can also be used to support the dissemination of CH literacy and skills, enabling wider audiences to engage with such resources. However, ensuring that such openness does not exacerbate inequities or introduce vulnerabilities requires thoughtful implementation.

In addition to the principles of Open Science, the concept of Open Scholarship has been introduced by [Tennant et al. \[2020\]](#) as a broader approach that encompasses the arts and humanities and goes beyond the research community to the wider public. Open Scholarship emphasises the importance of making research and scholarship accessible to a wider audience, including non-experts, educators and policy makers. It can be particularly relevant to the arts and humanities, as they often deal with complex cultural materials and narratives that have wider societal implications. By making their work openly accessible and engaging with non-experts, humanities researchers can contribute to public discourse, promote cultural understanding, and inform policy and decision-making. Open scholarship can also support greater collaboration and innovation within the Arts and Humanities by enabling researchers to work collaboratively across disciplines and with a wide range of constituents. For instance, open educational resources can be used to develop collaborative teaching and learning materials that draw on the expertise of scholars and practitioners from different disciplines, while open data practices can facilitate the sharing and reuse of CH materials.

Conversely, [Knöchelmann \[2019\]](#) advocates for the term Open Humanities as a dedicated discourse that would *‘bring together the variety of currently disconnected endeavours and proficiencies’* within the humanities. Notably, he argues that Open Humanities should adapt key Open Science elements to the Humanities’ unique context. In the case of preprints, the challenges in the humanities, such as limited discipline-specific preprint servers and linguistic diversity, require tailored solutions to encourage adoption. Open peer review in the humanities should accommodate the field’s subjectivity and diverse perspectives. Concerns about liberal copyright licenses revolve around potential misrepresentation and plagiarism, highlighting the importance of maintaining scholarly integrity regardless of the chosen license. Knochelmann’s proposal underscores the need for context-sensitive approaches to promote openness and collaboration while respecting humanities’ distinct characteristics.

Overall, the principles of Open Science provide a framework for promoting greater collaboration, transparency and accessibility in research practices. Yet, the challenges discussed underscore the need for careful adaptation to address inequities, cybersecurity concerns, and field-specific nuances. The concept of Open Scholarship, which stresses the importance of making research and scholarship accessible to wider audiences, can be instrumental in broadening the impact of research in both natural sciences and the humanities, as Open Science encourages greater collaboration and innovation across disciplines. Ultimately, this underscores the need for adaptation and positions all academic disciplines as essential contributors to societal understanding, cultural preservation and informed decision-making, while ensuring the sustainability and integrity of open practices.

3.3.2.2 Citizen Science, Citizen Humanities

Citizen Science and Citizen Humanities are approaches that involve the public in scientific and humanities research, respectively. They have become increasingly popular in recent years as a means of democratising research and engaging the public in academic initiatives.

Citizen Science, as articulated by Irwin [1995], embodies a fundamental commitment to sourcing knowledge beyond the confines of academia, with a deliberate focus on addressing the concerns and interests of the public. This perspective underscores the transformative power of Citizen Science, making it a catalyst for a more democratic approach to scientific endeavours.

Bonney [1996]’s perspective complements this vision by framing Citizen Science as a collaborative process where amateur enthusiasts actively participate in data collection for academic science, all the while gaining a deeper understanding of scientific principles and processes. In this light, Citizen Science emerges as an ideal vehicle for science education and a potent tool for enhancing public appreciation of scientific pursuits.

These viewpoints loosely align with the Oxford English Dictionary’s definition, characterising Citizen Science as ‘*scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions*’ and traces the earliest evidence of the term in 1989 [Oxford English Dictionary 2023c]. As such, Citizen Science stands as a harmonious intersection of public engagement, education, and scientific inquiry, amplifying the voice of non-academic contributors and democratising the scientific landscape.

The public can play a vital role in data collection, analysis, and interpretation. This involvement can take the form of participating in wildlife sightings tracking, monitoring water quality, or assessing air pollution. By participating in these activities, citizens become direct contributors to the generation of valuable scientific data. The transformative power of Citizen Science extends across a wide spectrum of scientific disciplines, emphasising its capacity to democratise and broaden the reach of scientific endeavours [see Pelacho et al. 2021].

Citizen Science is a form of co-creation, whether viewed as an innovation-oriented means of value creation [Jansma et al. 2022] or as a more radical form of empowerment, reinforces the democratisation of the research process [Metz et al. 2019]. It amplifies the voice of non-academic participants in scholarly pursuits, reflecting a profound shift in the way science is conducted. This collaborative model demonstrates how public engagement enriches the scientific landscape, allowing for the inclusion of different perspectives and a wider range of voices in the pursuit of knowledge. Furthermore, engaging in participatory practices also involves elements of ‘phronesis’⁹⁸ [see Mehlenbacher 2022], encompassing moral, affective, and care-oriented dimensions.

Trust is also a foundational and indispensable element in the landscape of participatory initiatives [see Dahlgren and Hansson 2020]. The success and sustainability of projects within Citizen Science heavily rely on establishing and maintaining trust among all stakeholders involved. This trust extends in multiple directions. First and foremost, participants must trust the project organisers and platforms that host these initiatives. They must have confidence that their contributions will be used responsibly and ethically, with respect to their time and effort. When contributors are assured that their involvement is valued and that the data they provide serves a meaningful purpose, their motivation to participate and

⁹⁸ Phronesis in philosophy is related to ‘*practical understanding; wisdom, prudence; sound judgement.*’ [Oxford English Dictionary 2023e]

provide accurate information is bolstered. Conversely, project organisers and institutions also need to instil trust in participants. Transparency in project objectives, methodology and data use is paramount. Clear and consistent communication is essential to address participants' concerns and provide feedback on the impact of their contributions. This two-way trust is the foundation of successful participatory projects and facilitates long-term engagement.

Citizen Humanities where members of the public can participate in activities such as crowdsourced transcription, tagging, and annotation of digital CH materials. These activities can help to uncover new knowledge and insights, as well as to make CH materials more accessible to a wider audience [Strasser and Haklay 2018]. It is important to note that within the context of these terms, Citizen Science is often regarded as the broader concept, encompassing both Citizen Science and Citizen Humanities. While the primary distinction between the two may, in some cases, appear to be terminological, in practice, they both exemplify the principles of open and inclusive research, akin to the concepts of Open Science and Open Humanities discussed in the preceding subsection. These approaches foster collaboration and engagement between researchers and the public, deepening the public's understanding and appreciation of the research process as a whole [Zourou and Ziku 2022]. This inclusive perspective, even if those participatory activities have been more widely used in natural sciences than in the humanities [Lowry and Stepenuck 2021], underscores Citizen Science as an umbrella term encompassing both scientific and humanities endeavours, each enriched by the active participation of the public.

While Citizen Science involves the public in research, they differ from crowdsourcing projects in several ways. Crowdsourcing typically involves the outsourcing of tasks to a large group of people, often through online platforms, with the aim of completing a specific task or project [Ridge 2017]. In contrast, Citizen Science focuses more on engagement and collaboration, with the goal of involving the public in the research process and generating new knowledge. That being said, there is also a convergence between Citizen Science with crowdsourcing projects. In many cases, Citizen Science initiatives may also involve crowdsourcing tasks, such as collecting or annotating data. Similarly, crowdsourcing projects may involve elements of Citizen Science, particularly when they aim to engage the public in scientific or CH research [Ridge et al. 2021b].

For instance, Haklay [2013, pp. 115-116] distinguish four categories or levels of participation in Citizen Science projects, each serving as a rung on the ladder of public engagement. The levels are as follows:

- Level 1. Crowdsourcing** In this level, citizens act as sensors and volunteered computing resources.
- Level 2. Distributed intelligence** Here, citizens serve as basic interpreters and volunteered thinkers.
- Level 3. Participatory science** At this stage, citizens actively participate in problem definition and data collection.
- Level 4. Extreme Citizen Science** In the highest level, citizens engage in collaborative science that encompasses problem definition, data collection, and analysis.

When applied in the context of Citizen Humanities, public participation takes diverse forms. This involvement can encompass activities such as the public's engagement in archaeological finds recording, as demonstrated by the Finnish Archaeological Finds Recording Linked Open Database (SuALT) project [Wessman et al. 2019]. Another illustration is the case of the Citizen-Led Urban Environmental Restoration project where '*young citizen scientists [in Jamaica and the United States] worked closely with museum scientists to restore two environmentally degraded urban sites*' [Commock and Newell 2023]. In terms of crowdsourcing of CH data or more broadly in the humanities, Owens [2013, p. 121]

discusses two primary challenges associated with integrating the concept. He highlights that both the terms ‘crowd’ and ‘sourcing’ pose certain problems. Successful crowdsourcing initiatives in libraries, archives, and museums, as he notes, typically do not rely on extensive crowds, and they are far from resembling traditional labour outsourcing endeavours. Furthermore, Owens emphasises that the central focus of such initiatives is not on amassing large crowds but rather on cultivating engagement and participation among individuals in the public who have a genuine interest.

As Citizen Humanities broadens its scope to encompass a wider public engagement in DH and CH research, successful collaborations between DH and relevant research infrastructures have shown promising results [Fišer and Wuttke 2018, Simpson et al. 2014]. Furthermore, the integration of scientific and curatorial knowledge plays a pivotal role in CH and humanities studies, uncovering previously unknown contextual information within original materials [France and Toth 2014]. As illustrated by institutions like the National Library of Estonia, the shift towards human-centred approaches and the development of DH services exemplify the expansion of Citizen Humanities [Andresoo 2018].

Incorporating user-generated or user-enhanced metadata still presents several challenges [Raemy 2021]. One major challenge is ensuring the quality and consistency of the data. Another challenge is managing the large volume of data generated by users. With increasing numbers of participants and contributions, it can become difficult to process and organise the data in a way that is useful for research and for the broader public. As Dahlgren and Hansson [2020] argue:

‘Participatory metadata production has been valued for its potential to reduce the workload of the heritage institutions and make possible speedier digitization. However, in practice, little of the resulting metadata has been reinserted into the institutions databases and used in-house by information specialists.’ [Dahlgren and Hansson 2020, p. 249]

This challenge is compounded by the fact that user-generated metadata may be unstructured, making it more difficult to analyse and interpret. To address these challenges, it can be helpful to have a robust data curation strategy, maintained by a team that can communicate with participants on a regular basis, as well as tools and technologies that enable efficient data processing and analysis. LOD can also be a useful approach for organising and linking diverse sources of information, enabling researchers to incorporate different perspectives and opinions into their analysis. This form of participation often involves micro-tasks, akin to ‘puzzle-like’ tasks, connecting users closely with the subjects they are describing [see Ridge 2023]. The dynamics of participatory projects are intriguing and multifaceted. As expressed by Dahlgren and Hansson [2020]:

‘[It] is the often tightly curated top-down design of crowdsourcing platforms where participation is wide in terms of numbers of participants but small in terms of what those participants are allowed to do. The second involves the preconception that the crowd per se, because of its sheer size, in some ways represents a diversity of perspectives and experiences, an idea which is often put forward as one of the benefits of participatory metadata production.’ [Dahlgren and Hansson 2020, p. 250]

Five recommendations outlined by Ridge et al. [2023], specifically geared towards the CH domain, can serve as valuable guidance for various participatory endeavours. These recommendations encompass:

Infrastructure ensuring platform sustainability by supporting existing tools alongside new developments;

Evidencing and Evaluation creating an evaluation toolkit to emphasise impact and wider benefits;

Skills and Competencies establishing a self-guided skills assessment tool and workshops for up-skilling;

CoP funding international knowledge-sharing events like informal meetups, low-cost conferences, and peer review panels, ensuring inclusivity beyond limited regional funding projects;

Incorporating Emergent Technologies and Methods providing support for educational resources and workshops to anticipate the opportunities and implications of emerging technologies.

These recommendations offer a versatile framework that can be applied to various participatory efforts, transcending the boundaries of specific domains and promoting a more inclusive and effective approach to public engagement in research and collaborative initiatives. By adhering to these measures, Citizen Science projects can better flourish, fostering a collaborative and proficient community of practitioners. Rather than creating new infrastructure, research projects should leverage and extend existing ones, such as Zooniverse⁹⁹, a generic Citizen Science portal and FromThePage¹⁰⁰, a transcribing platform.

In summary, both Citizen Science and Citizen Humanities represent participatory methods of inquiry. While they have gained popularity, critical discussions regarding their potential limitations, notably in terms of diversity, are integral to their ongoing development. These critical discussions encompass issues like the challenge of addressing notions of volunteer — thus unpaid — labour, lack of diversity, and countering the dominance of traditional, often exclusive scientific practices [see [Stengers and Muecke 2018](#)]. These conversations serve as essential drivers for the evolution of participatory approaches, prompting a reevaluation and refinement of their methodologies to ensure greater inclusivity and equity [[Lewenstein 2022](#)].

3.3.2.3 FAIR Data Principles

The FAIR data principles¹⁰¹ were developed to ensure that three types of entities – namely data, metadata, as well as infrastructures – are Findable, Accessible, Interoperable, and Reusable. The four key principles of FAIR and their underlying 15 sub-elements or facets are as follows [[Wilkinson et al. 2016](#)]:

F. Findable (Meta)data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

F1 (Meta)data are assigned a globally unique and PID

F2 Data are described with rich metadata (defined by R1)

F3 Metadata clearly and explicitly include the identifier of the data they describe

F4 (Meta)data are registered or indexed in a searchable resource

A. Accessible Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1 (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2 Metadata are accessible, even when the data are no longer available

⁹⁹ Zooniverse: <https://www.zooniverse.org/>

¹⁰⁰ FromThePage: <https://fromthepage.com/>

¹⁰¹ FAIR Principles: <https://www.go-fair.org/fair-principles/>

I. Interoperable The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1 (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2 (Meta)data use vocabularies that follow FAIR principles
- I3 (Meta)data include qualified references to other (meta)data

R. Reusable The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1 (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1 (Meta)data are released with a clear and accessible data usage license
 - R1.2 (Meta)data are associated with detailed provenance
 - R1.3 (Meta)data meet domain-relevant community standards

Originally introduced to improve data management and sharing in the life sciences, the FAIR principles have evolved into a widely adopted framework that transcends research disciplines. They have been adopted in a wide range of fields, including astronomy [O’Toole and Tocknell 2022], genomics [Corpas et al. 2018], environmental science [Crystal-Ornelas et al. 2022] and the humanities. In particular, FAIR principles have been applied to make historical archives, artworks or linguistic datasets more openly available for human users and search engines. Moreover, CHIs have embraced FAIR principles as guidelines and best practices, employing them in the deployment of repositories, virtual research environments or data platforms [Beretta 2021, Hahnel and Valen 2020].

Yet, the concept of FAIR data management practices in the humanities is not always straightforward as demonstrated by Gualandi et al. [2022] at the University of Bologna in Italy. The study, involving 19 researchers from the Department of Classical Philology and Italian Studies were interviewed to investigate the concept of ‘data’ in the humanities, particularly in relation to the FAIR principles. The study identified 13 types of research data based on participant input, such as publications, primary sources (manuscripts, artworks), digital representations of CH resources, but also websites, events, or standards. Thus, suggesting that within FAIR, ‘data’ should encompass all inputs and outputs of humanities research. The research also emphasised the importance of methodologies and collaboration in managing research effectively, emphasising the need for clarity and consensus in applying FAIR data principles within the field.

Indeed, implementing FAIR can be complex due to the variety of data types and existing practices. Such complexity requires structured methods. As [Jacobsen et al. 2020, p. 11] point out, FAIR is not a standard, it is a guide that needs implementations based on interpretations.

Similarly, Dunning et al. [2017, pp. 187-188] emphasise the multifaceted nature of the FAIR Data Principles and the need to view compliance as an aspirational objective. Their research reveals challenges in achieving full compliance, with specific difficulty in the Interoperable and Re-usable facets. They advocate for basic policy implementation in areas like PID, metadata, licensing, and protocols, alongside transparent documentation. Additionally, the authors stress the importance of using Hypertext Transfer Protocol Secure (HTTPS) – an extension of HTTP (see subsection 3.4.1) – to ensure secure data transmission and accessibility. Finally, the authors stress the importance of collaboration between (data) archivists and researchers.

GO FAIR [2016] – an initiative that aims to implement the FAIR data principles – outlined a seven-step FAIRification process, which includes essential stages in the transformation of data, as illustrated by Figure 3.11. These steps begin with the retrieval of non-FAIR data, followed by in-depth analysis to understand the content and structure of the data. The process then requires the creation of a semantic model that accurately defines the meaning and relationships of the data in a computational way, often involving the integration of existing ontologies and vocabularies. The fourth step involves linking data through the application of Semantic Web technologies, thereby improving interoperability and integration with disparate data sources. In addition, the assignment of a clear licence is highlighted as a separate step, emphasising its key role in enabling data reuse and open access. As a sixth step, metadata need to be assigned in order to support data discovery and access. Finally, the FAIRified data is deployed or published with its associated metadata and licence, ensuring that it can be accessed and discovered by search engines, even if authentication and authorisation requirements are in place. As a result, the new FAIR dataset can now be more conveniently aggregated with other data sources, making it more straightforward to raise research questions across multiple sources.

A further illustration of how FAIR can be deployed is the conceptualisation of the FAIR Digital Object (FDO), which includes a strong binding of various types of metadata [Schultes and Wittenburg 2019, pp. 7-9]. The members of the European Commission. Directorate General for Research and Innovation. [2018, p. 35] underline that the establishment of a FAIR-compliant ecosystem hinges on the FDO concept, an implementation framework to develop scalable cross-disciplinary capabilities. As illustrated in Figure 3.12¹⁰², data must be assigned PIDs and accompanied by detailed metadata to ensure reliable discoverability, usability and citation. They also argue to use widely accepted file formats and adhering to community-specific metadata standards as well as vocabularies to support interoperability and reuse.

Soiland-Reyes et al. [2022] highlight the potential of LOD to drive the adoption of FDO within research infrastructures. While this approach provides specifications and tools, the proliferation of standards and metadata vocabularies poses challenges to interoperability and implementation. To address these hurdles, the authors present the use of FAIR Signposting¹⁰³, which enables straightforward

¹⁰² This figure is also available in an enlarged format in Figure B.3.

¹⁰³ FAIR Signposting: <https://signposting.org/FAIR/>. Signposting focuses on expressing the topology of digital objects on the web with a view to increasing the FAIRness of scholarly objects in a distributed manner [Van de Sompel 2023].

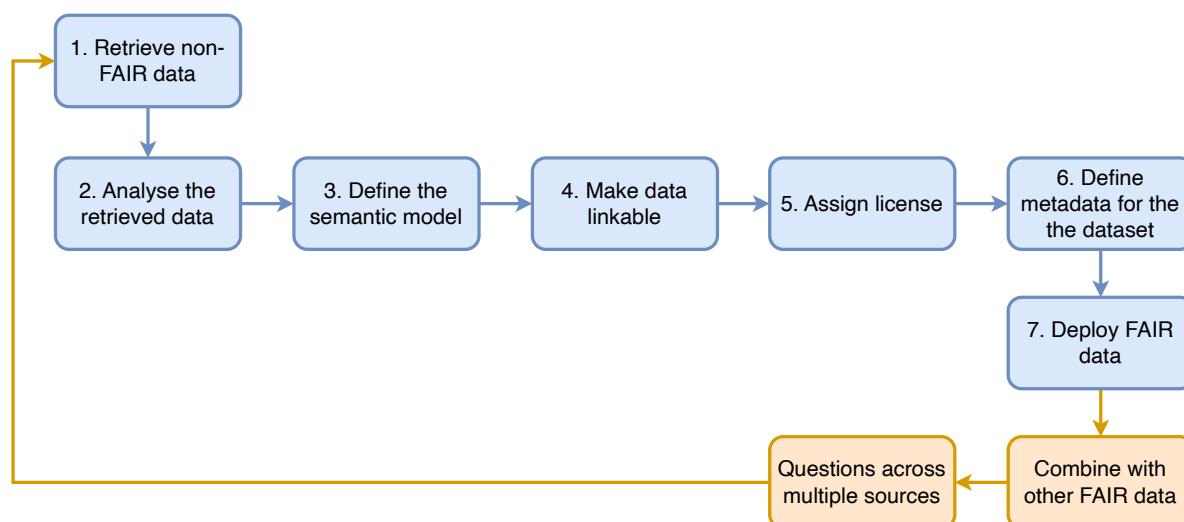


Figure 3.11: The FAIRification Process. Adapted from GO FAIR [2016]

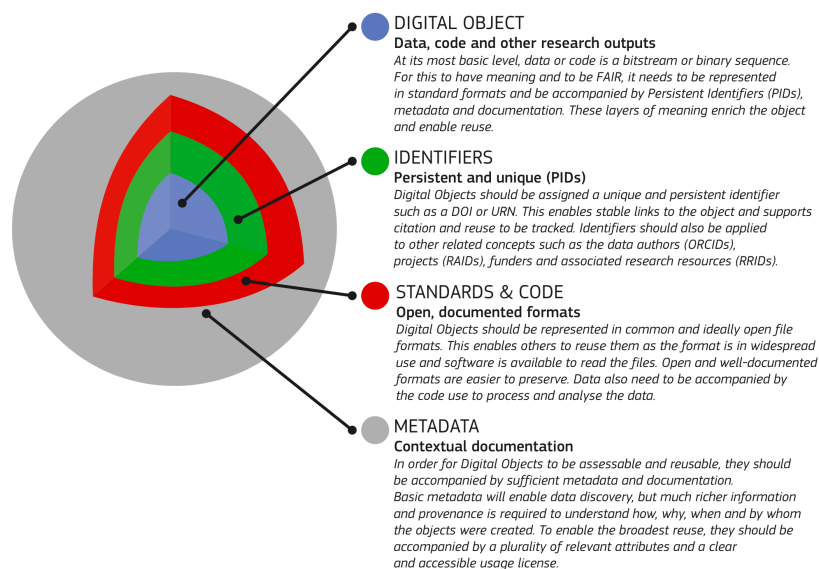


Figure 3.12: The FAIR Digital Object (FDO) Model [European Commission. Directorate General for Research and Innovation. 2018]

navigation to core FDO properties, without the need for complex content negotiation heuristics.

In summary, the FAIR data principles comprise four key principles or 15 facets that provide a comprehensive framework for data management and sharing. While processes are in place to facilitate their implementation, the path to FAIRness can be complex, with interoperability and compliance challenges. A key element is the thoughtful mapping of different metadata standards and the strategic incorporation of Linked Data technologies. The FDO approach is equally relevant to the CH sector, supporting the preservation, accessibility and sharing of CH data and resources. The sharing of code, accompanied by comprehensive documentation, also enhances such an ecosystem by facilitating the exchange of valuable technical knowledge and resources.

3.3.2.4 CARE Principles for Indigenous Data Governance

The Collective Benefit, Authority to Control, Responsibility, Ethics (CARE)¹⁰⁴ Principles were developed to protect Indigenous data sovereignty [Carroll et al. 2020] as complementary guidelines to FAIR. The principles are as follows:

- C. Collective Benefit** Data should be collected and used in a way that benefits the community as a whole, rather than just individuals or organisations.
- A. Authority to Control** Indigenous communities should have control over their own data, including how it is collected, stored, and used.
- R. Responsibility** Those who collect and use Indigenous data have a responsibility to ensure that it is used ethically and responsibly.
- E. Ethics** Indigenous data governance should be guided by ethical principles that reflect the values and beliefs of the community.

The concept of adhering to the CARE principles is vital for promoting equitable data practices. CARE are built upon existing data reuse principles like FAIR but also integrate the efforts of Indigenous-led networks focused on Indigenous data governance and research control. While FAIR emphasise data

¹⁰⁴ CARE Principles for Indigenous Data Governance: <https://www.gida-global.org/care>

accessibility, CARE go beyond that by considering actions aligned with the needs and intentions of individuals and communities connected to the data [Carroll et al. 2021]. By embedding CARE-informed data practices into project design, the ethical and responsible use of Indigenous data can be enabled to improve inclusive policies and services [Robinson et al. 2021].

3.3.2.5 Collections as Data

In the same vein of FAIR and CARE should be mentioned the ‘Vancouver Statement on Collections as Data’ that originated from a meeting of GLAM practitioners in Vancouver, Canada in April 2023 that builds on the ‘Santa Barbara Statement on Collections as Data’¹⁰⁵ done in 2017 [Padilla et al. 2017]. The statement highlights the growing global engagement with collections as data. It promotes the responsible computational use of collections to empower memory, knowledge and data practitioners. It emphasises ethical concerns, openness and participatory design, as well as the need for transparent documentation and sustainable infrastructure. The statement, comprising of ten recommendations, also recognises the potential impact of data consumption by AI, and the importance of considering climate impacts and exploitative labour [Padilla et al. 2023b]. More specifically, the following ten principles have been established for anyone with (meta)data stewardship responsibilities:

1. Collections as Data development aims to encourage computational use of digitised and born digital collections.
2. Collections as Data stewards are guided by ongoing ethical commitments.
3. Collections as Data stewards aim to lower barriers to use.
4. Collections as Data designed for everyone serve no one.
5. Shared documentation helps others find a path to doing the work.
6. Collections as Data should be made openly accessible by default, except in cases where ethical or legal obligations preclude it.
7. Collections as Data development values interoperability.
8. Collections as Data stewards work transparently in order to develop trustworthy, long-lived collections.
9. Data as well as the data that describe those data are considered in scope.
10. The development of collections as data is an ongoing process and does not necessarily conclude with a final version.

In a final report, Padilla et al. [2023a] underscore the transformative potential of the Collections as Data paradigm, particularly in the context of GLAMs. The principles and case studies highlighted in the report offer a roadmap for organisations to responsibly and ethically engage with their collections in the digital era. It is imperative to recognise that the journey towards fully realising its potential is ongoing and requires a commitment to continual evaluation and adaptation. This involves not only adhering to established principles but also being responsive to emerging technological trends, societal changes, and evolving ethical considerations. The role of AI in shaping the future of Collections as Data is particularly noteworthy. As AI continues to advance, it offers both opportunities and challenges in terms of enhancing access and insights into collections while also necessitating careful consideration of ethical implications, such as bias and privacy. Furthermore, the growing emphasis on climate impacts and sustainable practices in data stewardship is a crucial aspect that aligns with global efforts towards environmental responsibility.

¹⁰⁵ The Santa Barbara Statement on Collections as Data: <https://collectionsasdata.github.io/statement/>

Building on the discussion of the principles and initiatives surrounding Collections as Data, an in-depth analysis was carried out to assess the compliance of repositories, projects and platforms from six organisations with the checklist, namely the British Library¹⁰⁶, the National Library of Scotland¹⁰⁷, LoC¹⁰⁸, the Royal Danish Library¹⁰⁹, Meemoo¹¹⁰, and the Miguel de Cervantes Virtual Library¹¹¹ [Candela et al. 2023a, p. 13]. Although several institutions have opened access to their collections through APIs, such as IIF's capabilities, challenges remain in fully embracing the Collections as Data principles. Barriers include resource limitations and the balance between making collections widely available through simplified access and downloads. In addition, different items within the checklist may require different levels of maturity and prioritisation, often requiring collaborative efforts. Initial results show that the checklist is a valuable tool for identifying relevant issues for individual institutions, although prioritisation may vary according to context and user needs. Collaborative initiatives between institutions are underway to improve the practical implementation and user experience, particularly in the structuring of datasets [Candela et al. 2023a, pp. 20-21].

While there are still relatively few examples of institutions that have fully adopted the Collections as Data principles, several case studies – such as at the Royal Library of Belgium, which is materialised through DATA-KBR-BE¹¹² [see Chambers et al. 2021] – and initiatives offer valuable insights. For instance, Candela et al. [2023a, p. 7] outline a checklist tailored to GLAM institutions to publish Collections as Data¹¹³. They devised 11 criteria, including the provision of clear licensing for dataset reuse without restrictions, citation guidelines, comprehensive documentation, the use of public platforms, sharing examples of dataset use, structuring the data, providing machine-readable metadata, participation in collaborative edition platforms, offering API access to the repository, developing a dedicated portal page, and defining clear terms of use. These recommendations serve as a structured framework to enhance accessibility, usability, and interoperability, fostering engagement with cultural and historical collections.

A further notable advancement that has been done in the area of publishing Collections as Data is the contribution of Alkemade et al. [2023]. They have outlined a series of recommendations for developing datasheets or modular templates designed for CH datasets. This initiative holds significant importance for GLAMs, facilitating the structured organisation of their data, notably for seamless integration with ML tools where they propose to provide a description of how content have been influenced by digitisation. Their work highlights the need for documentation, focusing on tailored metrics, biases, and system integration. The proposed datasheets aim to detail the creation, selection, and digitisation processes, enhancing transparency and addressing the distinctive challenges of digital CH data. Emphasising a narrative approach to articulating biases, the author acknowledges the complex historical context and ethical implications.

3.4 Open Web Platform and Linked Data

The web, created at the European Organization for Nuclear Research (CERN) in 1989 by Tim Berners-Lee¹¹⁴, has enabled scholars and CH practitioners to access and analyse vast amounts of data in new ways, thereby opening the door to the creation of federated datasets and KGs. At the heart of this

¹⁰⁶ British Library's Research Repository: <https://bl.iro.bl.uk/>

¹⁰⁷ Data Foundry – Data collections from the National Library of Scotland: <https://data.nls.uk/>

¹⁰⁸ LoC Labs Data Sandbox: <https://data.labs.loc.gov/>

¹⁰⁹ Royal Danish Library's Mediestream: <https://www2.statsbiblioteket.dk/mediestream/avis>

¹¹⁰ Meemoo's Art in Flanders: <https://artinflanders.be/>

¹¹¹ BVMC Labs: <https://data.cervantesvirtual.com/>

¹¹² DATA-KBR-BE: <https://www.kbr.be/en/projects/data-kbr-be/>

¹¹³ A Checklist to Publish Collections as Data in GLAM Institutions: <https://glamlabs.io/checklist/>

¹¹⁴ The birth of the Web: <https://home.cern/science/computing/birth-web>

transformation are two pivotal concepts: the Open Web Platform and Linked Data. The Open Web Platform refers to a set of technologies and standards that allow for the creation and sharing of content on the web. Linked Data, on the other hand, refers to a set of principles and technologies that enable the publication and interlinking of data on the web, creating a web of data that can be easily navigated and used by humans and machines alike.

Recognising the Web as an environment that supports a wide range of applications beyond traditional browser-based interactions is becoming increasingly important. Platforms such as social networks like Facebook, Twitter, Instagram, and Mastodon, streaming services such as Netflix and Disney+, as well as cloud-based applications, all leverage web technologies even when not accessed via a traditional web browser. These platforms are integral to the web ecosystem, highlighting the web's role as a foundational platform for diverse digital interactions and data exchanges.

In the context of the Internet, it is important to note that much of what we know today about it is the result of developments by many individuals and organisations. However, a significant milestone was the development of the TCP/IP protocol by Vinton Cerf and Robert E. Kahn in the 1970s [see [Cerf and Kahn 1974](#)]. This protocol became the standard networking protocol on the ARPANET in 1983, marking the beginning of the modern Internet [[Leiner et al. 1997](#)]. Understanding the differentiation between the Internet and the web is crucial. The former is a global network of interconnected computers that communicate using Internet protocols, forming the infrastructure that enables online communication. The web, or World Wide Web, is a service built on top of the Internet, leveraging Hypertext Transfer Protocol (HTTP) to transmit data. While the Internet provides the underlying connectivity, the web offers a way to access and share information through websites and links. This differentiation is vital in comprehending how the web, as a part of the Internet, has evolved into a versatile and ubiquitous platform supporting a wide array of applications.

This section, divided into five subsections, explores some of the key concepts underlying the Open Web Platform and Linked Data, and their applications in the CH field.

First, subsection 3.4.1 examines the foundational principles and technologies that underpin the Open Web Platform. This includes an overview of principles, protocols such as HTTP, and the use of URIs to identify resources on the web. This part also explores the different types of web architectures such as the client-server model or the concept of web services, which allow for the exchange of data and functionality across different applications and systems.

Subsection 3.4.2 explores the vision of the web as a giant, interconnected database of structured data that can be queried and manipulated by machines. The subsection examines the technologies and standards that make up the Semantic Web, including RDF, RDFS, OWL, and SPARQL.

Subsection 3.4.3 examines the set of principles designed to promote the publication and interlinking of data on the web. The subsection explores the four principles of Linked Data - using URIs to identify resources, using HTTP to retrieve resources, providing machine-readable data, and linking data to other data.

Subsection 3.4.4 examines the set of criteria for publishing data on the web in a way that makes it easily discoverable, accessible, and usable. The subsection describes the Five-Star and Seven-Star deployment schemes, which include criteria such as providing data in a structured format, using open standards, and providing a machine-readable license.

Finally, subsection 3.4.5 explores the specific application of LOD in the CH domain. The subsection provides examples of how CHIs such as museums, libraries, and archives are using Linked Data to make their datasets more accessible and discoverable on the web.

Overall, this section provides a comprehensive overview of the key concepts and technologies underlying the web as an open and linked platform, and their applications in the CH field and more broadly for any scientific endeavours as that the web started with ‘*the philosophy that much academic information should be freely available to anyone*’ [Nelson and Van de Sompel 2022, citing Berners-Lee [1991]]. Through exploring these concepts, we can gain a deeper understanding of how the web is evolving into a more open, interconnected, and data-driven platform, and how this evolution is transforming the way we access, use, and share information.

3.4.1 Web Architecture

The web architecture has played a very important role in the development of scholarly research and CH practices, enabling new forms of collaboration, data sharing and interdisciplinary research. By providing a standardised and interoperable framework based on open standards for sharing and accessing data [Berners-Lee 2010], it has facilitated the open exchange of information, even if citation, i.e. ‘*the ability to unambiguously refer to named objects*’ has always been an issue, particularly for scholarly outputs [Lagoze et al. 2012, p. 2223].

Web architecture is a conceptual framework led by the W3C that underpins and sustain the World Wide Web [Jacobs and Walsh 2004], created to be ‘*a pool of human knowledge*’ [Berners-Lee et al. 1994]. It encompasses the architectural bases of identification, interaction, and format – also referred to as representation where HTTP provides the technical mechanisms for transmitting and accessing information.

The web architecture is based on a set of identifiers, such as URIs, which are used to uniquely identify resources on the web. These identifiers play a crucial role in enabling users to find, access, and share information on the web, and they help to ensure that web-based systems are both user-friendly and interoperable. Here, it is valuable to distinguish between three key concepts: URI, URL, and URN, as a URI can be further classified as a locator, a name, or both [Berners-Lee et al. 2005, p. 7] – as shown in Figure 3.13.

URI It is the overarching term encompassing both URLs and URNs. It serves as a generic identifier for any resource on the web. URIs can be used to uniquely identify resources, regardless of the specific naming or addressing scheme employed.

URL It is a subset of URIs and refers to web addresses that specify not only the resource’s identity but also its location or how to access it. URLs often include the protocol (such as HTTP) and the resource’s specific location (e.g., a domain and path).

URN It is another subset of URIs that emphasise the resource’s identity rather than its location or how to access it. URNs are designed to be persistent and unique, making them suitable for resources that are intended to be recognised and referenced over time. While URLs may change as resources move or evolve, URNs should remain constant.

Interaction between web agents, i.e. a person or a piece of software acting in the information space on behalf of a person, entity or process, over a network involves URIs, messages and data. Web protocols, such as HTTP, are message-based. Messages can contain data, resource metadata, message

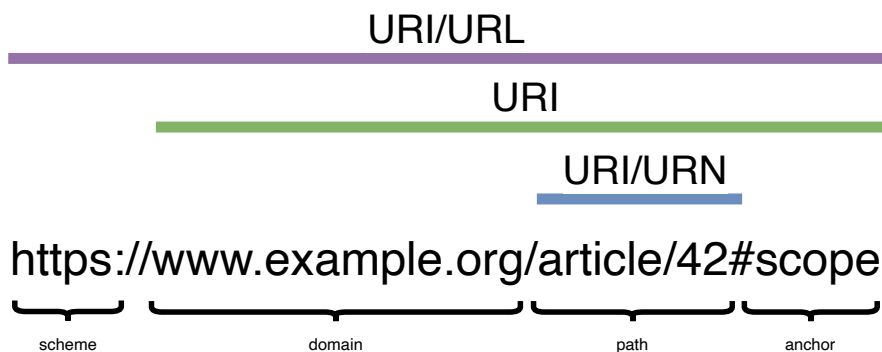


Figure 3.13: Overlap and Difference between URI, URL, and URN

data, and even metadata about the metadata of the message, typically for integrity checking [Jacobs and Walsh 2004].

The Web Architecture allows for multiple Representations of a Resource. In this context, a data format specification becomes pivotal, encapsulating an agreement on how to correctly interpret the representation of data, as articulated by [Jacobs and Walsh 2004]:

‘A data format specification embodies an agreement on the correct interpretation of representation data. The first data format used on the Web was Hypertext Markup Language (HTML). Since then, data formats have grown in number. Web architecture does not constrain which data formats content providers can use. This flexibility is important because there is constant evolution in applications, resulting in new data formats and refinements of existing formats. Although Web architecture allows for the deployment of new data formats, the creation and deployment of new formats (and agents able to handle them) is expensive. Thus, before inventing a new data format (or “meta” format such as XML), designers should carefully consider re-using one that is already available.’

Access can also be mediated by content negotiation, which is a mechanism employed in web communication to determine the most appropriate representation of a resource to be sent to a client based on the client’s preferences and the available representations [Lagoze et al. 2012, pp. 2223-2224].

At its core, web architecture is based on a set of architectural principles that guide the design and development of web-based systems and applications. These principles include concepts such as orthogonality, extensibility, error handling, and protocol-based interoperability. Orthogonality allows the evolution of identification, interaction, and representation independently. Extensibility is key, enabling technology to adapt without compromising interoperability. Error handling addresses diverse errors, from predictable to unpredictable, ensuring seamless correction. Finally, the web’s protocol-based interoperability fosters communication across varied contexts, outlasting entities and facilitating the longevity of shared technology [Jacobs and Walsh 2004]. Overall, these principles help to ensure that the web remains robust, reliable, and flexible.

Web architectures can be categorised into several types, each offering a specific approach to designing and structuring web-based systems. Here, I will focus on the following three types of architectures – shown in Figure 3.14: the client-server model, the three-tier model, and service-oriented architecture (SOA).

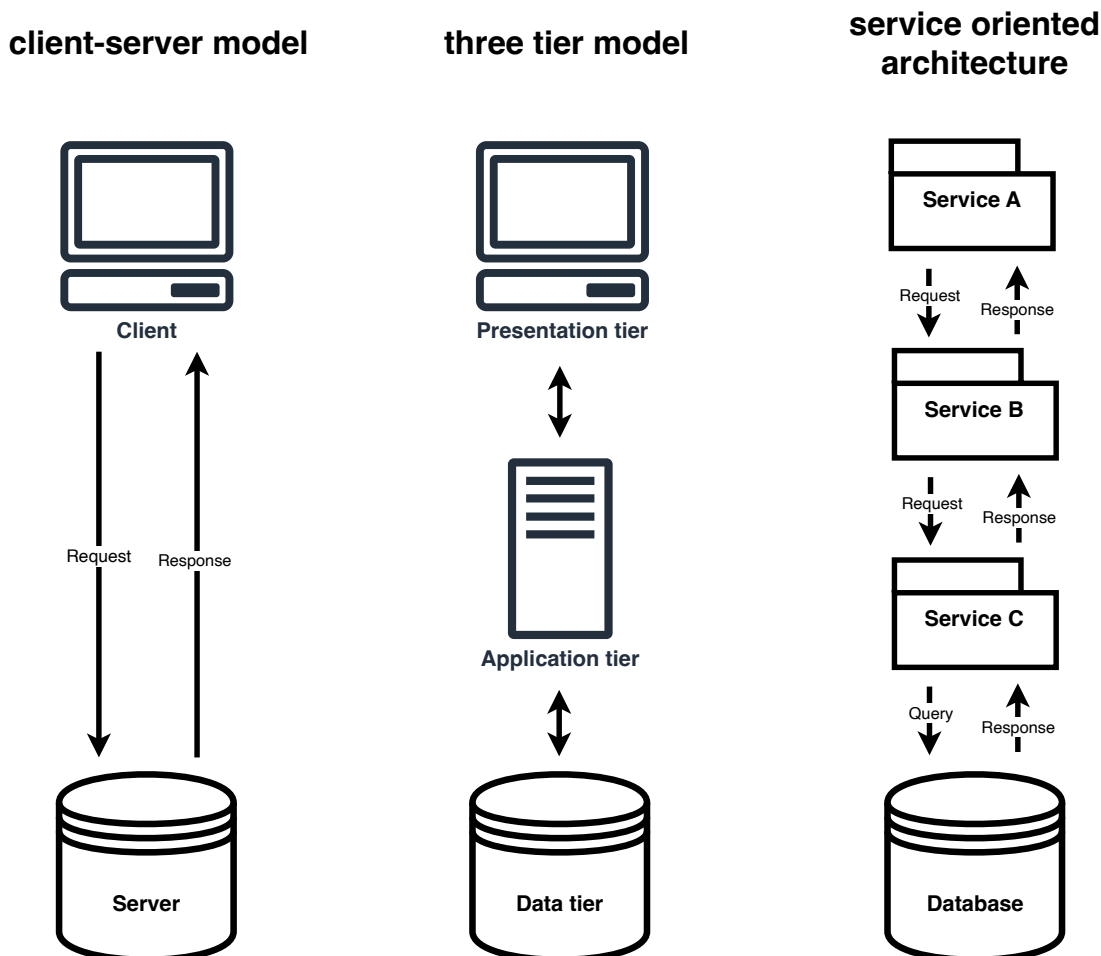


Figure 3.14: Types of Web Architectures: Client-server Model, Three Tier Model, SOA

The client-server model partitions the responsibilities between two key components: the client, which represents the user interface or user-facing part of the system, and the server, which is responsible for storing and serving data. In this model, clients and servers communicate to perform various functions, such as requesting and delivering information [Oluwatosin 2014, pp. 67-68].

The three-tier model is another significant web architecture that introduces an additional layer between the client and server, resulting in a three-part structure. This architecture is designed to further segregate and manage the system's components [Wijegunaratne and Fernandez 1998, pp. 41-42]. The three tiers typically consist of the presentation tier (the user interface), the application tier (responsible for logic and processing), and the data tier (where data storage and retrieval occur).

SOA is a web architecture that emphasises the creation and utilisation of services as the central building blocks of a system. Services in this context are self-contained, modular units of functionality that can be accessed and used independently by various components of a web application. These services are designed to be loosely coupled, meaning they can interact with other services without a deep dependency on one another. Overall, '*SOA is a paradigm for organizing and packaging units of functionality as distinct services, making them available across a network to be invoked via defined interfaces, and combining them into solutions to business problems.*' [Laskey and Laskey 2009, p. 101].

SOA can encompass various communication protocols, such as Representational state transfer (REST), which is a prominent architectural style for designing networked applications [Fielding

2000], primarily leveraging HTTP. RESTful services, i.e. applications that complies with the REST constraints, are designed to work with existing capabilities rather than creating new standards, frameworks and technologies [Battle and Benson 2008, p. 62]. These services are built around a set of constraints, including statelessness, a uniform interface, resource-based identification, and the use of standard request methods such as GET, POST, PUT, and DELETE [Tilkov 2017].

The following are all the specified request methods enabling clients to perform a wide range of operations on resources¹¹⁵ [Fielding et al. 2022, p. 72]:

GET Transfer a current representation of the target resource.

HEAD Same as GET, but only transfer the status line and header section.

POST Perform resource-specific processing on the request payload.

PUT Replace all current representations of the target resource with the request payload.

DELETE Remove all current representations of the target resource.

CONNECT Establish a tunnel to the server identified by the target resource.

OPTIONS Describe the communication options for the target resource.

TRACE Perform a message loop-back test along the path to the target resource.

RESTful services, with their emphasis on using standardised HTTP methods and resource-based identification, offer a versatile means of designing web services and APIs. Their simplicity and compatibility with the web’s core protocols make them a practical choice for implementing various web-based applications. In the context of the Semantic Web, RESTful services can serve as a crucial component for accessing and exchanging graph data [see Lee and Kim 2011].

3.4.2 The Semantic Web

The Semantic Web is ‘*[a] form of Web content that is meaningful to computers*’ [Berners-Lee et al. 2001, p. 35]. It was already in Berners-Lee [1999]’s vision and prediction that the web, in its next phase, could be understood by machines, i.e. shifting from a traditional web of documents to a web of data. Bauer and Kaltenböck [2012, p. 25] articulates that ‘*[t]he basic idea of a semantic web is to provide cost-efficient ways to publish information in distributed environments. To reduce costs when it comes to transferring information among systems, standards play the most crucial role*’.

At the heart of the Semantic Web lies the foundation of RDF. The original RDF specification, known as the RDF Model and Syntax, serves as the underlying mechanism that establishes the basic framework of RDF. This framework provides the cornerstone to facilitate the exchange of data among automated processes [Lassila and Swick 1999]. A fundamental component within RDF is the RDF triple as shown in Equation 3.1, comprising three essential elements: the subject (*s*), the predicate (*p*), and the object (*o*). In an RDF triple, the subject is the resource or entity about which a statement is made, the predicate is the relationship or property describing that statement, and the object is the value or resource associated with the statement.

$$s \xrightarrow{p} o \quad (3.1)$$

¹¹⁵ All general-purpose servers must support the methods GET and HEAD. All other methods are optional.

RDF statements are reminiscent of the semiotic triangle of [Ogden and Richards \[1930, p. 11\]](#) — as illustrated in Figure 3.15 — where the referent is tantamount to the predicate of a triple. This analogy emphasises the intrinsic relationship between communication, representation and knowledge organisation. It highlights how both language and structured data rely on the establishment of connections and relationships to effectively convey meaning.

Figure 3.16 is an RDF graph about myself and where I was born leveraging mostly Schema.org¹¹⁶, a collaborative project and Linked Data vocabulary used to create structured data markup on websites. This graph consists of vertices and edges, where vertices can be either URIs or literal values, and the edges represent relationships between them. In plain language, the graph asserts that there is a person represented by the URL ‘<https://www.example.org/julien-a-raemy>’, who has a given name ‘Julien Antoine’ and a family name ‘Raemy’. The person’s birthplace is specified as an URL from Wikidata, which is of type `schema:Place`. Additionally, there’s a statement indicating that the birthplace is named ‘Fribourg’.

In the subject-predicate-object syntax of RDF, the subject can be either a URI or a blank node¹¹⁷. The predicate is an URI, like `schema:givenName`, and its aim is to establish connections between subjects and objects, describing the nature of the relationship. The object is either an URI, a blank node or a literal, such as ‘Julien Antoine’ or ‘Raemy’. Objects can also act as subjects if they are identifiable, allowing for the expansion and interconnection of RDF graphs.

The original specification proved too broad, leading to confusion and a subsequent effort yielded an updated specification and new documents such as RDF/XML [\[Beckett 2004\]](#), which express an RDF graph as XML, a syntax specification recommendation in 2004 and later revised in 2014 as part of the RDF 1.1 document set [\[Gandon and Schreiber 2014\]](#), which also introduces the notion of an RDF dataset that can represent multiple graphs [\[Cyganiak et al. 2014\]](#). Code Snippet 3.4 is the RDF/XML serialisation of the earlier graph.

¹¹⁶ Schema.org: <https://schema.org/>

¹¹⁷ ‘A blank node is like a URI without a name and is also sometimes called an anonymous resource. They can be useful when you need to link to a collection of items but don’t want to bother making up a URI for it.’ [\[Wood et al. 2014, p. 35\]](#)

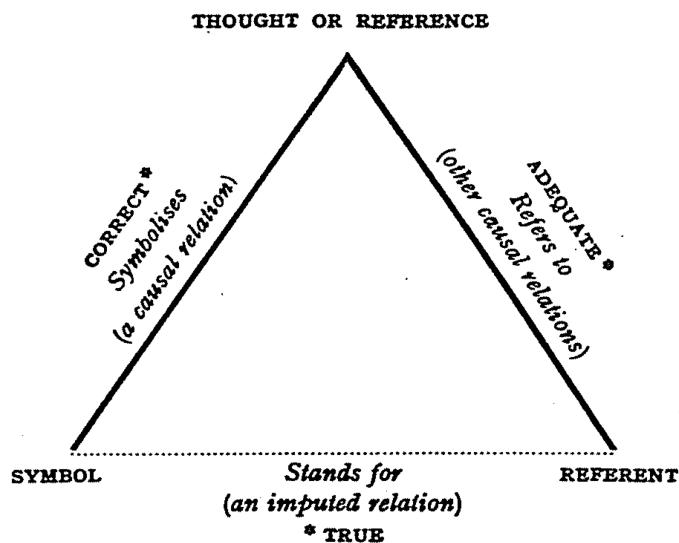


Figure 3.15: The Semiotic Triangle by [\[Ogden and Richards 1930\]](#).

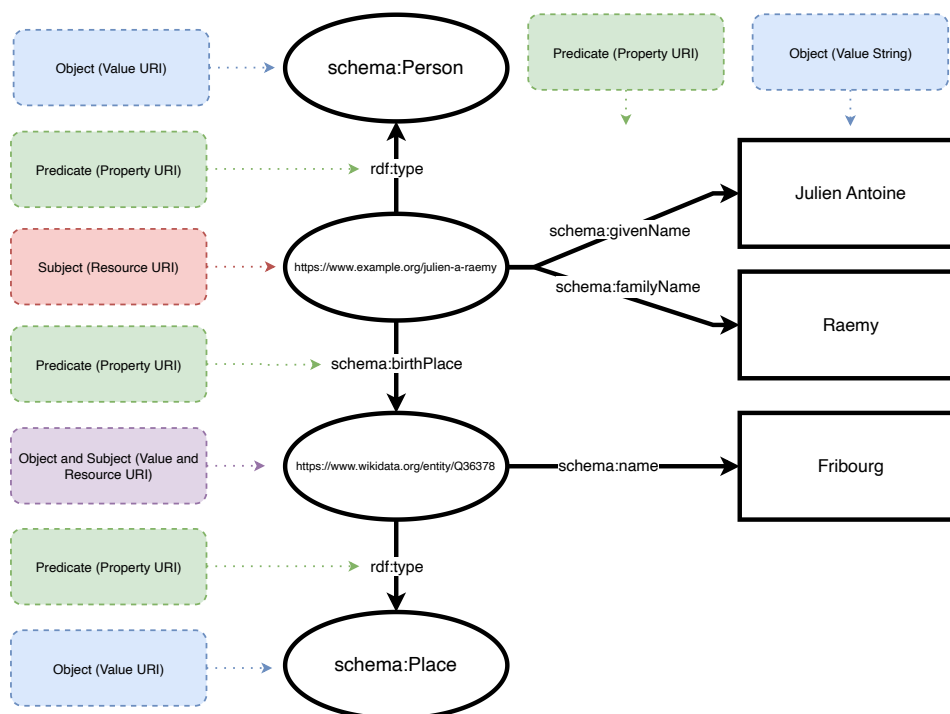


Figure 3.16: Example of an RDF Graph

Code Snippet 3.4: RDF/XML Serialisation of the RDF Graph

```

1 <?xml version="1.0" encoding="utf-8" ?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:schema="http://schema.org/"
4   <schema:Person rdf:about="https://www.example.org/julien-a-raemy">
5     <schema:givenName>Julien Antoine</schema:givenName>
6     <schema:familyName>Raemy</schema:familyName>
7     <schema:birthPlace>
8       <schema:Place rdf:about="https://www.wikidata.org/entity/Q36378">
9         <schema:name>Fribourg</schema:name>
10      </schema:Place>
11    </schema:birthPlace>
12  </schema:Person>
13 </rdf:RDF>

```

Idehen [2017] highlights a significant concern regarding the earlier representations of the Semantic Web and how it is portrayed. These portrayals often place undue emphasis on the pivotal role of XML as an ostensibly obligatory component in Semantic Web development. To him, this historical perspective, particularly prominent around the year 2000, erroneously positioned XML as a superior alternative to HTML for constructing the Semantic Web.

As illustrated by Figure 3.17, Idehen [2017]’s revision embodies a Semantic Web layer cake that encompasses several technical or conceptual components.

Smart Applications and Services These systems are constructed declaratively, as opposed to using an imperative approach, with a flexible integration of data models, interaction, and visualisation.

Trust It is established through verifiable claims regarding identity, the source of content, and related issues.

Proof It provides a basis for building trust, such as leveraging authentication tiers.

Transmission Security It pertains to safeguarding data during its transit over networks. This protection is achieved by implementing established protocols, such as Transport Layer Security (TLS),

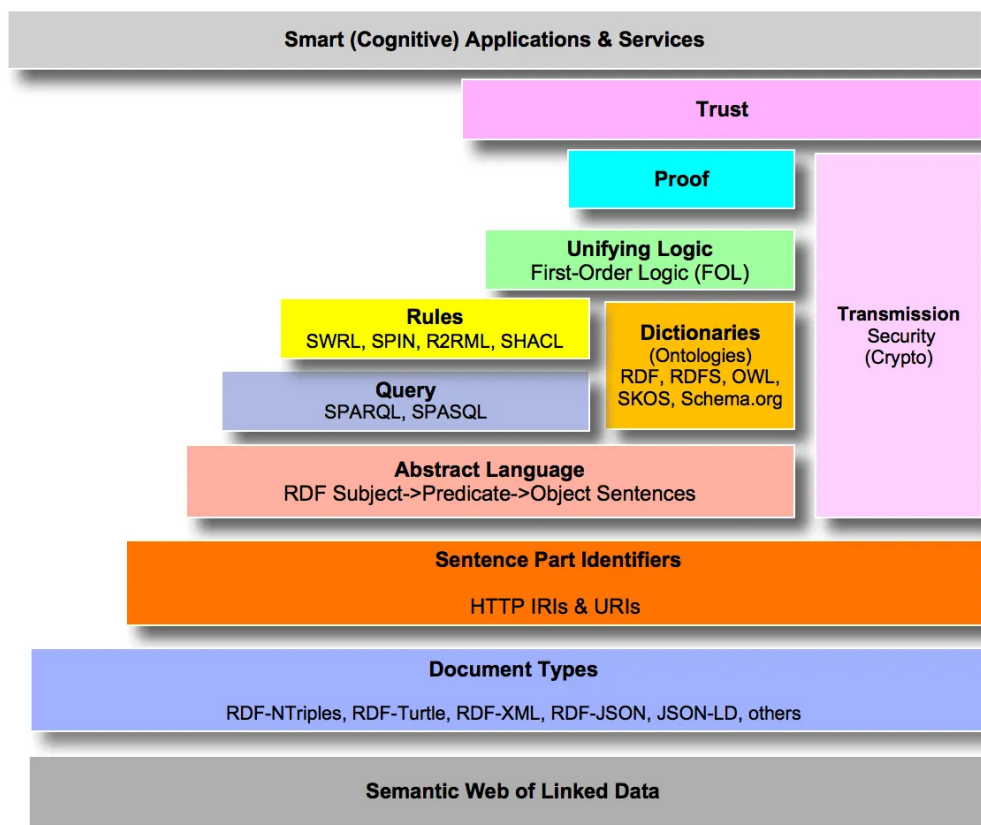


Figure 3.17: Tweaked Semantic Web Technology Layer Cake by Idehen [2017]

which includes inherent cryptographic support for ensuring the confidentiality and integrity of data as it travels across communication channels.

Unifying Logic In this component, First-order logic (FOL)¹¹⁸ assumes a central role by providing the fundamental schema for modelling and comprehending data. These propositions serve as the core building blocks for problem-solving and decision-making, offering a universal and abstract structure that can be applied across various domains and applications¹¹⁹.

Rules They serve as the foundation for conducting reasoning and inference, such as using SHACL, which is used to specify integrity constraints for data entry when constructing structured data with RDF. At this level, mapping languages such as RDB to RDF Mapping Language (R2RML) and RDF Mapping Language (RML) have a part to play. R2RML is a powerful language for expressing customised mappings from relational databases to RDF datasets [see Das et al. 2012]. On the other hand, RML extends these capabilities beyond relational databases, allowing for the transformation of various types of data sources, including Comma-separated values (CSV), XML, and JSON files, into RDF [see Dimou and Vander Sande 2022]. Both languages are instrumental in bridging the gap between non-RDF data sources and the Semantic Web.

¹¹⁸ FOL, also known as first-order predicate logic or first-order predicate calculus, is a formal system of symbolic logic used in mathematics, philosophy, and computer science. It is a logical framework for expressing and reasoning about statements involving objects and their properties and relationships. In FOL, statements are represented using variables, constants, functions, and predicates. It allows for the quantification of variables and the formulation of statements such as \forall (for all) and \exists (there exists), which enable the expression of universal and existential quantification. As such, FOL can express facts concerning some or all of the objects in the universe. Its epistemological commitment, i.e. what an agent believes about facts, is concentrated of what is true, false, or unknown [see Russell and Norvig 2010, pp. 285 ff.]

¹¹⁹ It must be noted that DL, a subset of FOL – briefly introduced in subsection 3.2.4.4, has a more restricted syntax and semantics tailored for ontology modelling.

Query This can be achieved by employing SPARQL for retrieving and manipulating data on structured RDF statements.

Dictionaries Collections of formal definitions (also referred to as vocabularies or ontologies) that describe entity and entity relationship types. They provide a structured framework for defining and organising concepts, properties, and relationships, which aids in modelling knowledge in a systematic and machine-readable manner. Some of the commonly used dictionary languages include RDFS, which is a simple vocabulary language used to define basic schema information for RDF, SKOS designed for representing and organising some KOSs, particularly in the context of thesauri and taxonomies, as well as OWL.

Abstract Language It is done by leveraging the RDF syntax ($s \xrightarrow{p} o$) as a basis.

Sentence Part Identifiers To identify resources on the web, Internationalized Resource Identifiers (IRIs)¹²⁰ or URIs can be used.

Document Types Different serialisations of RDF exist within the Semantic Web, such as RDF/XML, Turtle or JavaScript Object Notation for Linked Data (JSON-LD)¹²¹.

Semantic Web of Linked Data The final component that holds the rest together; originating from Tim Berners-Lee's vision (see subsection 3.4.3).

Following on from the components outlined in Figure 3.17, I will look in more detail at further RDF features, serialisations, and RDF-based standards for representing, querying or validating graphs. In doing so, I will touch on some considerations related to the inference and reasoning of RDF graphs.

Code Snippet 3.5 is a Turtle serialisation of Figure 3.16. Turtle, a W3C standard, is a notation is a way to express this data in a structured and machine-readable format [Wood et al. 2014, p. 44]. It is a common syntax used for representing RDF data. It allows people to create statements in a more friendly manner than in an RDF/XML serialisation [Beckett et al. 2014]. Here are some of the most important features:

@prefix It is used to declare namespace prefixes at the beginning of a Turtle document. These prefixes are used to shorten URIs, making the RDF data more readable and helps to avoid repeating long URIs.

Resource Notation In Turtle, resources (URIs) are typically enclosed in angle brackets: `< >`.

Literals Literal values like strings, numbers, and dates are represented directly without angle brackets.

Comments Turtle allows for comments, which are preceded by a hash symbol.

¹²⁰ IRI is an extension of URI that allows for the use of international characters and symbols in web addresses.

¹²¹ JSON-LD will be discussed through examples in Section 3.5.

Code Snippet 3.5: Turtle Serialisation of the RDF Graph

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix schema: <http://schema.org/> .
3
4 # A small RDF graph about myself
5 <https://www.example.org/julien-a-raemy>
6   rdf:type schema:Person ;
7   schema:givenName "Julien Antoine" ;
8   schema:familyName "Raemy" ;
9   schema:birthPlace <https://www.wikidata.org/entity/Q36378> .
10
11 <https://www.wikidata.org/entity/Q36378>
12   rdf:type schema:Place ;
13   schema:name "Fribourg" .
```

To query RDF-based graphs, SPARQL is leveraged. It is a query language as well as a protocol designed for querying and manipulating RDF data. It allows users to retrieve specific information from RDF datasets, making it a fundamental tool for working with Linked Data [Wood et al. 2014, pp. 99-100]. The SPARQL query provided in Code Snippet 3.6 is aimed at extracting the name of a person from a given RDF graph. The query uses RDF predicates associated with the given name and family name properties, concatenating them to form the person’s complete name. The answer to the query, based on the provided RDF data, would be the complete name of the person, which in this case is ‘Julien Antoine Raemy’. It showcases the versatility of SPARQL, offering a flexible and expressive means of interacting with RDF data. By specifying patterns and conditions, this query identifies and combines relevant information.

Code Snippet 3.6: SPARQL Query Against the RDF Graph

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX schema: <http://schema.org/>
3
4 SELECT ?name
5 WHERE {
6   <https://www.example.org/julien-a-raemy> schema:givenName ?givenName ;
7                                           schema:familyName ?familyName .
8   BIND(CONCAT(?givenName, " ", ?familyName) AS ?name)
9 }
```

While RDF’s simplicity is one of its strengths, it can also pose limitations when it comes to expressing nuanced details, tracing the origins of statements, and addressing the intricacies of logical reasoning. RDF’s foundational model, which relies on triples to establish relationships between resources, may not always capture the full depth of knowledge and inferences that more complex KR languages or systems can achieve.

As such, RDF reification, as expounded by Massari et al. [2023], emerges as a pivotal mechanism to enrich RDF. It stands as a mechanism that allows for the explicit representation of statements about statements in RDF. This technique facilitates the modelling of metadata or provenance information about RDF triples. In other words, RDF reification allows you to describe when and by whom a particular statement was made, adding a layer of context and traceability to RDF data. This is particularly valuable in scenarios where data lineage, trustworthiness, and attribution need to be maintained.

Furthermore, RDF 1.2 introduces the concept of ‘quoted triples’, i.e. ‘an RDF triple that is used as the subject or object of another triple’ [Hartig et al. 2023]. Quoted triples are a powerful extension to the RDF model, enabling the inclusion of entire RDF triples within other triples. This feature adds a layer of expressiveness to RDF, making it more versatile in representing complex relationships and data structures.

Deduction of entity class membership is an essential aspect of KR, and in this regard, RDFS plays a pivotal role [Brickley and Guha 2014]. RDFS, with its foundation in formal semantics, provides a framework for verifying properties like `rdfs:domain` and `rdfs:range` associated with given properties [Bruns et al. 2023].

Code Snippet 3.7 showcases how RDFS can be leveraged. In this example, RDFS class and property are specified. For instance, `schema:Person` and `schema:Place` are classes using `rdf:type rdfs:Class`, and property characteristics, including domains and ranges, using `rdf:Property`, `rdfs:domain`, and `rdfs:range` have been defined to create basic RDFS.

Code Snippet 3.7: Example of using RDFS to Determine Class Membership and Property Definitions

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
3 @prefix schema: <http://schema.org/> .
4
5 # RDFS Class Definitions
6 schema:Person rdf:type rdfs:Class .
7 schema:Place rdf:type rdfs:Class .
8
9 # RDFS Property Definitions
10 schema:givenName rdf:type rdf:Property ;
11     rdfs:domain schema:Person ;
12     rdfs:range rdfs:Literal .
13
14 schema:familyName rdf:type rdf:Property ;
15     rdfs:domain schema:Person ;
16     rdfs:range rdfs:Literal .
17
18 schema:birthPlace rdf:type rdf:Property ;
19     rdfs:domain schema:Person ;
20     rdfs:range schema:Place .
21
22 schema:name rdf:type rdf:Property ;
23     rdfs:domain schema:Place ;
24     rdfs:range rdfs:Literal .
```

RDFS, when used in conjunction with other reasoning methods, enhances the overall inferential capabilities of RDF, enabling more profound insights and KR. For example, OWL allows authors to decide how expressive they want to be, given the computational realities involved, which is not the case with RDFS.

Indeed, for modelling ontologies in the Semantic Web, OWL is the preferred language. It is a powerful and widely used language for developing ontologies that can be leveraged as metadata vocabularies on the web [Zeng and Qin 2022, p. 63] and provides a framework for creating, sharing, and reusing ontologies. OWL is a key component in the Semantic Web, enabling the formal representation of knowledge and complex relationships. OWL possesses several key features that make it a powerful language for modelling ontologies. Firstly, it provides a well-defined formal semantics, allowing computers to infer new knowledge and make logical deductions based on the information encoded in the ontology. Additionally, OWL supports a rich vocabulary of constructs for defining classes, properties, and individuals, enabling the creation of complex and structured ontologies.

Code Snippet 3.8 is an example of an ontology using OWL in Turtle. Two classes, `ex:Person` and `ex:Place`, along with two properties: `hasName` and `hasBirthPlace` have been defined. The `ex:Person` class is constrained to have a name, which is specified as a datatype property with a string value. Additionally, each `ex:Person` is required to have a `hasBirthPlace`, which is defined as an object property with a reference to the `ex:Place` class.

Code Snippet 3.8: Example of a OWL-based Ontology

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .
4 @prefix ex: <http://example.org/ontologies#> .
5
6 ex:Person rdf:type owl:Class .
7 ex:Place rdf:type owl:Class .
8
9 ex:hasName rdf:type owl:DatatypeProperty .
10 ex:hasBirthPlace rdf:type owl:ObjectProperty .
11
12 ex:Person rdfs:subClassOf [
13     rdf:type owl:Restriction ;
14     owl:onProperty ex:hasName ;
15     owl:someValuesFrom xsd:string
16 ] .
17
18 ex:Person rdfs:subClassOf [
19     rdf:type owl:Restriction ;
20     owl:onProperty ex:hasBirthPlace ;
21     owl:someValuesFrom ex:Place
22 ] .
    
```

OWL is grounded in DL [Ma and Hitzler 2009, p. 197], striking a balance between expressiveness and computational complexity, making it well-suited for automated reasoning. There are different versions of OWL, including OWL Lite, OWL DL, and OWL Full, which vary in terms of their expressive power and reasoning capabilities. OWL 2, introduced in 2009, extended the original OWL specification with additional features and improvements [W3C OWL Working Group 2012]. Figure 3.18 provides an overview of OWL, highlighting its main building blocks in terms of syntax as well as semantics.

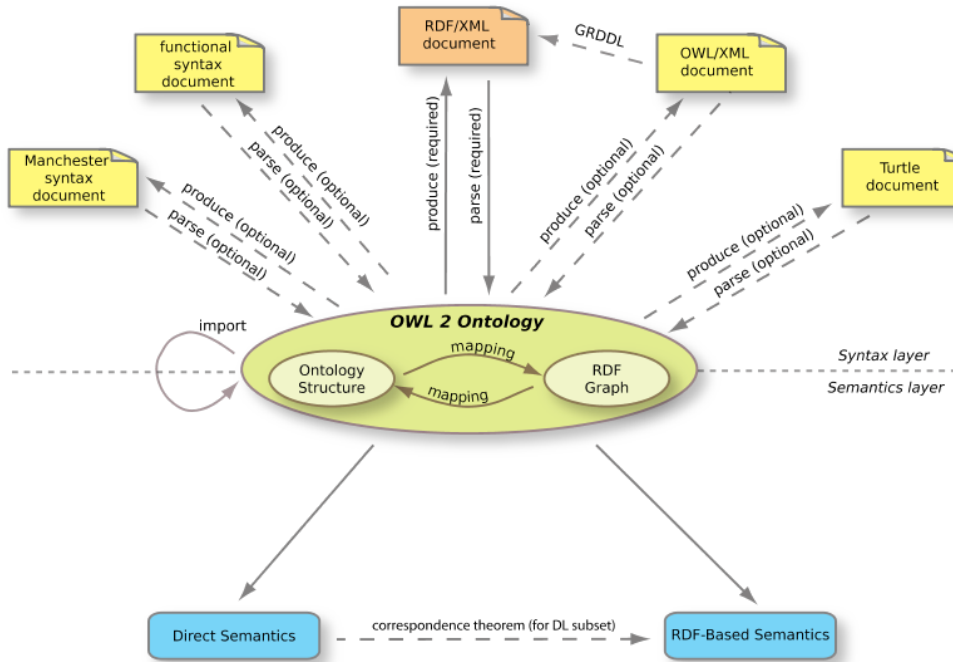


Figure 3.18: Structure of OWL 2 [W3C OWL Working Group 2012]

Turning to SHACL, a W3C standard defined in 2017, it provides a valuable framework for validating data graphs against specific shapes and constraints. Shapes are templates that specify the expected structure, properties, and validation rules for a particular class of resources, while constraints are rules

or conditions applied to data instances to ensure they conform to the structure and validation criteria defined by a shape. Overall, SHACL is primarily used for data validation and quality assessment [Knublauch and Kontokostas 2017].

Code Snippet 3.9 is an example of SHACL could be used to validate the RDF data graph of Code Snippet 3.5. The following shapes and constraints have been defined:

- `schema:PersonShape` which corresponds to `schema:Person`;
- `schema:PlaceShape` which relates to `schema:Place`;
- `sh:targetClass` to specify the target class;
- `sh:path` to specify the property path;
- `sh:minCount` to ensure that the specified properties are required;
- `sh:nodeKind sh:IRI` to specify that the `schema:birthPlace` property should have IRI values.

Code Snippet 3.9: Example of a SHACL Document in Turtle that can be Used to Validate the RDF Graph

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix schema: <http://schema.org/> .
3 @prefix sh: <http://www.w3.org/ns/shacl#> .
4
5 # Define a SHACL shape for a Person
6 schema:PersonShape
7   a sh:NodeShape ;
8   sh:targetClass schema:Person ;
9   sh:property [
10     sh:path schema:givenName ;
11     sh:minCount 1 ;
12   ] ;
13   sh:property [
14     sh:path schema:familyName ;
15     sh:minCount 1 ;
16   ] ;
17   sh:property [
18     sh:path schema:birthPlace ;
19     sh:minCount 1 ;
20     sh:nodeKind sh:IRI ;
21   ] .
22
23 # Define a SHACL shape for a Place
24 schema:PlaceShape
25   a sh:NodeShape ;
26   sh:targetClass schema:Place ;
27   sh:property [
28     sh:path schema:name ;
29     sh:minCount 1 ;
30   ] .
```

One thing to realise, however, is that there are limitations when it comes to reasoning with controlled vocabularies and constraints in SHACL. While it excels in data validation and quality assurance, SHACL is not inherently a comprehensive reasoning framework. In their study, Sacramento et al. [2022, p. 129] noted that there is still a challenge in harnessing the potential of ontologies' axioms to deduce information and derive implicit facts. While ontologies are instrumental in structuring knowledge and relationships between concepts, the efficient utilisation of axioms for reasoning remains an ongoing challenge.

3.4.3 Linked Data Principles

By exploiting the core structure of the web's architecture, Linked Data maximises its potential to enable the widespread distribution of information on a global scale. This approach makes effective use of the fundamental elements of the Web and is consistent with its principles of simplicity, decentralisation and openness. As a result, Linked Data exploits the optimal structural equilibrium of the web, allowing content from disparate servers to be assembled seamlessly into a unified repository of global knowledge. This harmonious convergence of principles has driven the remarkable growth of the Web over the past few decades, demonstrating its ability to provide an integrated platform for information exchange. The four principles of Linked Data, as outlined by [Berners-Lee \[2006\]](#) are:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (e.g. RDF, RDFS, SPARQL, etc.)
4. Include links to other URIs so that they can discover more things.

By following these principles, Linked Data enables structured data to be published and interlinked on the Web, creating a graph of data that can be accessed and navigated by both humans and machines. These principles are closely related to the Semantic Web, which is an extension of the web that aims to give information well-defined meaning, enabling computers and people to work together more effectively [[Heath and Bizer 2011](#)].

'The glue that holds together the traditional document Web is the hypertext links between HTML pages. The glue of the data web is RDF links.' [[Bizer et al. 2008](#), p. 1265]

Some of the challenges highlighted by [Bizer et al. \[2009\]](#) as a retrospective ten years after the launch of RDF were and still remain an issue:

User Interfaces and Interaction Paradigms Challenges persist in designing user interaction paradigms for applications based on dynamically assembled data. These challenges include navigating between entities and managing data sources within an integrated view.

Application Architectures The potential for scalability issues in on-the-fly link traversal and federated querying suggests that extensive crawling and caching could become the standard for timely data access.

Schema Mapping and Data Fusion While most applications display data from diverse sources, they often lack deeper integration. Achieving integration involves mapping terms from different vocabularies and resolving data conflicts. Data sources can publish correspondences to help clients transform data between schemata.

Link Maintenance Link maintenance in evolving sources is a research challenge. Outdated links can affect system efficiency. Solutions include periodic link recalculation, data source update feeds, and central registries.

Licensing Applications using web data need access to specifications governing data reuse and republishing. Frameworks for publishing such specifications are vital for encouraging data owners' participation and providing legal assurance to data consumers.

Trust, Quality and Relevance Representing data trustworthiness in integrated views is a research challenge. Heuristic methods for assessing data quality should be leveraged.

Privacy Safeguarding privacy may require technical and legal measures, as well user awareness.

These challenges serve as a foundation for exploring LOD practices in the subsequent subsection, where I introduce two deployment schemes.

3.4.4 Deployment Schemes for Open Data

LOD is a potent blend of Linked Data and Open Data that is both linked and uses open sources. LOD is defined as Linked Data released under an open licence that does not prevent it from being reused free of charge.

In 2010, Tim Berners-Lee introduced the Five-Star Open Data Deployment Scheme¹²², also known as Five-Star Open Data or Five-Star Linked Data, to provide a structured framework for publishing and promoting data on the web. The scheme, as illustrated by Figure 3.19, comprises five progressively more demanding criteria that define the level of openness and accessibility of data.

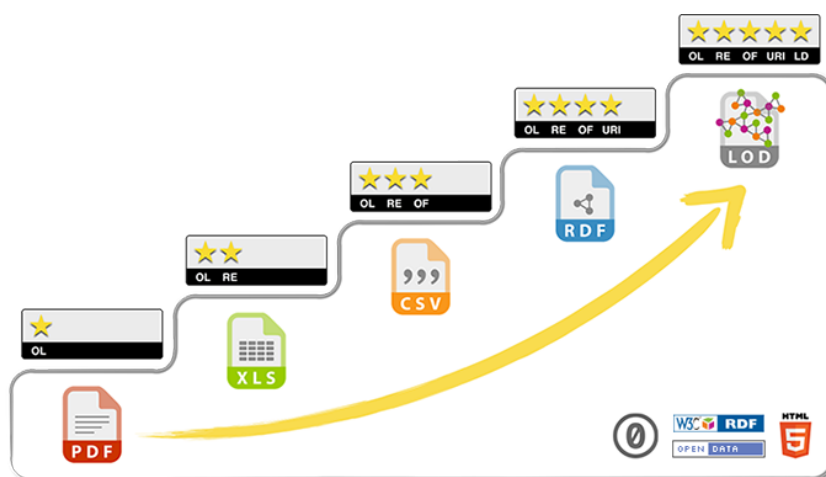


Figure 3.19: Five-Star Deployment Scheme for Open Data

- ★ The 1st star indicates that the data are available on the web in any format, setting a basic level of accessibility.
- ★★ Moving on to the 2nd star, data should be made available in a machine-readable, structured format that allows for easier processing and analysis. This eliminates the need to interpret unstructured information and promotes efficient use of data.
- ★★★ The 3rd star builds on the second by emphasising the importance of using non-proprietary formats such as CSV rather than vendor- or platform-specific formats such as Microsoft Excel. This promotes interoperability and avoids vendor lock-in, ensuring wider access and use of the data.
- ★★★★ The 4th star takes the criteria further by requiring that data be published using open standards specified by the W3C, in particular RDF and SPARQL. These standards improve the integration and linking of data across the Web, enabling richer and more meaningful analysis.
- ★★★★★ The highest standard, represented by the 5th star, incorporates all of the previous criteria and requires that data not only be in a non-proprietary format and conform to W3C standards, but also be linked to other LOD. This interconnectedness enables a broader and richer web of data, increasing its value and utility in a global context.

¹²² 5-star Open Data: <https://5stardata.info/>

By adhering to these progressively stringent criteria, data providers can contribute to a more open and connected web, fostering a collaborative environment for data consumption and analysis. The Linked Data Principles and the Five-Star Open Data Deployment Scheme share commonalities, including adherence to open standards and linking data with other sources. However, the rating system offers greater specificity and provides a well-defined path for data publishers to improve openness and interconnection of data.

Yet, [Hasnain and Rebolz-Schuhmann \[2018\]](#) noted in their study around assessing FAIR against the Five-Star Deployment Scheme that while LOD emphasises data openness, FAIR additionally requires a stated licence for access, incorporating the concept of reusability with consideration of licensing agreements and contextual information such as provenance. As such, FAIR can be seen as an extension of LOD, focusing on user needs and incorporating aspects of separation of core data and metadata, including licensing considerations, while LOD remains an idealistic approach centred on open data.

In complement and in response to the Five-Star model, [Hyvönen et al. \[2014\]](#) proposed an extended Seven-Star model to address some further key challenges in dataset reuse: the difficulty in assessing how suitable the data are for a particular application purpose. They highlighted that datasets often lack clear definitions or descriptions of the schemas or vocabularies they use, making it difficult to effectively understand the characteristics of the data. Furthermore, even when both the data and its schema are available, assessing the data's conformance to the schema is hampered by the data quality issues that are prevalent in the Semantic Web. To mitigate these challenges and to encourage data publishers, they introduced two additional stars.

- ★ ★ ★ ★ ★ The 6th star is awarded if the schemas or vocabularies used in the dataset are explicitly described and published alongside the dataset, unless these schemas are already available elsewhere on the Web. This documentation improves the understanding of the structure and organisation of the data.
- ★ ★ ★ ★ ★ ★ To attain the 7th star, the quality of the dataset in terms of the schemas it uses must also be explicitly described. This information enables users to assess whether the quality of the data meets their specific requirements and provides a valuable tool for determining the suitability of the dataset for their intended applications.

The Seven-Star model thus extends Berners-Lee's Five-Star Open Data Deployment Scheme to provide a comprehensive framework for improving the clarity, usability and applicability of open data on the web¹²³.

3.4.5 Linked Open Data in the Cultural Heritage Domain

In this subsection, I will present several illustrative cases of CH projects or CHIs that have engaged or are engaging with LOD, offering insights into some of the processes, benefits or challenges, such as issues related to sustainability and community engagement.

The application of LOD in the CH domain has received considerable attention due to its ability to improve the quality and visibility of data provided by institutions, especially from GLAMs [[Candela et al. 2018](#), p. 481]. Equally, the development of DH has seen a growing interest in Linked Data, where research projects have increasingly adopted semantic enrichment approaches to improve the quality,

¹²³ It is worth mentioning a related initiative to illustrate all datasets available as LOD in a so-called LOD cloud, which attempts to show the extent of LOD by domain or type on the web, but the website (<https://lod-cloud.net/>) is sometimes inaccessible. The sequence of diagrams, from 2007 up to 2020, can still be accessed on Wikidata: <https://www.wikidata.org/entity/Q43984865>

annotation, and visibility of their data such as the Pelagios network¹²⁴ [see [Isaksen et al. 2014](#)], while at the same time encouraging a more self-sustaining environment [see [Zeng 2019](#)].

According to [Davis and Heravi \[2021, p. 21:5\]](#), the CH sector's involvement in Linked Data research began approximately a decade to 15 years ago, with a notable peak around 2015. [Davis and Heravi \[2021, pp. 21:7-21:8\]](#) identified some key players within the CH sector that have been involved in the implementation of Linked Data, among the most prominent: universities, university libraries, national libraries, government bodies, as well as museums. Within this context, the authors identified eight prominent themes that emerged to signify the motivations behind this interest. These themes included meeting research needs, exploring Linked Data, meeting the needs of users, improving discoverability, fostering interoperability, educational initiatives, meeting the needs of GLAM institutions, as well as prioritising preservation efforts [[Davis and Heravi 2021, pp. 21:9-21:10](#)]. Moreover, [Candela et al. \[2023b\]](#) assert that LOD is also a motivation using advanced methods and techniques for publishing and reusing digital collections.

The deployment of LOD is primarily evident through publication of datasets, which I will highlight here in three different ways, ranging from the most rudimentary form of publication to the most complex. First, records may be accompanied by a representation conforming to the RDF syntax, which can be downloaded in RDF/XML or Turtle. A second approach involves establishing a content negotiation mechanism that enables both humans and machines to access metadata. Thirdly and on a more occasional basis, platforms have implemented a SPARQL service allowing for federated queries, either on a dedicated web interface or through a Command-line interface (CLI) [[Raemy 2022a, pp. 132-133](#), citing [Papadakis et al. \[2015\]](#)].

This publication effort is carried out not only by libraries, archives, and museums but also by aggregation platforms such as Europeana¹²⁵ [[Purday 2009](#)], which provides access to millions of digitised books, images, and audiovisual resources from European CHIs. Europeana uses LOD to create rich, interconnected descriptions of these resources, making it easier for users to discover and explore related resources across different institutions and collections [[Poulopoulos and Wallace 2022, p. 6](#)].

Europeana is a project funded by the European Union that aims to promote a sense of European identity by aggregating metadata from national and local heritage institutions onto a single digital platform [[Freire et al. 2018c, Raemy 2020](#)]. The EDM is the core of the project, which leverages standardised thesauri and vocabularies to provide semantic contextualisation for CH resources and allow for semantic operations on metadata [[Charles and Isaac 2015](#)]. However, the overarching nature of EDM may not deliver the level of granularity that all CHIs require to accurately document their resources. Despite this, memory institutions accept the sacrifice of accuracy to be part of a Europe-wide collection that promotes a sense of Europeaness. The study by [Capurro and Plets \[2020\]](#) traces how a standardised European metadata structure plays a role in governing local and national heritage institutions and enacting a European mindset. The EDM may allow heritage stakeholders to benefit from Europeana's online exposure while also contributing to the creation of a European identity.

The landscape of LOD research project predominantly originates from prominent institutions and national aggregators within the Europeana network. For instance, the Bibliothèque nationale de France (BnF) has made substantial contributions through its data service¹²⁶ [[Simon et al. 2013](#)].

¹²⁴ Pelagios network: <https://pelagios.org>

¹²⁵ Europeana: <https://www.europeana.eu/>

¹²⁶ BnF Data: <https://data.bnf.fr/>

Additionally, the Swedish National Heritage Board¹²⁷ for metadata aggregation has played a crucial role in metadata aggregation, as outlined in the work of Smith [2021, pp. 67-69]. Furthermore, the Dutch Digital Heritage Network¹²⁸ has been actively engaged in the management of KOSs [Scharnhorst et al. 2023].

Another example of LOD in the CH domain that has already been mentioned is the Getty Research Institute's controlled vocabularies, which provides access to data on art history, architecture, and related fields. The initiative uses Linked Data principles to create a network of interlinked resources, including artists, artworks, and publications, allowing researchers to explore complex relationships between different CH resources [Cobb 2015, Harpring 2010]. Similarly, authority records such as the Gemeinsame Normdatei (GND), the Library of Congress Subject Headings (LCSH), or a collaborative structured service such as Wikidata, a collaboratively edited KG, have gained significance and have become hubs connecting various identifiers and promoting the use of related concepts [Chardonnes 2020, Thalhath et al. 2021]. Moreover, Lodi et al. [2017] highlight that interconnected CH resources should be managed separately due to diverse definitions and classifications in the GLAM sector.

While quite a few (well-funded) libraries have played a crucial role in advancing Linked Data in the CH domain, there remain substantial challenges. Raza et al. [2019, p. 11] underline these challenges, including issues related to cataloguing, the adoption of new standards, a proliferation of vocabularies and ontologies, and the absence of data access agreements. Additionally, challenges include a scarcity of Linked Data tools and expertise, mapping difficulties, ownership disputes, and the management of multilingual and heterogeneous data. Similarly, Banerjee [2020]'s critique of LOD highlights that Linked Data terminology often seems unnecessarily complex and obscure. He emphasises that Linked Data is most effective for technical challenges with well-maintained metadata, ontologies and vocabularies. He considered for instance that the application of Linked Data in general library use quite inappropriate due to the complexity mismatch and limited practical benefits. The author suggests that its use should be reserved for specific domains with comprehensive vocabularies, and warns against its indiscriminate adoption for tasks better suited to simpler, established methods.

For instance, in Switzerland, efforts were made to enhance data accessibility and interoperability through the development of linked.swissbib.ch. This platform was designed to serve as the LOD counterpart to Swissbib, a comprehensive meta-platform of bibliographic resources [Prongué and Schneider 2015, pp. 125-126]. The initial process of creating this data complement involved transforming bibliographic MARC/XML records into JSON-LD, a structured data format conducive to Linked Data integration [Hipler et al. 2018, p. 166]. This transformation played a foundational role in preparing the data for inclusion in linked.swissbib.ch. However, it is important to note that the meta-platform Swissbib, along with its envisioned Linked Data service, linked.swissbib.ch, is no longer in existence and at the time of writing, there is no active Linked Data service directly equivalent to linked.swissbib.ch for libraries in Switzerland.

Transitioning from the conventional logic of data publication, Hyvönen [2020, p. 187-190] suggests that there is a need for a transition in the use of Semantic Web portals, shifting from a data publication logic to one of analysis and the serendipitous discovery of knowledge. He categorises three generations of portals:

1. Portals for search and browsing which can handle data harmonisation, aggregation, search, and navigation.

¹²⁷ Swedish National Heritage Board: <https://www.raa.se/>

¹²⁸ Dutch Digital Heritage Network: <https://netwerkdigitaalerfgoed.nl>

2. Portals with tools for distant reading that provides users with an integrated set of tools for interactive research problem-solving.
3. Portals for serendipitous knowledge discovery based on AI and which may automatically address research queries based on constraints set by scholars and savvy users.

While Hyvönen [2020] asserts that while DH and CHIs have contributed to the deployment of second-generation systems, the path forward necessitates a greater emphasis on source criticism and the proficient use of sophisticated computing tools to reach the third stage. This progression is essential to expand the scope and capabilities of DH and beyond.

Bernasconi et al. [2023, p. 12] corroborate this viewpoint some extent as they assert that LOD-driven GUIs should be able to integrate different interaction paradigms ‘*to support the creation and exploration for data models by domain experts who may not pose a strong technical background on ontologies and RDF*’. For example, the LOD4Culture project¹²⁹ is an initiative aimed at accessing CH data. LOD4Culture has implemented a RESTful API as detailed by Vega-Gorgojo [2023]. However, it is important to note that, even though an API is used, the underlying technology driving the system remains SPARQL.

As highlighted earlier, Linked Data’s journey in the CH domain has witnessed significant attention, with its potential benefits and capabilities as well as its limitations. Davis and Heravi [2021, p. 21:2, citing Linden [2015]] discuss that Gartner’s 2015 Hype Cycle¹³⁰ placed Linked Data in the ‘trough of disillusionment’ and predicted that it would reach the ‘plateau of productivity’ within the next 5 to 10 years. Linked data has the potential to improve interoperability, information sharing and standard practices in the CH sector, in line with its goals of sharing resources and advancing knowledge.

Furthermore, Davis and Heravi [2021, pp. 21:13-21:14]’s literature review on CH leveraging LOD reflects the alignment of collaboration, research needs, education, discoverability and interoperability with the ideals of the Linked Data movement and the overarching vision of the Semantic Web. However, their emphasis on the need for CHIs to prioritise collaboration for effective Linked Data project implementation underscores a critical aspect of this landscape.

In the context of my literature review, a notable research gap emerges, highlighting the importance of interactions with the data and the paramount role of collaboration. Furthermore, it is crucial to acknowledge that a number of LOD projects or platforms have not been actively maintained, and those that have are more likely to belong to wealthy institutions. Therefore, a comprehensive exploration of the LOUD design principles and their underlying specifications, is necessary. This exploration is particularly valuable because beneath the surface of LOUD standards lies the potential for community-driven practices enabling different degrees of semantic interoperability, ranging from basic to sophisticated.

3.5 Linked Open Usable Data

According to Sanderson [2016], the Semantic Web remains a compelling concept, with RDF graphs offering powerful capabilities for modelling reality, enabling computers to infer novel information more

¹²⁹ LOD4Culture: <https://lod4culture.gsic.uva.es/>

¹³⁰ ‘Gartner’s Hype Cycle, introduced in 1995, characterizes the typical progression of an emerging technology from overenthusiasm through a period of disillusionment to an eventual understanding of the technology’s relevance and role in a market or domain’ [Linden and Fenn 2003, p. 5]. The *x* axis is concerned with maturity and the *y* axis with visibility. The five phases of the curve, on the *y* axis, are as follows: technology trigger, peak of inflated expectations, trough of disillusionment, slope of enlightenment, and plateau of productivity.

effectively. However, Sanderson contends that querying RDF graphs is complex, requiring a prior understanding of the structure. He notes that ‘*data [is] currently treated as [a] second class citizen*’, highlighting challenges such as the lack of a clear start and end in XML serialisation compared to graphs, the difficulty of handling documents in triplestore storage, and sub-optimal visualisation practices. This aligns with Target [2018]’s argument that achieving the Semantic Web on a global scale is quite challenging. Yet, despite these challenges, some of the Semantic Web’s intent is now achievable thanks to the creation and deployment of JSON-LD. Notably, this serialisation can be treated as regular JSON, a lightweight data interchange format heavily used by software developers, or as a graph [Raemy and Sanderson 2023], providing flexibility in representation and addressing some of the issues mentioned previously.

Transitioning into the domain of community-driven and JSON-LD-based specifications and their transformative potential, Linked Open Usable Data (LOUD) emerges as a key player. It is a term coined by Sanderson [2019b], who has been involved in the conception and maintenance of the standards discussed in this section, which is organised into six parts.

Beginning with subsection 3.5.1, I explore the five core design principles or characteristics of LOUD. Then, subsection 3.5.2 presents a systematic review of LOUD. Subsection 3.5.3 provides an overview of IIIF, encompassing its inception, community, specifications, and the IIIF ecosystem. Subsection 3.5.4 delves into the specifics of WADM, a W3C standard provides a structured framework for representing annotations on the web. Then, subsection 3.5.5 goes into the history, conceptual model, API and standards for Linked Art. It also explores the community around Linked Art and its current or planned implementations. Finally, subsection 3.5.6 discusses the implications and opportunity that LOUD brings to the field of CH and the humanities at large.

3.5.1 LOUD Design Principles

One of the main purposes of LOUD is to make the data more easily accessible to software developers, who play a key role in interacting with the data and building software and services on top of it, and to some extent to academics. As such, striking a delicate balance between the dual imperatives of data completeness and accuracy, which depend on the underlying ontological construct, and the pragmatic considerations of scalability and usability, becomes imperative. Similar to Tim-Berners Lee’s Five Star Open Data Deployment Scheme, five design principles underpin LOUD¹³¹.

- A. The right Abstraction for the audience** Developers do not need the same level of access to data as ontologists, in the same way that a driver does not need the same level of access to the inner workings of their car as a mechanic. Use cases and requirements should drive the interoperability layer between systems, not ontological purity.
- B. Few Barriers to entry** It should be easy to get started with the data and build something. If it takes a long time to understand the model, ontology, SPARQL query syntax and so forth, then developers will look for easier targets. Conversely, if it is easy to start and incrementally improve, then more people will use the data.
- C. Comprehensible by introspection** The data should be understandable to a large degree simply by looking at it, rather than requiring the developer to read the ontology and vocabularies. Using JSON-LD lets us to talk to the developer in their language, which they already understand. In other words, the data can be treated as a document serialised in JSON and/or as a graph.

¹³¹ Design Principles of LOUD: <https://linked.art/loud/>

D. Documentation with working examples You can never intuit all of the rules for the data. Documentation clarifies the patterns that the developer can expect to encounter, such that they can implement robustly. Example use cases allow contextualisation for when the pattern will be encountered, and working examples let you drop the data into the system to see if it implements that pattern correctly.

E. Few exceptions, instead many consistent patterns Every exception that you have in an API (and hence ontology) is another rule that the developer needs to learn in order to use the system. Every exception is jarring, and requires additional code to manage. While not everything is homogeneous, a set of patterns that manage exceptions well is better than many custom fields.

The concerns articulated by [Hyvönen et al. \[2014\]](#) in their Seven-Star Model in terms of schema and data validation are also indirectly addressed by the LOUD design principles. Conceptualisations of LOUD specifications and their representation, mostly through usable APIs – echoing human-centred approach for developing them as articulated by [\[Myers and Stylos 2016\]](#), address schema concerns. For data validation, best practices and validators developed by the IIF and Linked Art communities come into play. Moreover, LOUD indirectly respond to the needs of scientists who advocate that ‘*Linked Data wasn’t enough*’ and emphasise treating research objects, data, as first class citizens for reproducibility purposes [see [Bechhofer et al. 2013](#)].

Highlighting the success of the practices that guided IIF, [Sanderson \[2020b\]](#) identifies three systems adhering to the LOUD design principles: IIF and specifically the third version of the Presentation API, WADM, and Linked Art. These three systems are complementary and can be used either separately or in conjunction. Figure 3.20 illustrates an high-level overview of an infrastructure combining all three of the LOUD specifications.

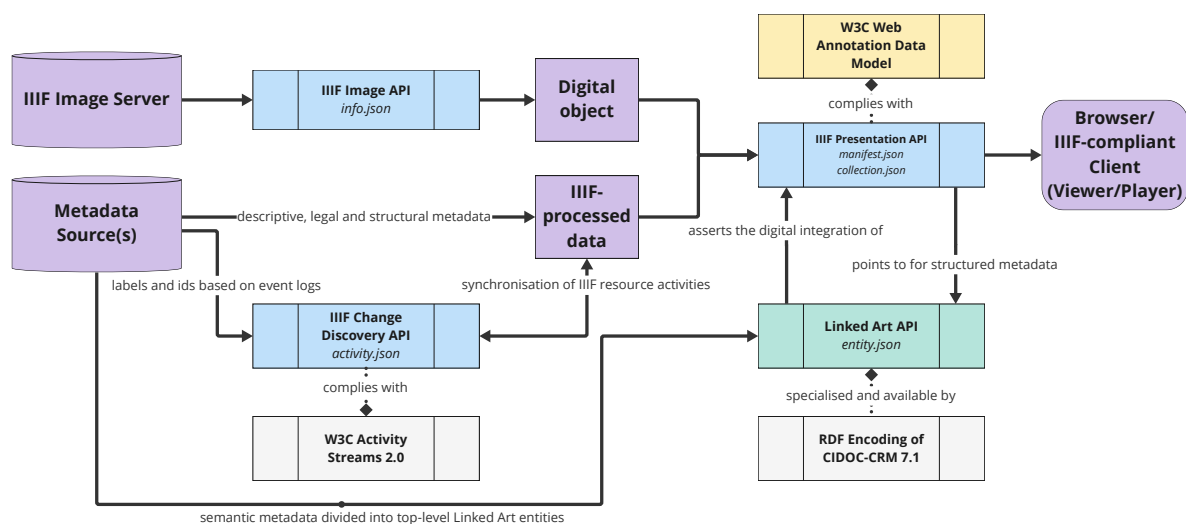


Figure 3.20: Example of a LOUD-Driven Infrastructure [[Felsing et al. 2023](#), p. 43]

In summary, the LOUD design principles, guided by considerations of accessibility, ease of use, comprehensibility, documentation, and consistency, not only address crucial concerns raised by Linked Data practitioners but also respond to the evolving needs of the CH and scientific communities emphasising data reproducibility. In the following subsection, I will explore the presence and impact of LOUD and their underlying principles in the scholarly landscape.

3.5.2 LOUD: Systematic Review

This subsection deals with a systematic review of academic references and ongoing projects related to LOUD. It examines how the concept is mentioned in the academic literature.

To conduct this systematic review, I employed the weight of evidence framework¹³² developed by Gough [2007, p. 218-219]. This framework consists of nine criteria, starting with formulating the review question and developing a protocol. In this case, the review question was formulated as a Boolean query, as shown in Equation 3.2, due to the absence of an existing systematic review for LOUD. The criteria also include defining inclusion and exclusion criteria for papers and conducting a systematic search strategy across academic databases. Screening, mapping, data extraction, quality and relevance appraisal, synthesis, and communication and engagement are other essential steps in this framework.

For the systematic review, I used a specific Boolean search query¹³³ to identify relevant research papers and resources. The query, as presented in Equation 3.2, was applied across academic databases, including Google Scholar¹³⁴, Semantic Scholar¹³⁵, Web of Science, and Scopus. Specifically, the query seeks references containing either ‘Linked Open Usable Data’ or ‘LOUD Design Principles’ or ‘LOUD Principles’ or instances of ‘LOUD’ combined with any of the three recognised standards adhering to its design principles: IIF, Linked Art, and WADM.

$$Q = \text{“Linked Open Usable Data”} \vee \text{“LOUD Design Principles”} \vee \text{“LOUD Principles”} \\ \text{“LOUD”} \wedge \text{“IIF”} \vee \text{“LOUD”} \wedge \text{“Linked Art”} \vee \text{“LOUD”} \wedge \text{“Web Annotation”} \quad (3.2)$$

In Google Scholar, 60 results were returned, with 41 deemed relevant. Semantic Scholar’s results were not available due to difficulties with the search process as it yielded more than 800k results. Web of Science returned no relevant results, and Scopus provided four results, of which three were unique items. I also queried the Zenodo repository¹³⁶, resulting in 25 hits, including two conference papers, one journal article, and one working paper. The remaining results included presentations or datasets, mostly associated with my own work.

Database	Total Items	Relevant Entries	Unique Entries
Google Scholar	60	41	41
Semantic Scholar	N/A	N/A	N/A
Web of Science	0	0	0
Scopus	4	4	3
Zenodo	25	4	2
Total	89	48	46

Table 3.2: Summary of Database Search Results on LOUD

As of mid-November 2023, the systematic review yielded a total of 46 results, covering the period between 2018 and 2023. These results are summarised in Table 3.2. For more detailed information about each reference, including their type, language, source database, and categorisation, refer to Table 3.3. For instance, the references were predominantly written in English (37), with additional contributions in

¹³² Initially, a snowballing method was considered to find additional references [see Wohlin 2014], but it yielded limited results. Most papers referred to only one or two references related to LOUD, primarily citing Sanderson [2018] and the design principles on the Linked Art website.

¹³³ \vee stands for OR and \wedge for AND.

¹³⁴ Google Scholar: <https://scholar.google.com/>

¹³⁵ Semantic Scholar: <https://semanticscholar.org/>

¹³⁶ Zenodo: <https://zenodo.org/>

German (5), French (2), and one reference being bilingual in German and French. Additionally, one paper was in Japanese.

To categorise the papers, I employed a classification scheme with three main categories: α , β , and γ . Within α , there are four subcategories: α_1 (**Mention**) for papers mentioning LOUD, α_2 (**Description**) for those providing descriptions, α_3 (**Principles**) for explanations of the design principles, and α_4 (**Analysis**) for comparative analyses. β (**Standards**) represents papers mentioning recognised LOUD standards, i.e. IIF, WADM, and Linked Art, while γ (**Application**) covers LOUD applications or reflection to other standards. Some papers fall into multiple categories. Figure 3.21 displays the distribution of references across these categories, categorised by year. I also highlighted in yellow the nine references I took part in.

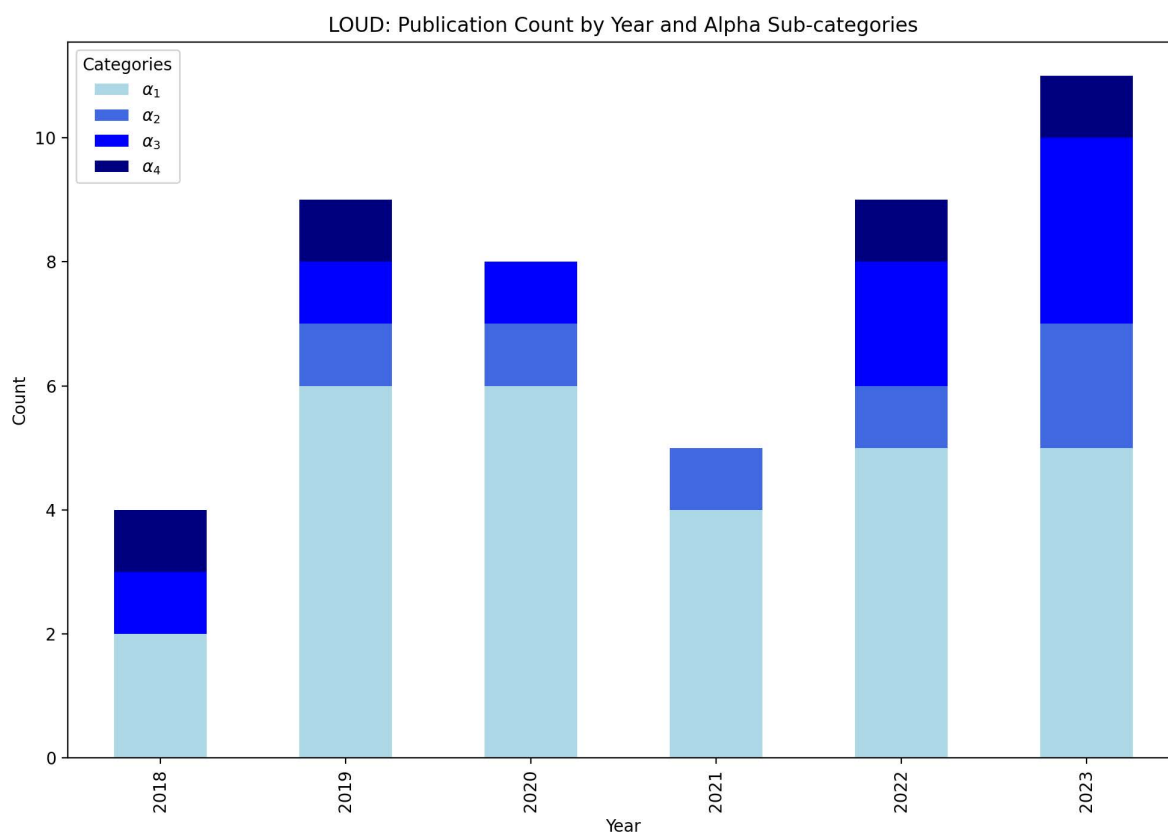


Figure 3.21: LOUD References Ordered by Year and α Sub-categories

α_1 – **Mention** Papers that mention LOUD as a new concept, accounting for 30 out of 46 papers. For instance, [Alexiev \[2018\]](#) introduced LOUD as a novel moniker or approach to Linked Data.

α_2 – **Description** Papers providing descriptions of what LOUD is, including varying perspectives on its target audience, such as human-readability or data practitioners [[Adamou 2022](#), [Schmidt et al. 2022](#)].

α_3 – **Principles** Papers explaining LOUD design principles, often authored by a small group that know each other such as [Delmas-Glass and Sanderson \[2020\]](#) or myself. They also frequently referenced IIF and Linked Art.

α_4 – **Analysis** Papers that interpret and reuse LOUD design principles for diverse applications, offering comparative analyses or presenting case studies. For instance, [Thiery \[2019\]](#) proposed a model of seven data spheres, with LOUD as the 6th sphere and LOUD and FAIR together in the 7th sphere.

In another case, Pohl et al. [2018] highlights the importance of the principles for lobid¹³⁷, a LOD service in Germany that consists of GUIs and APIs.

Figure 3.22 depicts Venn diagrams illustrating all six intersections of these categories, revealing how different aspects of LOUD are covered in the literature.

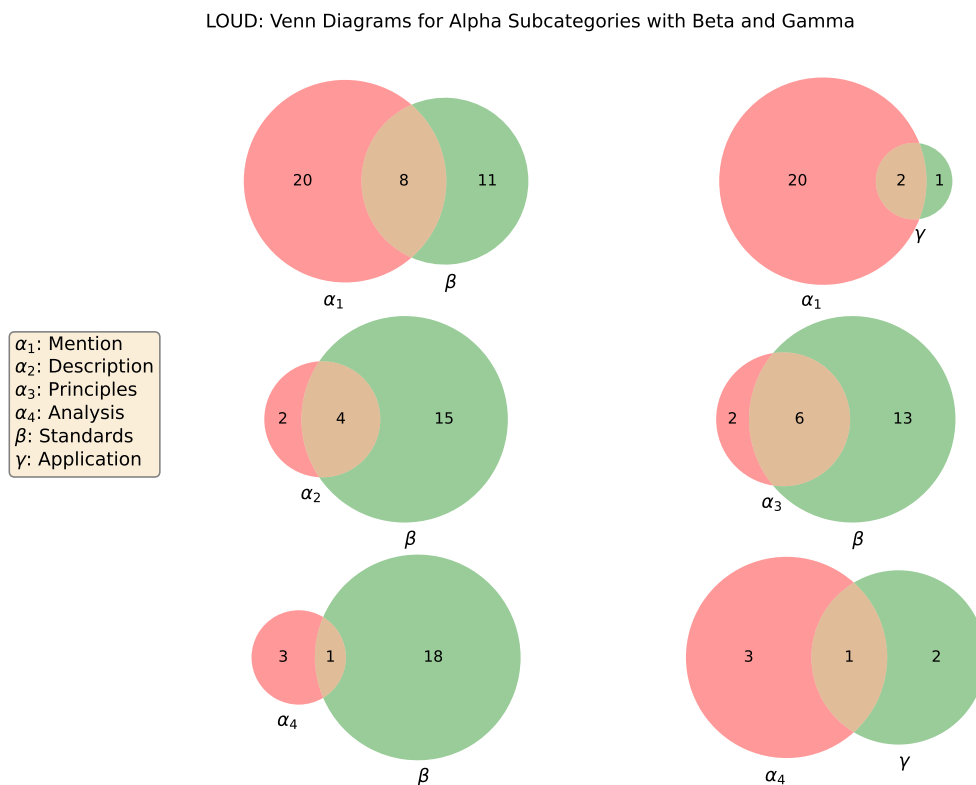


Figure 3.22: LOUD References as Venn Diagrams for Each of the Six Intersections

Several papers related to archaeology also made references to LOUD in the context of the CAA Data Dragons Interest Group’s work on semantics and LOUD¹³⁸. Furthermore, a search on popular search engines, specifically Bing and Google, revealed the existence of a research project named simply LOUD. This project, initiated in 2019 by the Royal Museum for Central Africa in Belgium¹³⁹, was delayed due to the COVID-19 pandemic and is now being considered for future re-initiation.

Further research should explore projects that implement a combination of the above standards, without necessarily mentioning LOUD. It is important to note that a quick search will show many IIF implementations that reference WADM, but it is mainly due to the Presentation API data model’s dependence on this specification. However, such mentions of WADM often do not indicate active adoption of web annotations for broader purposes. In these cases, while these projects still adhere to the design principles of LOUD by implementing IIF, I would argue that the full potential of LOUD is realised through a combination of specifications, especially when it comes to achieving semantic interoperability¹⁴⁰.

¹³⁷ lobid: <https://lobid.org>

¹³⁸ CAA Data Dragons: <https://datadragon.link/>

¹³⁹ Project at the Royal Museum for Central Africa: <https://fine-arts-museum.be/en/research/research-projects/loud>

¹⁴⁰ A good example of projects combining different specifications without explicitly mentioning LOUD is the O’Keefe Museum, where IIF and Linked Art have been applied for cross-collection exploration [see Degler et al. 2020].

Reference	Type	Lang.	DB	Cat.
Alexiev [2018]	Journal Article	eng	Google Scholar	α_1
Harpring [2018]	Conference Paper	eng	Scopus	α_1
Newbury [2018]	Conference Paper	eng	Google Scholar	α_3, β
Pohl et al. [2018]	Journal Article	ger	Google Scholar	α_4, γ
Cossu [2019b]	Journal Article	eng	Google Scholar	α_1, β
France and Forsberg [2019]	Conference Paper	eng	Scopus	α_1, β
Hofmann et al. [2019]	Journal Article	ger	Google Scholar	α_1
Klammt [2019]	Journal Article	ger, fre	Google Scholar	α_2
Klic [2019]	PhD Thesis	eng	Google Scholar	α_1, β
Nakamura [2019]	Conference Paper	jpn	Google Scholar	α_1, β
Sanderson [2019b]	Conference Paper	eng	Google Scholar	α_3, β
Thiery [2019]	Working Paper	eng	Zenodo	α_4
van Ginhoven and Rasterhoff [2019]	Journal Article	eng	Google Scholar	α_1
Brown and Martin [2020]	Conference Paper	eng	Google Scholar	α_1
Delmas-Glass and Sanderson [2020]	Journal Article	eng	Google Scholar	α_3, β
Paquet [2020]	Journal Article	eng	Google Scholar	α_2
Romein et al. [2020]	Journal Article	eng	Google Scholar	α_1
Schmidt and Thiery [2020]	Conference Paper	eng	Google Scholar	α_1
Thiery [2020]	Journal Article	ger	Google Scholar	α_1
Thiery et al. [2020a]	Conference Paper	eng	Google Scholar	α_1, γ
Thiery et al. [2020b]	Conference Paper	eng	Google Scholar	α_1, γ
Brown et al. [2021]	Conference Paper	eng	Google Scholar	α_1
Delmas-Glass [2021]	Book Chapter	eng	Google Scholar	α_2, β
Pohl [2021]	Journal Article	eng	Google Scholar	α_1
Raemy [2021]	Conference Paper	eng	Google Scholar	α_1, β
Roberts et al. [2021]	Journal Article	eng	Google Scholar	α_1
Adamou [2022]	Journal Article	eng	Google Scholar	α_3
Binding et al. [2022]	Journal Article	eng	Google Scholar	α_1
Hopkins [2022]	Journal Article	eng	Google Scholar	α_1
Lemos et al. [2022]	Journal Article	eng	Scopus	α_1
Middle [2022]	PhD Thesis	eng	Google Scholar	α_4
Raemy [2022a]	Conference Paper	fre	Google Scholar	α_3, β
Roke and Tillman [2022]	Journal Article	eng	Google Scholar	α_1
Schmidt et al. [2022]	Journal Article	eng	Google Scholar	α_2, β
Wigg-Wolf et al. [2022]	Journal Article	ger	Google Scholar	α_1
Cornut et al. [2023]	Journal Article	eng	Google Scholar	α_3, β
Felsing et al. [2023]	Conference Paper	eng	Google Scholar	α_2, β
Llewellyn et al. [2023]	Conference Paper	eng	Zenodo	α_1
Manz et al. [2023]	Conference Paper	eng	Google Scholar	α_3, β
Petz [2023]	Journal Article	ger	Google Scholar	α_3
Raemy [2023a]	Report	eng	Google Scholar	α_2, β
Raemy and Gautschi [2023]	Conference Paper	fre	Google Scholar	α_1, β
Raemy et al. [2023b]	Conference Paper	eng	Google Scholar	α_1, β
Raemy and Sanderson [2023]	Pre-print	eng	Google Scholar	α_4, β
Tóth-Czifra et al. [2023]	Report	eng	Google Scholar	α_1, β
Vitale and Rainer [2023]	Book Chapter	eng	Google Scholar	α_1

Table 3.3: Summary of the LOUD Systematic Review

As I move on to explore IIF, WADM, and Linked Art in the following subsequent subsections, it is worth noting that these systems, as exemplified by [Sanderson \[2020b\]](#), embody and adhere to the foundational principles of LOUD, showcasing their practical applicability in diverse contexts.

3.5.3 International Image Interoperability Framework (IIIF)

In essence, the International Image Interoperability Framework (IIIF)¹⁴¹ serves as both a model for presenting and annotating content, as well as a global community that develops shared APIs, implements them in software, and exposes interoperable content.

To explore these aspects, this subsection on IIIF is divided into four parts. First, in subsubsection 3.5.3.1, I look at the genesis of the framework. Next, subsubsection 3.5.3.2 unravels the IIIF community. Moving forward, subsubsection 3.5.3.3 offers an overview of the existing and forthcoming APIs, extensions, and the development process of IIIF standards. Finally, in subsubsection 3.5.3.4, I outline the ecosystem, encompassing some of the compliant software, as well as the institutions and projects that have implemented IIIF specifications – sometimes referred to as the IIIF universe.

3.5.3.1 Inception

IIIF was initially established in 2011. Its birth emerged as a community-based initiative that crystallised from the convergence of two pivotal endeavours. One strand of this narrative revolved around the imperative to facilitate the seamless exchange of high-definition images over the internet [Snydman et al. 2015, p. 16]. This aspiration arose as a practical solution to mitigate the proliferation of duplicated images required for distinct projects. The desire to avert the necessity of sending substantial volumes of image data via conventional methods, such as mailing TB of data on hard drives, led to the contemplation of a web-based approach for sharing images that could break down silos.

In his talk during the 2019 IIIF Annual Conference, Cramer [2019] underscored the genesis of this facet of IIIF’s inception, also highlighting the ‘paper tablecloth story’¹⁴² where a plan was hatched informally in 2011 over dinner between technologists from the Stanford University Library, Oxford University and the British Library [Raemy 2017, p. 13] for distributing at scale image-based resources. As Emanuel [2018, p. 125] asserts: ‘*a clear, focused emphasis on collaboration and interoperability, between both institutions and technical approaches, can provide significant relief from this conundrum*¹⁴³, and can provide value for organisational members and technological end-users alike’.

The second strand, interwoven with the first, emanated from the explorations and experiments surrounding the interoperability of digitised medieval manuscripts. The Digital Manuscript Technical Working Group (DMSTech) at Stanford, operational from 2010 to 2013, provided the fertile ground for these reflections [Robineau 2019]. These deliberations ultimately coalesced into the formulation of the Shared Canvas Data Model¹⁴⁴, a model which employs a Linked Data approach based on Open Annotations and Open Archives Initiative Object Exchange and Reuse (OAI-ORE)¹⁴⁵ [see Sanderson et al. 2011] – a set of standards and protocols to facilitate the description and exchange of aggregations of web resources – to collaboratively describe digital facsimiles of physical objects, primarily in the CH domain where instances of the model are consumed by rendering platforms with a view to ‘*understanding the relationships between the constituent text, image, audio or other resources*’ [Sanderson and Albritton 2013]. The roots of this modelling effort can be traced back to Sanderson [2003]’s PhD thesis, where the first manuscript description was crafted based on an electronic edition of Froissart’s Chronicles, a prose history of the Hundred Years’ War written in the 14th century.

¹⁴¹ IIIF: <https://iiif.io> – the acronym is pronounced ‘triple-eye-eff’.

¹⁴² No, it wasn’t a napkin.

¹⁴³ The conundrum refers to the challenge of information and effort being compartmentalised within organisations or repositories.

¹⁴⁴ Shared Canvas Data Model: <https://iiif.io/api/model/shared-canvas/1.0/>

¹⁴⁵ OAI-ORE: <https://www.openarchives.org/ore/>

Robert Sanderson’s recollections, shared on the IIF Slack Workspace¹⁴⁶ in April 2022, provide an invaluable account of the early phases. The narrative unfolds in the mid-2000s when discussions between Tom Cramer, Stuart Snyderman (both at Stanford University at the time), and himself at conferences such as the DLF Forum commenced. This discourse centred on the challenges and prospects of interoperability for medieval manuscripts. Concurrently, Los Alamos National Laboratory (LANL) was engaged in developing Djatoka, an image server leveraging JPEG2000 and an OpenURL-based API¹⁴⁷ [Chute and Van De Sompel 2008]. Sanderson’s move to Los Alamos in 2008 marked a turning point, as the convergence of ideas and resources paved the way for tangible progress. Funding from the Mellon Foundation¹⁴⁸ catalysed the evolution of IIF. A parallel Mellon grant bolstered the development of Open Annotation, a precursor to the Web Annotation standard. This confluence set the stage for IIF’s embryonic phase. As the project progressed, collaboration expanded to include a wide range of partners. Lessons learned from manuscript interoperability were extrapolated and refined, culminating in pragmatic adjustments. The evolution included key refinements, including the adoption of JSON-LD as the sole manifest format and the establishment of a coherent API structure [Snyderman et al. 2015, p. 18].

Tracing the development of IIF, as highlighted in this timeline¹⁴⁹, reveals a sequence of pivotal milestones. Among these, the establishment of the Shared Canvas Data Model, a subject previously touched upon, stands out as a significant step. This timeline, while informative, is not all-encompassing but effectively underscores the essential moments shaping IIF. The journey from visionary dialogues to the concerted collaborative efforts showcases the collective ingenuity pivotal in enhancing digital image accessibility and interoperability. A notable milestone in this journey was the first officially listed IIF working meeting, which took place in Cambridge, England, in September 2011. This event symbolises the global commitment and the evolutionary stride of the IIF initiative.

Stanford University played a pivotal role in rebuilding the Mirador viewer¹⁵⁰, a key milestone that exemplifies the spirit of community-driven effort in the IIF initiative [Zundert 2018, p. 2]. This endeavour, later augmented through a collaborative partnership with Harvard University Library, highlights the cooperative dynamics and collective contributions that have been fundamental to the success and advancement of IIF. The Universal Viewer (UV)¹⁵¹ served as an additional client, validating IIF’s interoperability aspirations [Raemy 2017, p. 30]. The availability of the people behind OpenSeadragon (OSD)¹⁵², facilitated these advancements [Raemy and Sanderson 2023, p. 8].

The maturation of IIF was marked by incremental achievements, with Stanford’s active stewardship from 2014 to 2016 playing a pivotal role. The project’s multifaceted requirements found a harmonious synthesis within a supportive community. Financial resources, technical infrastructure, a coherent data model, and the collective dedication of engaged participants converged to transform IIF from an abstract concept into a tangible reality [Hadro 2019].

3.5.3.2 IIF Community

The IIF community stands as the cornerstone of the framework’s ongoing development and widespread adoption. It encompasses a broad spectrum of institutions and individuals. Their collaborative efforts

¹⁴⁶ IIF Slack: <https://iif.slack.com>

¹⁴⁷ OpenURL: https://developers.exlibrisgroup.com/sfx/apis/web_services/openurl/

¹⁴⁸ Grants allocated to Stanford University by the Mellon Foundation to support the development of IIF: <https://www.mellon.org/grant-details/international-image-interoperability-framework-9268> and <https://www.mellon.org/grant-details/lam-convergence-via-iif-shared-canvas-devconx-and-bumblebee-10319>

¹⁴⁹ See https://training.iif.io/time_machine/iif_intro/timeline.html

¹⁵⁰ Mirador: <https://projectmirador.org/>

¹⁵¹ UV: <https://universalviewer.io/>

¹⁵² OSD: <https://openseadragon.github.io/>

are pivotal in crafting, upholding, and advocating for the IIF specifications and tools. This community-led approach guarantees that IIF aligns with the dynamic requirements of CH entities and the wider DH sphere.

While the majority of IIF participants hail from North America, the United Kingdom, and Western Europe, the framework's reach is continually extending globally, with new adopters implementing IIF-compliant solutions worldwide [Raemy 2017, pp. 14-15]. Initially centred around GLAMs, IIF is increasingly garnering interest from various domains, including organisations in the STEM sector [see Kiley and Crane 2016, Moutsatsos 2017], such as delivering digital pathology data by leveraging the IIF Image API behind the scenes through an in-browser tiling mechanism [Bhawsar et al. 2023] or by converting tiles from storage instances compliant with the Digital Imaging and Communications in Medicine (DICOM)¹⁵³ standard [Jodogne 2023].

Within the IIF community, several dedicated groups focus on both community engagement and technical specifications¹⁵⁴. These groups, fundamental to the discourse and evolution of IIF-related matters, convene regularly and welcome participation from all interested parties [Hadro 2022].

The structure of these groups falls into two categories. Community Groups engage in discussions on themes and topics pertinent to their focus areas, often showcasing demos of related work and recent implementations. Current active groups include 3D, Archives, Audio/Video (A/V), Maps, Museums, and Outreach, with the Manuscripts, Design, and Newspapers groups currently on hiatus. TSGs are concentrated on collaborative efforts towards specific objectives related to the IIF APIs. Current active TSGs are the 3D TSG, Maps TSG, Authorization Flow, and Content Search. Furthermore, IIF has witnessed the successful completion of several TSGs. The Discovery TSG completed its work in 2023, leaving behind a number of specifications and deliverables. The Text Granularity TSG completed its mission in 2019 with the publication of the Text Granularity Extension. The A/V TSG concluded its work with the release of the Presentation API 3.0 in June 2020, and subsequently evolved into the A/V Community Group. Both group types thrive on an ethos of open community interaction, encouraging active involvement from all members of the IIF community. New groups can emerge in accordance with the IIF Groups Framework¹⁵⁵.

IIF has also orchestrated various events¹⁵⁶, including working group meetings, training sessions, online gatherings, and annual in-person conferences rotating amongst continents (at the moment, IIF conferences have only been held in either North America or Europe), to facilitate ongoing development and broaden institutional engagement.

The IIF Consortium (IIF-C)¹⁵⁷, established in 2015 by 11 founding institutions — including the initial three original institutions — aims to achieve several objectives to steer and sustain the IIF initiative [Cramer 2015]: providing uniform and rich access to digitised resources, defining and maintaining APIs that support interoperability, and developing shared technologies for an enhanced UX in handling digitised materials. Although originating within libraries, the IIF community has expanded to encompass museums, archives, CH aggregators, commercial entities, and technology companies, thereby fostering novel opportunities for interaction and collaboration across different sectors.

¹⁵³ DICOM: <https://www.dicomstandard.org/>

¹⁵⁴ IIF Groups: <https://iif.io/community/groups/>

¹⁵⁵ IIF Groups Framework: <https://iif.io/community/groups/framework/>

¹⁵⁶ Past IIF-Hosted Events: <https://iif.io/event/>

¹⁵⁷ IIF-C: <https://iif.io/community/consortium/>

The IIF community's structure also includes key committees for editing the specifications and coordinating community processes, each playing a vital role in the framework's governance and progress.

- Editorial Committee¹⁵⁸: This committee convenes on a regular basis, working closely with both group types. Their primary responsibility involves proposing new technical specifications and updating existing ones.
- IIF Coordinating Committee (CoCo)¹⁵⁹: This committee is instrumental in ensuring the IIF community advances in a sustainable, coordinated manner, driven by community consensus. Comprising both members and non-members of IIF-C, CoCo represents a broad spectrum of the community. It advises on various IIF activities, including policy development, formation of new groups, and maintaining the overall health of community groups.

Although originating within libraries, the IIF community has expanded to encompass museums, archives, commercial entities, and technology companies, thereby fostering novel opportunities for interaction and collaboration across different sectors. As of the current writing, the IIF-C consists of 67 members, illustrating the growing global interest and investment in the IIF initiative. The governance of the IIF-C is overseen by three key committees, each playing a distinct role in the consortium's operations:

- IIF Executive Committee (Exec): Tasked with steering the strategic direction and overseeing budgetary decisions, this committee convenes biannually. Initially, it comprised representatives from the first 11 founding members of the Consortium, augmented by two additional elected members. Since 2021, Exec positions are elected annually by the consortium membership, with terms being staggered.
- Operating Committee: This committee, drawn from staff of Exec, includes roles such as a Chair, Vice-Chair, and Treasurer, along with three at-large members. Meeting monthly, the Operating Committee works to ensure that the strategic directives set by Exec are effectively implemented.
- TRC¹⁶⁰: Composed of appointees from IIF-C member institutions with technical expertise, this committee convenes monthly. Their primary function is to vote on the approval or rejection of proposed amendments to the specifications, cookbook recipes, and TSG initiatives.

These committees, through their collective efforts, play a crucial role in the structured and effective governance of IIF-C. Their guidance and strategic decision-making are instrumental in steering the mission of IIF and ensuring the fulfilment of its overarching goals. This foundation of leadership and collaboration within the consortium sets the stage for the development and refinement of IIF specifications, the subject of the next subsection. As I transition to this topic, I will explore the technical backbone of the framework, highlighting how these specifications have become essential in advancing interoperability across CH platforms.

3.5.3.3 IIF Specifications

Currently, IIF has defined six APIs¹⁶¹. The most prominent are the Image and Presentation APIs, both updated to version 3 in June 2020, often referred to as the core IIF APIs. Complementing these are the Content Search and Authorization Flow APIs, both in version 2 (released in 2022 and 2023, respectively), and slated for future updates to align with the core APIs. In addition, the Change

¹⁵⁸ IIF Editorial Committee: <https://iif.io/community/editors/>

¹⁵⁹ IIF Coordinating Committee (CoCo): <https://iif.io/community/coordinating-committee/>

¹⁶⁰ IIF Technical Review Committee (TRC): <https://iif.io/community/trc/>

¹⁶¹ IIF Specifications: <https://iif.io/api/>

Discovery and Content State APIs, both in version 1.0, are relevant to the discovery and aggregation and/or sharing of IIIF resources. All (recent) specifications are working with JSON-LD 1.1., which is for instance consistent with i18n language¹⁶².

The development of the IIIF specifications is firmly rooted in the IIIF design principles [Appleby et al. 2018]. These principles, which have influenced similar methodologies such as the LOUD design principles, are articulated across 13 main components. They encapsulate a methodology that prioritises clear, practical goals, aligning with shared use cases to ensure the specifications meet the developer's needs effectively. This approach, focusing on simplicity, minimal barriers to entry, and adherence to good web and Linked Data practices, has been crucial in making the APIs both user-friendly and technically sound. Notably, the design principles advocate for ease of implementation and internationalisation, promoting a flexible yet cohesive structure. These principles do not directly express usability but rather represent objective constructs that naturally lead to usable and accessible data¹⁶³ [Raemy and Sanderson 2023, pp. 7-8].

The Editorial Committee of IIIF, responsible for writing these specifications, adheres to these principles and a specific process¹⁶⁴, ensuring that the resulting APIs are not only performant but also seamlessly integrate within existing technologies and standards. The IIIF Editorial Process encompasses two primary components: the Community Process and the Editorial Committee Process. The Community Process involves open discussions on IIIF-Discuss¹⁶⁵, GitHub¹⁶⁶, and various group meetings for proposals and feedback. Essential to this process is the demonstration of use cases by at least two institutions, rigorous evaluation and testing of new features, and a thorough community review overseen by TRC. Specifications are subject to continuous revision with a release frequency guided by semantic versioning, balancing community input with the need for stability.

The editors make use of web best practices, including terminologies as defined in Request for Comments (RFC) 2119 [see Bradner 1997] of the Internet Engineering Task Force (IETF), a standards organisation mainly responsible for the Internet protocol suite. Additionally, the process involves clear documentation of editorial participation, with a structured approach for suggesting changes primarily via GitHub. Changes undergo strict acceptance criteria, requiring consensus among editors before merging. Regular meetings, both in-person and virtual, are crucial for editors to discuss ongoing work and coordinate efforts. This dual-component process ensures thoughtful, inclusive, and responsive evolution of IIIF specifications, catering to the diverse needs of the community.

In the following paragraphs I will undertake an exploration of each of the IIIF APIs. Special emphasis will be placed on the core APIs due to their integral role in the IIIF ecosystem. While providing a rich insight into the Image and Presentation APIs, I will also discuss the particularities of the Content Search, Authorization Flow, Change Discovery, and Content State APIs, highlighting their contributions to the overall breadth of the IIIF specifications.

¹⁶² i18n: <https://jsonforms.io/docs/i18n/> – ‘The term internationalisation is often represented as i18n, where 18 is the number of letters between the word’s opening “i” and closing “n”’ [Sheldon 2023].

¹⁶³ For instance, if all IIIF APIs try to avoid dependencies on technology, per the fourth IIIF design principle, there are still individuals and institutions from the wider community that have done research on the current and best options in terms of image formats (JPEG2000 versus pyramidal TIFF), types of conversion, storage, and optimal software [see Cossu 2019a, Gomez et al. 2020, Robson et al. 2023, Rosenthaler 2023].

¹⁶⁴ IIIF Editorial Process: <https://iiif.io/community/policy/editorial/>

¹⁶⁵ IIIF-Discuss (mailing list): <https://groups.google.com/g/iiif-discuss>

¹⁶⁶ IIIF on GitHub: <https://github.com/iiif>

The IIIF Image API, defined as a RESTful web service, responds to standard HTTPS requests with an image [Appleby et al. 2020a], in other words, it is an agreement between an image client and an image server to get the pixels [Robson 2021].

The API uses its own URI syntax and offers two primary modes of interaction – both of which are hereafter exemplified with the syntax and an URL working with the specification – so that users can easily pan, deep zoom, manipulate, cite, and share regions of interest of an image.

The two modes of interaction are as follows:

Image Request URI Syntax Retrieving an image derived from the underlying image content. The parameters define the characteristics of the returned image, all separated by a slash.

`scheme://server/prefix/identifier/region/size/rotation/quality.format`

<https://iiif.dasch.swiss/0801/3eIbRgj63JP-CrShE6LBcTB.jpg/full/max/0/default.jpg>

Image Information Request URI Syntax Requesting details about the image service, including characteristics and available functionality. It is serialised as a JSON-LD document.

`scheme://server/prefix/identifier/info.json`

<https://iiif.dasch.swiss/0801/3eIbRgj63JP-CrShE6LBcTB.jpg/info.json>

Figure 3.23 shows the sequence of parameters of the URI syntax that needs to be interpreted by compliant software (**region** then **size** then **rotation** then **quality** then **format**), as well as showcasing different manipulations that are available. In this case, the image has been cropped to 125, 15, 120, 140 as its coordinate, has a width of 90 pixels, is mirrored and then rotated by 345 degrees, and returned in a grey – or rather ‘gray’ according to the specification – quality. This would give the following URL: <https://example.org/image42/125,15,120,140/90,/!345/gray.jpg>

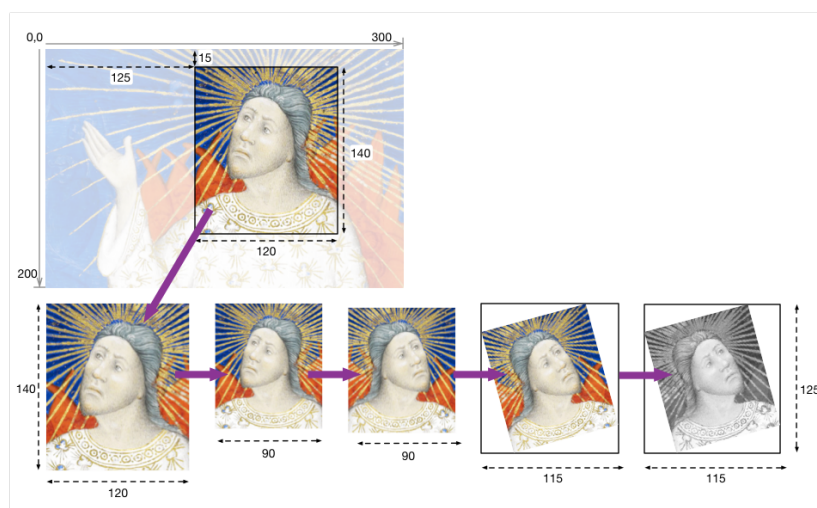


Figure 3.23: IIIF Image API Order of Implementation [Appleby et al. 2020a]

Furthermore, it is mandatory for the image information document to delineate the extent of API support, i.e. the set of parameters and features that are supported by the image service, by denoting the compliance level as the value of the profile property. This compliance level must align with one of the three levels (0, 1, or 2) outlined in the Image API Compliance document¹⁶⁷.

¹⁶⁷ IIIF Image API Compliance, Version 3: <https://iiif.io/api/image/3.0/compliance/>

The chosen compliance level should represent the highest level that meets all the requirements. Some of these go beyond image manipulation and include HTTP features such as server response behaviours and Cross-Origin Resource Sharing (CORS), which is a mechanism that allows or restricts resources on a web page to be requested from another domain outside the domain from which the first resource was served. CORS is a protocol that is particularly relevant for IIIF services, as it involves accessing and displaying images hosted on different servers or domains, necessitating careful management to ensure seamless access while maintaining web security protocols [see van Kesteren 2023]. To assist in achieving or verifying compliance with the IIIF Image API, an image service can refer to the Image API Validator¹⁶⁸, a tool designed to assist service providers in achieving compliance with the specification.

The result is that the IIIF Image API provides a well-defined image retrieval format that is easy to implement. It provides sufficient detail to drive zoomable image viewers. The URLs are designed to be intuitive, allowing easy manipulation by users. In addition, these URLs are optimised for caching and scalability, improving performance and efficiency.

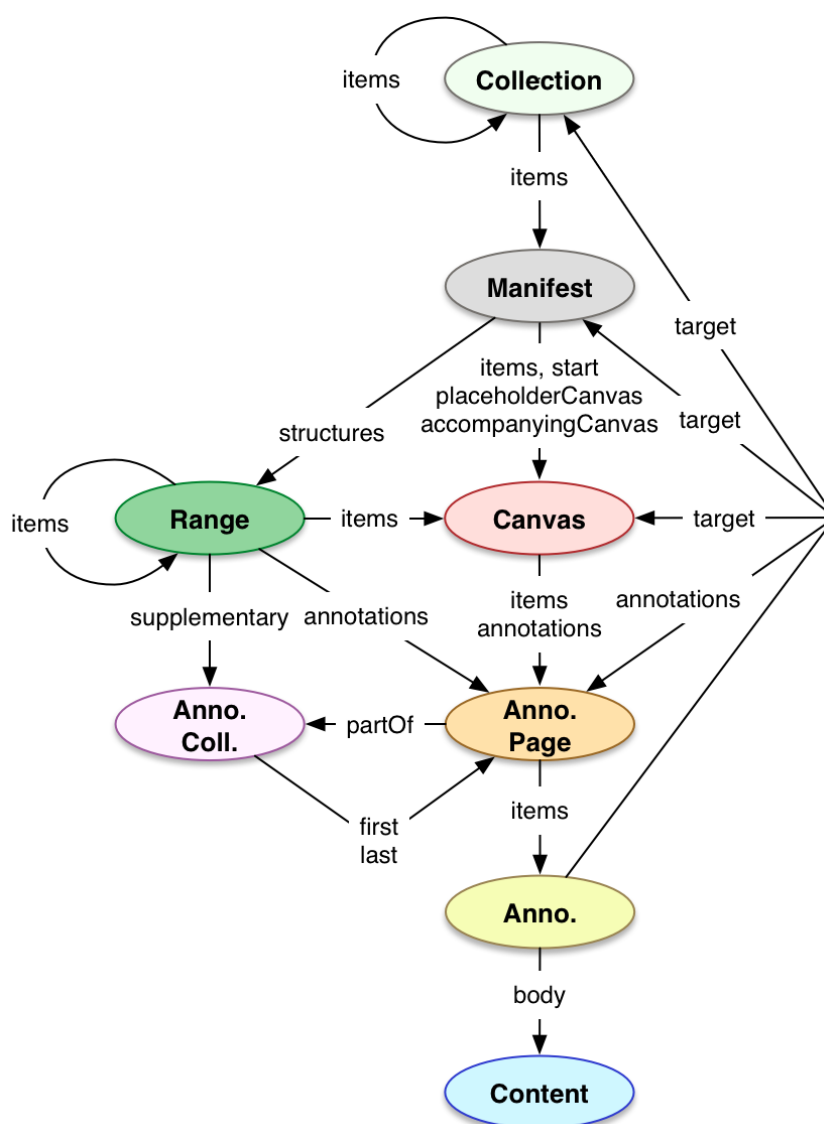


Figure 3.24: IIIF Presentation API Data Model [Appleby et al. 2020b]

¹⁶⁸ IIIF Image API Validator: <https://iiif.io/api/image/validator/>

As for the IIIF Presentation API, based on JSON-LD, it provides the necessary information about the structure and layout of objects or collections to drive a remote viewing experience. As such, its primary purpose is to display human-readable descriptive information, i.e. some minimal descriptive and legal metadata, rather than offering semantic metadata for search engines¹⁶⁹ [Appleby et al. 2020b]. Loosely based on the Shared Canvas Data Model, it incorporates four main resource types: `Collection`, `Manifest`, `Canvas`, and `Range`, each serving distinct functions within the API framework. In addition, the specification makes use of types that are defined by WADM, such as `AnnotationCollection`, `AnnotationPage`, `Annotation`, and `Content` (see subsection 3.5.4). A diagram of the Presentation API data model is shown in Figure 3.24.

The IIIF Manifest plays a key role as it encapsulates the structure and attributes of a compound object. It contains essential data required for a client application to accurately display the content to users. Typically, each Manifest is dedicated to delineating the presentation of a single complex entity, which could be diverse in nature, ranging from a book or a sculpture to a musical album. The Manifest thus serves as a crucial intermediary, translating complex object properties into a format that is both accessible and comprehensible to users engaging with digital representations of these objects. In the context of the Presentation API, the Manifest is perhaps one of the most important resource types, acting as the cornerstone for how digital content is structured and presented within the IIIF ecosystem. Code Snippet 3.10 shows an example of a basic Manifest from the cookbook recipes¹⁷⁰, which are published by the community to encourage publishers to adopt common patterns when modelling IIIF resources. To provide a clearer understanding, the following is a detailed breakdown of the structure of this IIIF Manifest.

@context This defines the context for interpreting the JSON-LD document. In this case, it points to the IIIF Presentation API 3 context.

id This is the unique identifier of the Manifest, pointing to its location on the web.

type Specifies that this JSON-LD document represents a IIIF Manifest.

label It provides a human-readable label for the manifest in different languages. Here, it is labelled in English as ‘Picture of Göttingen taken during the 2019 IIIF Conference’.

items An array of items within the Manifest. This Manifest contains one **item**, a `Canvas`.

Canvas Represents a single canvas that can contain one or more images or content resources. It includes its own **id**, **type**, **label**, **height**, and **width**. The **items** array within the `Canvas` contains an `AnnotationPage`.

AnnotationPage Represents a page within the canvas that can contain annotations. It has its own **id**, **type**, and contains an array of **items**, each an `Annotation`.

Annotation An annotation associates a region of an image with descriptive or interpretive information. It has its own **id**, **type**, **motivation** (here as ‘painting’), and **body** which contains detailed information about the image being annotated, including **id**, **type**, **format**, **height**, **width**, and **service**.

service It contains information about the IIIF Image Service for the image.

target It specifies the target `Canvas` to which the annotation is associated.

¹⁶⁹ However, to provide such semantic metadata, IIIF adopters can leverage the `rdfs:seeAlso` property and point to structured metadata for aggregation and discovery purposes [Raemy 2020, pp. 15-16]. A registry of profiles is available for this specific use case: <https://iiif.io/api/registry/profiles/>.

¹⁷⁰ IIIF Cookbook: <https://iiif.io/api/cookbook/>

Code Snippet 3.10: Example of a IIIF Manifest

```

1 {
2   "@context": "http://iiif.io/api/presentation/3/context.json",
3   "id": "https://iiif.io/api/cookbook/recipe/0005-image-service/manifest.json",
4   "type": "Manifest",
5   "label": {
6     "en": [
7       "Picture of Göttingen taken during the 2019 IIIF Conference"
8     ]
9   },
10  "items": [
11    {
12      "id": "https://iiif.io/api/cookbook/recipe/0005-image-service/canvas/p1",
13      "type": "Canvas",
14      "label": {
15        "en": [
16          "Canvas with a single IIIF image"
17        ]
18      },
19      "height": 3024,
20      "width": 4032,
21      "items": [
22        {
23          "id": "https://iiif.io/api/cookbook/recipe/0005-image-service/page/p1/1",
24          "type": "AnnotationPage",
25          "items": [
26            {
27              "id": "https://iiif.io/api/cookbook/recipe/0005-image-service/annotation/p0001-image",
28              "type": "Annotation",
29              "motivation": "painting",
30              "body": {
31                "id": "https://iiif.io/api/image/3.0/example/reference/918
32                  ecd18c2592080851777620de9bcb5-göttingen/full/max/0/default.jpg",
33                "type": "Image",
34                "format": "image/jpeg",
35                "height": 3024,
36                "width": 4032,
37                "service": [
38                  {
39                    "id": "https://iiif.io/api/image/3.0/example/reference/918
40                      ecd18c2592080851777620de9bcb5-göttingen",
41                    "profile": "level1",
42                    "type": "ImageService3"
43                  }
44                ]
45              },
46              "target": "https://iiif.io/api/cookbook/recipe/0005-image-service/canvas/p1"
47            }
48          ]
49        }
50      ]
51    }
52  ]
53 }

```

In summary, this IIIF Manifest describes a Canvas with a single IIIF image, and the image is associated with an **Annotation** that provides additional information about the image. The Manifest also includes metadata such as a label and unique identifiers for various components.

In addition to the core functionalities, the IIIF Presentation API is augmented by three formally approved extensions:

navPlace Extension Enhancing the IIIF Presentation API with the `navPlace` property, defined using GeoJSON-LD¹⁷¹ for geographic coordinates.

Text Granularity Extension This extension offers a standard for indicating text granularity levels in annotations, especially useful for OCR software, manual transcriptions, or existing digitised text.

¹⁷¹ GeoJSON-LD: <https://geojson.org/geojson-ld/>

Georeference Extension Leveraging WADM, it provides a pattern for georeferencing IIIF Presentation 3 API **Canvases** and images from the IIIF Image API.

IIIF Presentation API resources such as Manifest and Collection can be validated against a dedicated Presentation API Validator¹⁷². This tool ensures that these resources conform to the specification, guaranteeing their correctness within the framework.

If Version 3.0 of the Presentation API was released only a few years ago, the need for a specification that can formally display and disseminate 3D objects has been widely recognised across various domains. Addressing this demand, ongoing work on an upgraded version of the IIIF Presentation API (Version 4.0) to accommodate 3D resources holds great promise [Haynes 2023]. While particularly relevant to the CH sector, as highlighted by Raemy and Gautschy [2023] and Manz et al. [2023] where conventional dissemination approaches struggle to effectively convey the depth and complexity of 3D artefacts, there are specific challenges beyond visual representation. These challenges include accurately representing the lighting conditions that affect the object’s appearance and capturing the spatial orientation, such as handedness (defining the coordinate system: where does the coordinate x , y , z have to be and in which direction). Such enhancements of the forthcoming IIIF Presentation API have the potential to empower institutions and industries beyond CH.

Certain key features have already been firmly established, particularly in regard to the integration of 3D content. A new resource type known as a **Scene** will be introduced to better accommodate 3D content and sits besides Canvas in the updated data model. This **Scene** is characterised by an infinite boundless space, where the origin (0,0,0) is at the centre rather than on the top-left corner as it is the case with still images, and is structured around a right-handed coordinate system. Within this system, movement along the Y-axis is interpreted as upward, while movement along the Z-axis signifies forward progression. Importantly, **Canvases** can be effectively nested within these **Scenes**, and likewise, **Scenes** have the capacity to be nested within other **Scenes** [Haynes et al. 2023]. As Rossenova [2023] importantly argues, ‘the 3D specification would need to ensure that viewers can both implement a “minimum viable product” for user interaction and remain interoperable, as well as include more complex interactions and still specify those in a standards-compliant way’. This advancement could potentially enable the presentation of collections in immersive formats, transforming access, engagement and cross-platform interoperability.

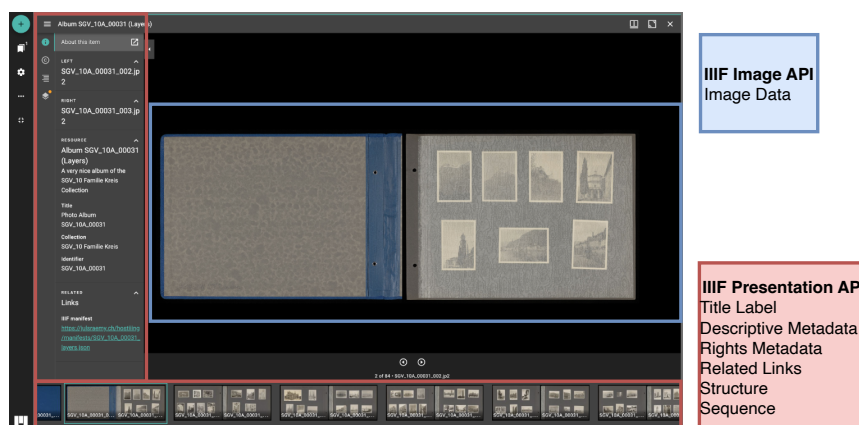


Figure 3.25: The Interaction of the Core IIIF APIs in Mirador

¹⁷² IIIF Presentation API Validator: <https://presentation-validator.iiif.io/>

Figure 3.25 illustrates the integration of the core APIs within Mirador. Elements highlighted in blue are rendered via the Image API, while those in red are facilitated by the Presentation API. It is important to note that the Image API can function independently or in conjunction with the Presentation API for additional viewing capabilities. Additionally, there is no direct counterpart to the Image API for A/V content; instead, A/V specifics are conveyed using the `duration` property of the Presentation API.

Shifting focus to the additional IIIF APIs, the Content Search API plays a vital role in enhancing the framework's functionality. The specification facilitates searching within textual annotations and provides the mechanism for performing these searches within the IIIF context [Appleby et al. 2022a]. To add a Content Search service to a IIIF Manifest, an URI must be provided through the `service` property. Then the client, such as Mirador or the UV, will use it to create a valid query. An auto-complete service can also be embedded, further enriching the UX by streamlining the search process.

The Authorization Flow API outlines four mechanisms for accessing restricted content to either initiate an interaction with an access control system or to give a client enough information about the user's state with respect to the content provider. These mechanisms are: access service, token service, probe service, as well as logout service. The access service typically manifests as a user interface provided by the content host, which the client opens in a new tab, usually leading to the provider's login page. The token service plays a crucial role wherein the client acquires an access token, necessitating presentation of the same authorising credentials to the token service as the browser would to the related access-controlled resource. The probe service, as defined in this specification, is an endpoint used by the client to ascertain the user's association with the access-controlled resource it represents. This is applicable for both access-controlled content resources and image information documents (`info.json`) that produce access-controlled image responses. Additionally, there is an optional logout service, designed to facilitate user logout [Appleby et al. 2023].

The Change Discovery API is designed for machine-to-machine interfaces, primarily serving IIIF-aware systems. It enables content providers to efficiently publish lists of links to their content, which facilitates easy discovery and synchronisation for consuming systems [Appleby et al. 2021]. The API is based on the W3C AS standard, ensuring a consistent and universally understandable pattern for publishing content changes. The API designates IIIF Collections and Manifests as primary access points for published content. Activities describing changes to these resources, and potentially to other types like IIIF Image API endpoints.

Central to the API's utility is its ability to allow content providers to articulate the specifics of how their content has evolved. This includes providing details on content that has been deleted or is no longer available. This optimisation significantly helps consuming systems in retrieving only those resources that have been modified since their last update, thereby streamlining the update process. The specification defines three distinct levels of conformance, each offering a different depth of information about the changes to IIIF resources

Level 0 It offers a straightforward list of resources along with their URIs, focusing on basic interoperability.

Level 1 It enhances the list by adding timestamps and ordering, allowing consuming applications to process changes more efficiently.

Level 2 It provides comprehensive information about the types of activities undertaken, such as the creation, modification, and deletion of resources, thus giving a detailed insight into the content changes. Code Snippet 3.11 shows an example of a Level 2 Activity.

Code Snippet 3.11: Example of a Level 2 Activity Compliant with the IIIF Change Discovery API [Appleby et al. 2021]

```
1 {
2   "type": "Create",
3   "object": {
4     "id": "https://example.org/iiif/1/manifest",
5     "type": "Manifest"
6   },
7   "endTime": "2017-09-20T00:00:00Z"
8 }
```

In the Change Discovery API, the `rdfs:seeAlso` property from the Presentation API is a vital feature that enhances the discoverability and understanding of IIIF resources [Raemy 2020, p. 16]. This property is used to reference external documents containing richer and more detailed information about the content being presented. The specification also provides a structured approach for the organisation and processing of `OrderedCollection` and `OrderedCollectionPage`. This includes clear directives on how activities within these structures are managed and accessed.

In terms of network considerations, the Change Discovery API addresses critical aspects such as handling media types and activities for access-restricted content. These considerations ensure that the network interactions remain efficient and reliable, even in scenarios where content accessibility is dynamically changing. Additionally, the IIIF community deployed a IIIF Registry¹⁷³ to give access to IIIF resources and streams that are compatible with the Change Discovery API.

Finally, the Content State API enables deep-linking into objects and annotations from search results, aiding users in sharing specific IIIF content [Appleby et al. 2022b]. Specifically, a content state is a JSON-LD data structure that can be leveraged by IIIF clients to load and present a particular part of a resource, such as region of a Canvas in a Manifest, playing a particular point in a recording, displaying multiple targets from a comparison view, and embedding search results. Because the content states are serialised in JSON, these must be encoded in the `base64url` format, a variant of the `base64` encoding designed to be safe for use in URLs as it avoids using special characters [see Josefsson 2006]. In summary, as Crane [2020] argues, the specification is ‘*how we point at things in IIIF*’.

Figure 3.26 displays a IIIF Manifest, specifically of *The Life of Buddha, first book (Shaka no Honji, jō)*, hosted by e-codices¹⁷⁴. This resource is accessed through Jalava¹⁷⁵, a directory tree-based interface that aggregates IIIF-compliant repositories. It also includes a custom viewer, which enables the sharing of resources based on the IIIF Content State API in JSON and `base64url` formats.

Collectively, these APIs, extensions and validators are not only fundamental components that enable the framework, they are also anchored by a set of well-defined design principles and an editorial process. This combination ensures improved interoperability and utility across different digital platforms, while maintaining consistent and usable standards that reflect the evolving needs of the community.

In order to effectively deploy IIIF, it is important for the system building blocks to have a comprehensive understanding of the different APIs involved. For a basic IIIF solution, the images need to be delivered in compliance with the Image APIs through a web service. Then, a series of scripts can

¹⁷³ IIIF Registry: <https://registry.iiif.io/>

¹⁷⁴ Cologny, Fondation Martin Bodmer, Cod. Bodmer 600a: *The Life of Buddha, first book (Shaka no Honji, jō)*. <https://doi.org/10.5076/e-codices-fmb-cb-0600a>

¹⁷⁵ Durham University’s Jalava: <https://iiif.durham.ac.uk/jalava/universe.html>

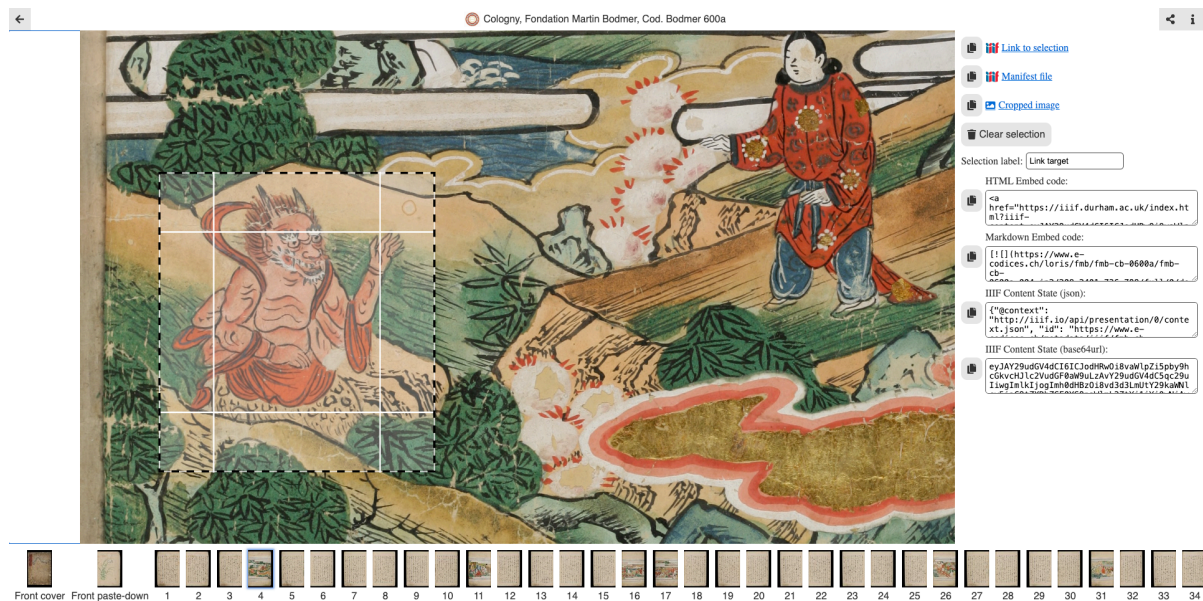


Figure 3.26: *Cologne, Fondation Martin Bodmer, Cod. Bodmer 600a: The Life of Buddha, First Book (Shaka no Honji, jō)* Embedded in Jalava, a Customised Client from Durham University Compliant with the IIIF Content State API

be used to generate IIIF resources, which combine the relevant metadata and data structure. These can be displayed within IIIF-compliant viewers. Deploying a more advanced IIIF architecture that can manage and save annotations, restrict access to certain objects, inform the IIIF community of changes made within its own collection, or search for OCR text requires a more complex environment. This involves deploying multiple types of servers and different libraries. Figure 3.27¹⁷⁶ illustrates an example of a IIIF architecture with the six APIs.

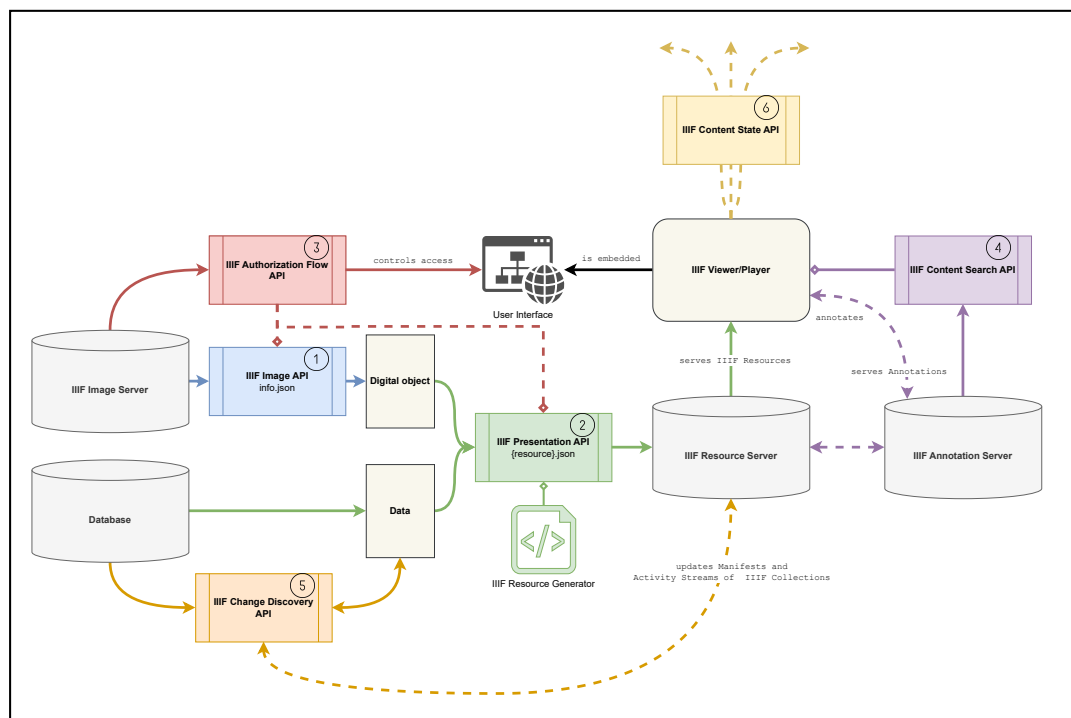


Figure 3.27: Example of a IIIF Architecture Implementing the Six APIs

¹⁷⁶ This figure is also available in an enlarged format in Figure B.4.

As I move from the underlying technical structure to its practical application, the next subsection takes a closer look at the IIIF ecosystem. This includes a diverse range of compliant software and implementations, each contributing to the widespread adoption and versatility of IIIF. By exploring these elements, it is possible to gain a comprehensive understanding of how IIIF is being operationalised and integrated into various environments.

3.5.3.4 IIIF Ecosystem

This segment explores the composition of the IIIF ecosystem, particularly focusing on how its suite of shared APIs has been integrated into software and the institutions that have made their resources compliant with the specifications.

The adoption of IIIF by diverse organisations marks a significant turning point plays a crucial role in dismantling silos. This fosters enhanced communication and interoperability between various software platforms. As a result, users gain access to a richer array of options for viewing and interacting with digital resources. This transformation, catalysed by the widespread adoption of IIIF, is illustrated in Figure 3.28, showcasing the dynamic and interconnected nature of the IIIF ecosystem.

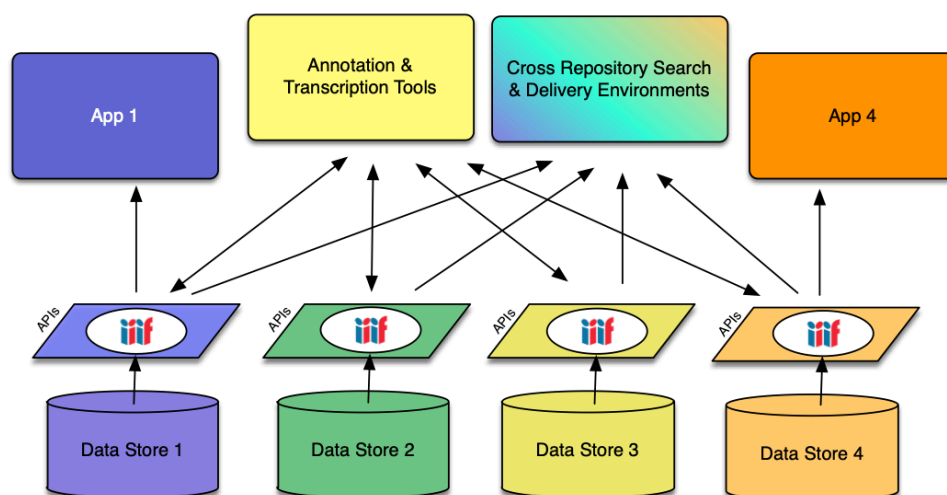


Figure 3.28: Example of a IIIF Ecosystem of Software

The suite of IIIF-compliant software encompasses a diverse range of tools, each catering to specific aspects of the framework. Central to this suite are image servers, predominantly tailored to align with the IIIF Image API, such as IIPImage¹⁷⁷, Cantaloupe¹⁷⁸, or SIPI¹⁷⁹. Complementing these servers are image viewers, designed to be compatible with either the Image API alone, like OSD and Leaflet-IIIF¹⁸⁰, or with both of the core IIIF APIs. With the release of the version 3.0 of the Presentation API, A/V players have also been developed, which frequently comply with the Image API as well. Examples of known clients that support the Image and Presentations APIs 3.0 are Mirador, the UV, Annona¹⁸¹,

¹⁷⁷ IIPImage: <https://iipimage.sourceforge.io/>

¹⁷⁸ Cantaloupe: <https://cantaloupe-project.github.io/>

¹⁷⁹ SIPI: <https://sipi.io/>

¹⁸⁰ Leaflet-IIIF: <https://github.com/mejackreed/Leaflet-IIIF>

¹⁸¹ Annona: <https://ncsu-libraries.github.io/annonna/multistoryboard/>

Clover¹⁸², and Ramp¹⁸³. As the support and behaviour vary, a dedicated viewer matrix¹⁸⁴ was created by Glen Robson, the IIIF-C Technical Coordinator.

Additionally, the ecosystem includes an array of tools and libraries which offer pre-written code segments essential for constructing and manipulating IIIF resources. For narrative and website creation, there are specialised exhibition and guided viewing tools such as Storiies¹⁸⁵ and Exhibit¹⁸⁶, adding a storytelling dimension to the experience. Moreover, validators play a critical role in ensuring compliance and accuracy, while annotation servers and clients offer advanced functionalities for enriching digital resources with contextual and interpretive information. Together, these components form a comprehensive and versatile toolkit, enabling a wide range of applications and enhancing the utility of IIIF resources.

The evolving software landscape has witnessed a notable expansion in IIIF compatibility, where a larger number of software applications now align with the IIIF specifications. This shift is evident in the growing support for one or more of the IIIF APIs, with a particular emphasis on the Image API or both core APIs [Rabun 2016]. In May 2016, recognising the need to provide guidance and support to both newcomers and software developers within the IIIF community, a curated collection titled ‘Awesome IIIF’¹⁸⁷ was initiated. This resource compiles a comprehensive array of IIIF-compliant tools, tutorials, and descriptive links, organised to facilitate easy navigation and understanding of the IIIF ecosystem. It serves as an invaluable resource, offering insights and directions for those embarking on their journey with IIIF, as well as helping developers in finding the necessary tools and information to integrate IIIF specifications effectively into their projects.

IIIF has been embraced by a variety of institutions worldwide, spanning from state and national libraries to aggregators, museums, archives, and universities. Examples include Austria’s national library, the British Library, BnF, Israel’s national library, aggregators like Europeana, the Digital Public Library of America, and the Wikimedia Foundation, museums such as the Victoria & Albert Museum, the alters Art Museum, and the YCBA, archives such as the Swiss Federal Archives, the Blavatnik Foundation Archive, and the National Film and Sound Archive of Australia, as well as academic institutions such as Leiden University, University of Edinburgh, and Université Paris Cité. These implementations showcase the versatility of IIIF, illustrating its ability to cater to a wide spectrum of CH needs.

In Switzerland, IIIF has seen significant uptake, primarily within the library and academic sectors. Institutions like DaSCH, e-manuscripta, e-rara, Kunstmuseum Basel, the Swiss Federal Archives, and the University of Geneva have all adopted IIIF to some extent. Notably, both core IIIF APIs, in their Version 2.0, saw their initial implementation by e-codices¹⁸⁸, the Virtual Manuscript Library of Switzerland, in December 2014 [Raemy 2017, p. 24]. It marked a strategic shift towards integrating modern open-source technologies, such as OSD, in the overhaul of their digital platform [Fritschi 2017, pp. 245-246]. This early adoption not only demonstrated e-codices’ commitment to using the latest state-of-the-art digital tools, but also paved the way for the wider adoption of IIIF standards in digital libraries.

¹⁸² Clover: <https://samvera-labs.github.io/clover-iiif/>

¹⁸³ Ramp: <https://iiif-react-media-player.netlify.app/>

¹⁸⁴ IIIF 3.0 Viewer Matrix: <https://iiif.io/api/cookbook/recipe/matrix/>

¹⁸⁵ Storiies: <https://www.cogapp.com/storiies>

¹⁸⁶ Exhibit: <https://www.exhibit.so/>

¹⁸⁷ Awesome IIIF: <https://github.com/IIIF/awesome-iiif>

¹⁸⁸ e-codices: <https://www.e-codices.ch/>

Scoping the whole so-called IIIF universe is complex as there is no mandatory requirement – and there probably shouldn't be one – to tell the IIIF community that you have implemented the APIs or that you have developed tools that are compliant with the specifications. In terms of specifications, mostly the core IIIF APIs are implemented, such as shown in different investigations [see [Raemy 2023b](#), [Raemy and Schneider 2019b](#)]. The IIIF community has also carried out several surveys, targeted at institutions implementing the standards in 2017, 2020, and 2023 [[Raemy 2023a](#), p. 2]. In 2017, the number of images compatible with the IIIF Image API was estimated to be over 335 million [[Raemy 2017](#), p. 27], probably under-representing the real number. At the time of writing, it is likely that there are over 2 billion images compliant with the Image API, as the Internet Archive alone hosts over 1 billion of them.

The latest 2023 IIIF Implementation Survey, receiving 80 responses, showcased a diverse range of institutions including 26 libraries, 11 museums (some encompassing combinations like museum and research library), 16 universities or research institutions, seven archives, five service providers, as well as four software providers. It is quite representative to what is known to IIIF-C and the wider community, with participation and implementations mostly from libraries, museums, and universities from North America and Europe. Most surveyed institutions currently support IIIF, with a few in the planning stage. The number of IIIF-compatible resources varies widely, ranging from fewer than a thousand to over a million in some cases. Popular IIIF-compliant image servers include Cantaloupe and IIIFImage. Mirador and UV are commonly used viewers. JPEG2000 and TIFF are the predominant image formats used across the surveyed institutions. A portion of the institutions actively contribute staff time or funding to shared IIIF tools and projects, highlighting a collaborative approach within the IIIF community.

Navigating the world of IIIF resources can be a challenging endeavour, given their vast and varied nature. To assist in this task, a comprehensive guide has been developed, providing valuable insights and directions on how to locate IIIF resources effectively¹⁸⁹. As an illustration, Figure 3.29 presents the entry of the Cultural Japan platform¹⁹⁰, which aggregates IIIF-compliant of and about Japan [see [Machiya et al. 2023](#)].

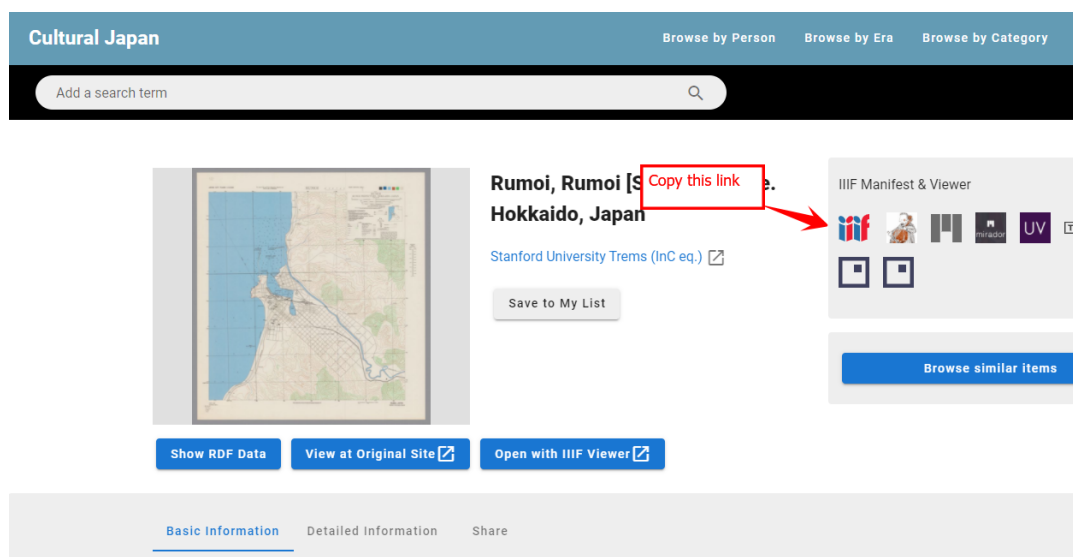


Figure 3.29: Screenshot of the Cultural Japan Platform in the IIIF Guides

¹⁸⁹ Guides to finding IIIF resources: https://iiif.io/guides/finding_resources/

¹⁹⁰ Cultural Japan: <https://cultural.jp/>

While the IIF ecosystem is dynamic and evolving, it is important to note that its expansion and development are predominantly observed within GLAM and academic institutions, mostly in wealthy countries. This skew indicates that while IIF specifications have gained traction and proving to be influential in the DH and CH domains, their adoption is not uniformly distributed globally. The concentration of IIF tools and implementations in certain regions underscores the need for broader outreach and support to enable more diverse and widespread adoption. Despite its current geographical and sectoral concentration, the framework's capability to enhance access, interaction, and preservation of digital resources remains significant, offering a promising avenue for the future of digital archival work on an international scale.

3.5.4 Web Annotation Data Model

The Web Annotation Data Model (WADM)¹⁹¹, a W3C standard established in 2017 and influenced by the outcomes of the Open Annotation Collaboration (OAC)¹⁹² resulting from several years of effort [see Haslhofer et al. 2011], provides an extensible and interoperable framework for creating and sharing annotations across various platforms.

WADM is designed to support both simple and complex use cases, such as linking content to specific data points or multimedia resources. It standardises annotations for interoperability and easy integration into existing collections, using JSON-LD for serialisation, which integrates annotations into the web's structured data ecosystem. The model adopts Linked Data principles, emphasising interoperable and flexible structures for annotations [Sanderson et al. 2013].

Annotations in WADM link resources, referred to as **body** and **target**. The **target** is the resource being annotated, while the **body** contains the annotation content, which can be text, images, or other media. This structure allows a single annotation to relate to multiple resources, ensuring compatibility with web architecture and facilitating cross-platform sharing [Sanderson et al. 2017]. Figure 3.30 illustrates this three-part model.

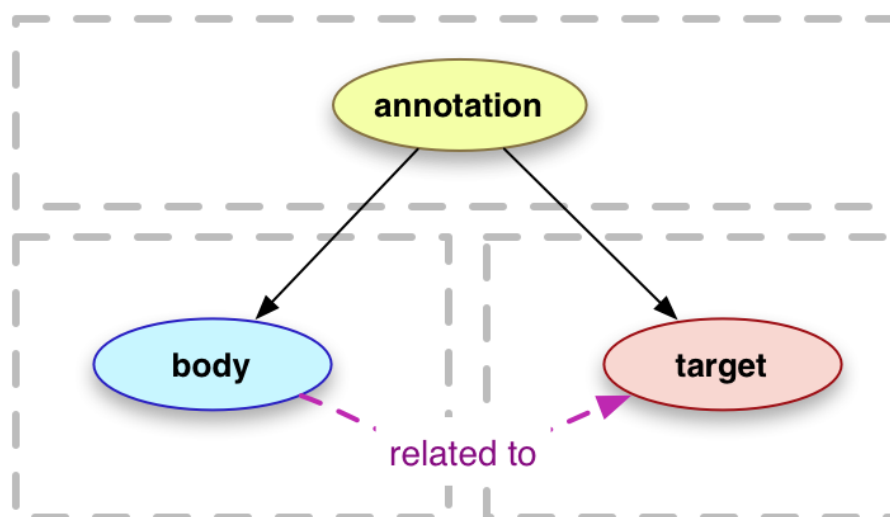


Figure 3.30: WADM High-level Overview [Sanderson et al. 2017]

The model outlines various components of an annotation [Sanderson et al. 2017]:

¹⁹¹ WADM: <https://www.w3.org/TR/annotation-model/>

¹⁹² OAC specified the Open Annotation Data Model, a precursor of WADM: <https://web.archive.org/web/20230328030315/http://www.openannotation.org/spec/core/>

@context A property defining the JSON's meaning as an **Annotation**. Each **Annotation** must include one or more **@context** values, with <http://www.w3.org/ns/anno.jsonld> as a mandatory inclusion. If only one value is present, it should be a string.

id A property identifying the **Annotation** with a unique IRI. Each **Annotation** must have exactly one identifying IRI.

type A relationship specifying the **Annotation**'s type. Each **Annotation** should have one or more types, including the **Annotation** class.

Annotation A class representing web annotations. Each **Annotation** must be specified with **type**.

body A relationship linking the **Annotation** to its body. An **Annotation** can have one or more body relationships, but it is possible to have none.

target A relationship linking the **Annotation** to its target. Each **Annotation** must have one or more target relationships.

Code Snippet 3.12 is an example of a basic annotation.

Code Snippet 3.12: Example of Basic Annotation According to WADM [Sanderson et al. 2017]

```
1 {
2   "@context": "http://www.w3.org/ns/anno.jsonld",
3   "id": "http://example.org/anno1",
4   "type": "Annotation",
5   "body": "http://example.org/post1",
6   "target": "http://example.com/page1"
7 }
```

Annotations can link to different types of resources, including external web resources, embedded text, or specific segments of a resource. They can also include lifecycle information, agent details, intended audience, accessibility features, motivations, rights information and other identities.

Code Snippet 3.13 shows an example of an annotation with a motivation and a purpose.

Code Snippet 3.13: Example of an Annotation According to WADM with a motivation and a purpose [Sanderson et al. 2017]

```
1 {
2   "@context": "http://www.w3.org/ns/anno.jsonld",
3   "id": "http://example.org/anno15",
4   "type": "Annotation",
5   "motivation": "bookmarking",
6   "body": [
7     {
8       "type": "TextualBody",
9       "value": "readme",
10      "purpose": "tagging"
11    },
12    {
13      "type": "TextualBody",
14      "value": "A good description of the topic that bears further investigation",
15      "purpose": "describing"
16    }
17  ],
18   "target": "http://example.com/page1"
19 }
```


Annotations can be organised into collections or pages for easy reference and management. An `AnnotationCollection` manages a potentially large number of annotations, maintaining its own descriptive and creation information, and includes references to the first page of annotations. Each `AnnotationPage` contains an ordered list of some or all annotations within the collection.

WADM has played a key role in several areas, particularly in extending the functionality of the IIF Presentation API data model. Beyond the IIF space, the influence of WADM extends into other areas, demonstrating its versatility and wide applicability.

In the field of knowledge management, [Rossenova et al. 2022, p. 10] uses WADM to annotate 3D content within an instance of Wikibase¹⁹³, a suite of KB software used by Wikidata for managing LOD. Additionally, WADM finds application in the field of musicology, where it aids in the annotation of musical notes, thus enriching the study and understanding of music through a digital lens [Weigl et al. 2021, pp. 26-27]. The importance of WADM has been also important in the domain of data quality assessment [Wei et al. 2016, p. 24]. Moreover, the VR gaming industry has also embraced WADM to enrich the generated KGs [Rousi et al. 2021, pp. 100-101]. This application demonstrates the standard's potential in creating immersive and interactive digital experiences.

Incorporating all the discussed aspects of WADM, it is evident that this model offers a comprehensive and flexible framework for web annotations. The WADM's capacity to handle diverse annotation types, from simple text to complex multimedia segments, and its consideration of aspects like resource lifecycle, rights information, and accessibility, make it a robust tool for web-based annotations. Envisioning WADM-driven storytelling, it could serve as a potent means to convey *hypotypes* — picturesque descriptions of scenes or events. This aligns with the concept of multimedia assemblage of archival objects, as discussed by Bachimont [2021] and revisited in the context of web annotations¹⁹⁴, highlighting the potential of WADM in transforming the presentation and interpretation of content into dynamic, contemporary narratives.

In contrast to IIF and Linked Art, WADM stands out as the only recognised LOUD standard that doesn't concurrently function as an active community. In the event that updates to WADM are required, a dedicated W3C working group would need to be re-established or revamped¹⁹⁵. This necessity contrasts with the community-driven evolution seen in IIF and Linked Art, where ongoing community engagement and collaboration play a crucial role in their development and updating processes.

3.5.5 Linked Art

Linked Art¹⁹⁶ is a community-driven initiative working together to define a LOUD specification for describing CH, primarily artworks, designed to operate across CHIs, facilitating the publication and use of a common body of knowledge. As such, Linked Art presents itself as '*providing a standards based metadata profile, which consistently solves problems from real data, is designed for usability and ease of implementation, which are prerequisites for sustainability*' [Sanderson 2023c]. Specifically, the community develops and maintains a shared data model, or profile, based on CIDOC-CRM as well as a JSON-LD-based API to interact with the data [Cornut et al. 2023, p. 10].

¹⁹³ Wikibase: <https://wikiba.se/>

¹⁹⁴ The idea for such formulation is not mine and can be credited to Kembellec [2023] on his tweet about Bruno Bachimont and IIF. I have revisited this thought to discuss Web Annotations.

¹⁹⁵ The Web Annotation Working Group was chartered from August 2014 to October 2016, and was extended to February 2017 when WADM was officially vetted as a standard.

See <https://www.w3.org/groups/wg/annotation/>.

¹⁹⁶ Linked Art: <https://linked.art>

Linked Art presents a layered framework that distinguishes between the conceptual and implementation aspects of its model. This stratification is key to understanding how the Linked Art profile and its API are situated within a broader context of shared abstractions and sustainable implementations [Sanderson 2020a].

Shared Abstraction Level The conceptual model is encoded by the ontology, which is refined by vocabularies.

Conceptual Model This represents an abstract way of conceptualising the world in a holistic, consistent, and coherent manner. A conceptual model can express just about anything anyone wishes to document. Linked Art is based on CIDOC-CRM, an event-based framework.

Ontology This involves a shared set of terms that encode the conceptual model in a logical and machine-actionable way. This layer aims for the correctness of data. Linked Art uses the RDF encoding of CIDOC-CRM 7.1.3, with a few other extensions like CRMdig and properties and classes created by the Linked Art community, as well as other commonly used RDF namespaces such as Dublin Core and RDFS for disambiguating closely related property names.

Vocabulary A curated set of terms specific to sub-domains, making the ontology more concrete and applicable. The Getty Vocabularies are used in Linked Art to further classify the main core entities from CIDOC-CRM.

Sustainable Implementations Level The shared abstraction level is specialised by a profile and available through an API.

Profile It entails a selection of appropriate abstractions from the conceptual model, aimed at encoding the scope of what can be described. In the case of Linked Art, its profile is mainly focused on describing artworks.

API This involves a selection of appropriate technologies to provide access to data managed using the profile. Like the other LOUD specifications, the Linked Art API is serialised in JSON-LD and follows the REST architectural principles.

Figure 3.31 illustrates these layers, delineating the transition from shared abstractions (in blue) to their sustainable implementations (in green) in the Linked Art ecosystem.

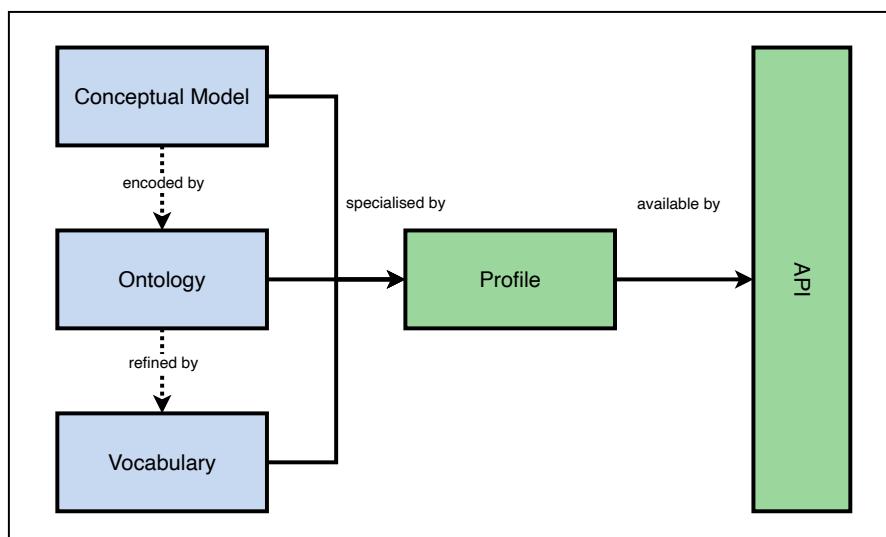


Figure 3.31: Abstraction and Implementation Level of Linked Art. Adapted from Sanderson [2021]

Embarking on an exploration of Linked Art, this subsection unfolds in four distinct yet interconnected segments. First, subsection 3.5.5.1 offers a historical perspective on the development and evolution of Linked Art. This is followed by subsection 3.5.5.2, which provides an in-depth examination overview of the model, detailing its structure and patterns. Subsequently, subsection 3.5.5.3 explores the architectural and design principles of the Linked Art API, including its protocol, core constructs, and the array of entity endpoints it recommends. The subsection culminates with subsection 3.5.5.4, where I look at the community dynamics surrounding Linked Art and how the specification has been, or will be, implemented.

3.5.5.1 History

Linked Art officially came into existence in 2017 [Raemy 2023a, pp. 1-2]. However, its roots can be traced back to November 2016 with the creation of its first GitHub repository¹⁹⁷, marking the beginning of the community. The initiative gained significant momentum in January 2019, when it commenced hosting open community calls, inviting broader participation and engagement¹⁹⁸.

In the same year, Linked Art received financial support in the form of grants from both the Kress Foundation¹⁹⁹, known for its dedication to advancing the field of art history, and the Arts and Humanities Research Council (AHRC)²⁰⁰, a major funding body in the United Kingdom. This support from two esteemed organisations has strengthened Linked Art’s capabilities and resources, allowing it to expand its scope and organise five face-to-face meetings; three in 2019 and two in 2023 – delayed due to the COVID-19 pandemic.

The official recognition of Linked Art as a CIDOC Working Group in 2020 marks another milestone in its development. This recognition not only confirms the importance and relevance of Linked Art within the international museum sector, but also increases its visibility.

The evolution of Linked Art was significantly influenced by two key precedents: the American Art Collaborative (AAC)²⁰¹ and Pharos²⁰², the International Consortium of Photo Archives. AAC, a consortium of fourteen American art museums, played a critical role in promoting the adoption of agreed-upon LOD practices for artworks [Knoblock et al. 2017]. On the other hand, Pharos brought together a network of photographic archives [Delmas-Glass and Sanderson 2020].

Fink [2018, p. 35] highlights the philosophical alignment between Linked Art and AAC. She notes that ‘*[Linked Art] describes the philosophy that shaped the AAC target model. It indicates the model will be updated by applying it to other data sets, such as the Getty Museum and Pharos, The International Consortium of Photo Archives, with the intention of having it serve as a resource for the broad museum community*’. While AAC’s alignment with Linked Art’s philosophy is evident, Pharos’ engagement with the community has followed a distinct path. Initially closely aligned with Linked Art, Pharos eventually branched off, adopting unique modelling practices and decided to deploy their own instance of ResearchSpace²⁰³, an open source semantic web platform, to aggregate and align data from partner institutions.

¹⁹⁷ Linked Art on GitHub: <https://github.com/linked-art/linked.art>

¹⁹⁸ Linked Art Call, 8am Pacific, 11am Eastern on 2019-01-30 – Message on the Linked Art Google Group by Robert Sanderson: <https://groups.google.com/g/linked-art/c/OCoz0lEdpYY>

¹⁹⁹ Kress Foundation: <https://www.kressfoundation.org/>

²⁰⁰ AHRC: <https://www.ukri.org/councils/ahrc/>

²⁰¹ AAC: <https://americanart.si.edu/about/american-art-collaborative>

²⁰² Pharos: <http://pharosartresearch.org/>

²⁰³ ResearchSpace: <https://researchspace.org/>

The creation of AAC's dates back to 2014, and in 2016 it published its target model based on CIDOC-CRM. This model embodied a design philosophy similar to that of Linked Art, characterised by its adaptability to different museum infrastructures. The well-structured and strategic design of the AAC target model laid a solid foundation for the development of functional applications and influenced the approach and methodology adopted by Linked Art in its subsequent evolution.

Linked Art's development and growth are also significantly intertwined with the trajectory of IIF [see Daga et al. 2022]. This connection is evident in the way Linked Art has emulated IIF's principles and methodologies. In addition, a few core people collaborate in both initiatives [Raemy 2023a, p. 11].

The release of the Linked Art API V1.0, scheduled for the course of December 2024, represents a significant turning point in the evolution of the standard, but it is worth noting that several earlier implementations had already laid the groundwork for its adoption. These implementations not only provided valuable insights into the data model and the API's potential but also served as a catalyst for its refinement and enhancement.

3.5.5.2 Linked Art Data Model

At its core, Linked Art is a data model²⁰⁴ or metadata application profile²⁰⁵ that draws extensively from the RDF implementation of the version 7.1.3 of CIDOC-CRM [see Bekiari et al. 2024]. The Getty Vocabularies – namely AAT, ULAN, and TGN – are leveraged as core sources of identity for domain-specific terminology. JSON-LD 1.1 is chosen as the preferred serialisation format, promoting clarity and interoperability. This framework constructs common patterns, integrating conceptual models, ontologies, and vocabulary terms. These elements are derived from real-world scenarios and contributions from the diverse participants and institutions within the Linked Art community [Sanderson 2019a].

In the area of data provenance, the work of Ram and Liu [2009] on the W7 model stands as an example. This ontological framework conceptualises data provenance through seven interconnected facets: 'what', 'when', 'where', 'how', 'who', 'which', and 'why'. Each of these elements serves as a means to monitor and understand events impacting data throughout its existence. Importantly, the W7 model demonstrates flexibility and adaptability in capturing the nuances of provenance for data across diverse domains. The Linked Art data model, in its domain-specific application, particularly resonates with five of these facets: 'what', 'where', 'who', 'how', and 'when' [Sanderson 2020b]. This focus facilitates a nuanced tracking and interpretation of the key aspects of data provenance, as visually captured in Figure 3.32²⁰⁶.

²⁰⁴ Linked Art Data Model: <https://linked.art/model/>

²⁰⁵ Linked Art Profile of CIDOC-CRM: <https://linked.art/model/profile/>

²⁰⁶ Notes on colours for data visualisation purposes: brown is used for physical objects, purple for digital objects, green for place, yellow and orange for conceptual entities, blue for activity, teal for timespan, and red for agents.

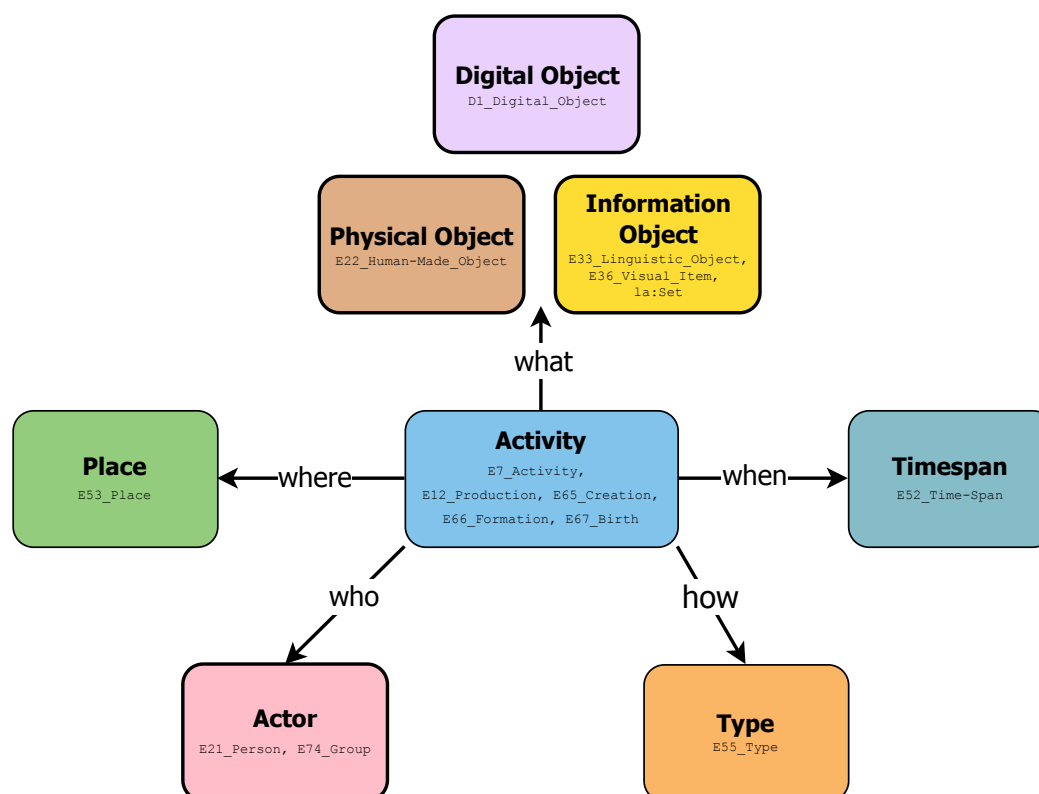


Figure 3.32: Linked Art from 50,000 Feet [Raemy et al. 2023a, adapted from Sanderson [2020b]]

There are a few base patterns²⁰⁷ that every resource should have for it to be a useful part of the world of Linked Data [Raemy and Sanderson 2023, p. 11]. The core properties, types and classifications are listed hereafter.

@context It contains a reference to the context mapping. It is a property of the document pointing to <https://linked.art/ns/v1/linked-art.json>.

id It captures the URI that identifies the object. Every resource must have exactly one **id**, and it must be an HTTP URI.

type It captures the class of the object, or `rdf:type` in RDF. Every resource must have exactly one class. This allows software to align the data model with an internal, object oriented class based implementation.

_label It captures a human readable label as a string, intended for developers or other people reading the data to understand what they are looking at. Every resource should have exactly one label, and must not have more than one. It is just a string, and does not have a language associated with it – if multiple languages are available for the content, then implementations can choose which is most likely to be valuable for a developer looking at the data.

classified.as CIDOC-CRM needs to be extended for effective application. To facilitate this, Linked Art employs the `classified.as` property, which is designed to link to a term within a controlled vocabulary, offering a mechanism for expansion and customisation. Specifically, it is `P2_has_type` with a different mapping in the `@context` document to avoid the collision with `rdf:type`.

²⁰⁷ Linked Art Baseline Patterns: <https://linked.art/model/base/>

The JSON-LD serialisation displayed in Code Snippet 3.14 represents a practical application of the properties previously discussed, exemplifying the modelling of Rembrandt's *The Night Watch*²⁰⁸. This particular use case was not only a central focus during the Linked Art face-to-face meeting held in Amsterdam in October 2023 [Sanderson 2023c] but also serves as an illustrative example in various core scenario modellings highlighted here. This extract and the ones that follow offer a glimpse of how Linked Art effectively captures artwork metadata and presents it in a structured format.

Code Snippet 3.14: The Night Watch: Representation in the Linked Art Data Model Including URI; Class; Label; and Classification

```

1 {
2   "@context": "https://linked.art/ns/v1/linked-art.json",
3   "id": "https://example.org/object/42",
4   "type": "HumanMadeObject",
5   "_label": "The Night Watch",
6   "classified_as": [
7     {
8       "id": "http://vocab.getty.edu/aat/300033618",
9       "type": "Type",
10      "_label": "Painting"
11    },
12    {
13      "id": "http://vocab.getty.edu/aat/300133025",
14      "type": "Type",
15      "_label": "Work of Art"
16    }
  ]
}
```

Linked Art categorises controlled vocabulary terms²⁰⁹ into three distinct classes to facilitate validation and interoperability. This classification is key to ensuring that data adheres to consistent rules across different implementations.

Required These terms must be used to be considered valid, as they form the bedrock of classification.

Recommended Terms in this category are strongly suggested for use. They contribute significantly to the quality and coherence of data. While their usage is encouraged, there may be valid reasons for opting out.

Listed These terms are put forward to promote uniformity and wider adoption of common practices. While their use is not mandatory, employing them aligns with the established practices of the Linked Art community.

Code Snippet 3.15 shows how names, identifiers and statements are represented in Linked Art. Names, distinct from the `_label` property which serves as internal documentation, are essential for user-facing resources. Every entity that should be visible to an end-user, be it an object, person, group, or event, is recommended to have at least one specific name. This is achieved using the `identified_by` property with a `Name` resource, where the actual name is specified in the `content` property. Identifiers are handled similarly but employ the `Identifier` class. They are often classified to distinguish between various types, like internal system numbers and accession numbers. This classification helps in differentiating and understanding the origins and nature of each identifier. Statements come into play where data does not support the specificity that the full ontology allows or when information is best conveyed in human-readable form. This could include descriptions like medium or materials, rights or usage statements, dimensions, or edition statements. In scenarios where there is descriptive text about

²⁰⁸ Van Rijn, Rembrandt (1642). *De Nachtwacht* [Oil on canvas]. Amsterdam Museum on permanent loan to Rijksmuseum, Amsterdam, The Netherlands. <http://hdl.handle.net/10934/RM0001.COLLECT.5216>

²⁰⁹ Vocabulary Terms in Linked Art: <https://linked.art/model/vocab/>

the resource, regardless of the type of description, the `referred_to_by` property is used to record this textual information.

Code Snippet 3.15: The Night Watch: Representation in the Linked Art Data Model Including Names; Identifiers; and Statements

```
1 {
2   "@context": "https://linked.art/ns/v1/linked-art.json",
3   "id": "https://example.org/object/42",
4   "type": "HumanMadeObject",
5   "_label": "The Night Watch",
6   (...)
7   "identified_by": [
8     {
9       "type": "Name",
10      "_label": "The Night Watch (English)",
11      "content": "The Night Watch"
12      "classified_as": [
13        {
14          "id": "http://vocab.getty.edu/aat/300404670",
15          "type": "Type",
16          "_label": "Primary Name"
17        }
18      ],
19      "language": [
20        {
21          "id": "http://vocab.getty.edu/aat/300388277",
22          "type": "Language",
23          "_label": "English"
24        }
25      ]
26    },
27    {
28      "type": "Identifier",
29      "_label": "Night Watch Object Identifier",
30      "content": "SK-C-5"
31      "classified_as": [
32        {
33          "id": "http://vocab.getty.edu/aat/300404621",
34          "type": "Type",
35          "_label": "Repository number"
36        }
37      ]
38    }
39  ],
40  "referred_to_by": [
41    {
42      "type": "LinguisticObject",
43      "classified_as": [
44        {
45          "id": "http://vocab.getty.edu/aat/300435429",
46          "type": "Type",
47          "_label": "Material Statement",
48          "classified_as": [
49            {
50              "id": "http://vocab.getty.edu/aat/300418049",
51              "type": "Type",
52              "_label": "Brief Text"
53            }
54          ]
55        }
56      ],
57      "content": "Oil on Canvas",
58      "language": [
59        {
60          "id": "http://vocab.getty.edu/aat/300388277",
61          "type": "Language",
62          "_label": "English"
63        }
64      ]
65    }
66  ]
67 }
```

The use of an intermediate **Activity** entity, rather than direct links, allows for richer descriptions and associations. This approach enables linking multiple actors, places, techniques, and timespans to a single activity, enhancing the depth of cultural heritage data representation. Here are the following properties:

carried_out_by It identifies who carried out the activity, linking it to people or groups.

took_place_at It specifies where the activity occurred, associating it with specific locations.

technique It describes how the activity was carried out, detailing methods or techniques used.

Furthermore, Linked Art incorporates a minimal **TimeSpan** model for activities, utilising properties **begin_of_the_begin** and **end_of_the_end**. These properties record the beginning and end of the **TimeSpan**, respectively. Both activities, all encapsulated within the **produced_by** property, and timespans are exemplified in Code Snippet 3.16.

Code Snippet 3.16: The Night Watch: Representation in the Linked Art Data Model Including Activities and Timespans

```

1 {
2   "@context": "https://linked.art/ns/v1/linked-art.json",
3   "id": "https://example.org/object/42",
4   "type": "HumanMadeObject",
5   "_label": "The Night Watch",
6   (...)
7   "produced_by": {
8     "type": "Production",
9     "carried_out_by": [
10      {
11        "id": "http://vocab.getty.edu/ulan/500011051",
12        "type": "Person",
13        "_label": "Rembrandt, Harmensz van Rijn"
14      }
15    ],
16    "took_place_at": [
17      {
18        "id": "http://vocab.getty.edu/tgn/7006952",
19        "type": "Place",
20        "_label": "Amsterdam"
21      }
22    ],
23    "technique": [
24      {
25        "id": "http://vocab.getty.edu/aat/300053343",
26        "type": "Type",
27        "_label": "Painting"
28      }
29    ],
30    "timespan": {
31      "type": "TimeSpan",
32      "begin_of_the_begin": "1642-01-01",
33      "end_of_the_end": "1642-12-31",
34      "identified_by": [
35        {
36          "type": "Name",
37          "content": "1642",
38          "classified_as": [
39            {
40              "id": "http://vocab.getty.edu/aat/300404669",
41              "type": "Type",
42              "_label": "Display Label"
43            }
44          ]
45        }
46      ]
47    }
48  }
49 }

```

In essence, Linked Art's basic patterns revolve around a few principles: URIs for web identification of entities and records, a small yet impactful set of classes (ontology) coupled with an extensive array

of classifications (vocabulary). Core elements such as names, identifiers, statements, and classifications are universally applicable to every entity. The concepts of activities and partitioning enable connections and detailed specificity between entities.

Within the data model, further key patterns have been recognised and established. These include object descriptions, people and organisations, places, digital integration (such as describing IIIF-compliant resources and pointing to their JSON-LD API context – as shown in Figure 3.33²¹⁰), provenance of objects, collections and sets, exhibitions of objects, primary sources of information, assertion level metadata, and dataset level metadata [Raemy and Sanderson 2023, p. 12]. This classification provides a robust framework to distinguish between the physicality of objects, the abstract nature of works, and the human or collective agents involved.

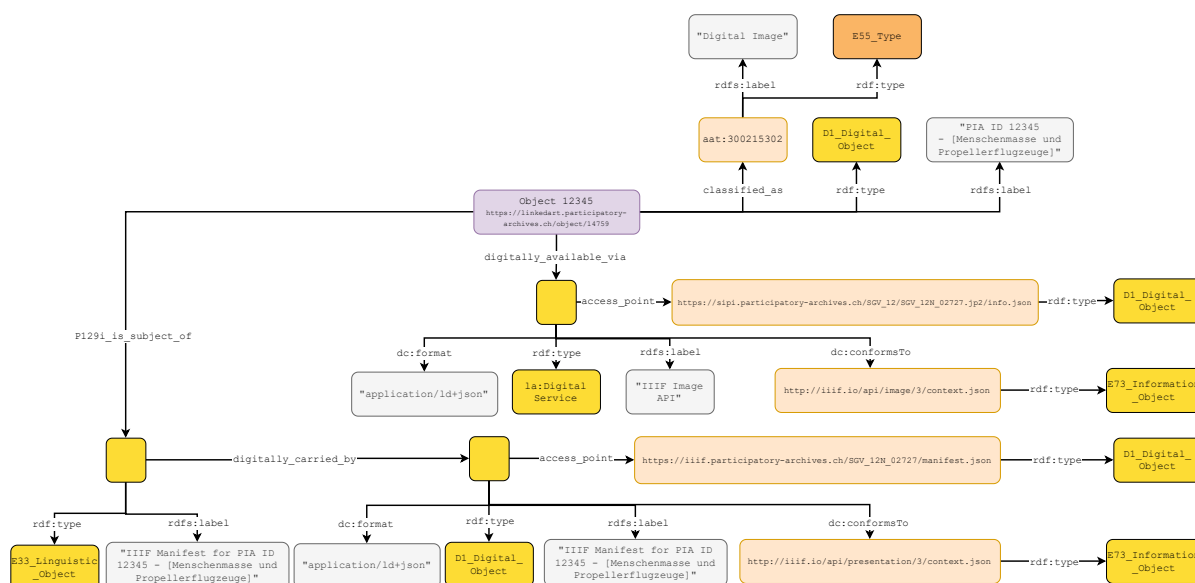


Figure 3.33: Linked Art Digital Integration: Possible Description of Available IIIF resources

Particularly, objects are differentiated as **HumanMadeObject**²¹¹, physical entities with tangible presence, and **DigitalObject**, representing digital files. Works include **PropositionalObject** (abstract concepts not tied to text or visuals), **LinguisticObject** (textual works like ‘Lord of the Rings’), and **VisualItem** (visual works, such as ‘The Night Watch’). Actors are classified as **Person** (individuals with intentional action) or **Group** (collectives).

Linked Art extends the notion of actors to non-humans who are responsible, albeit temporarily, for activities under the category of persons. **Place** entities are defined as fixed geographical locations with identifiable coordinates. For concepts, **Type** represents broad categories or classifications, supplemented by specific concepts such as **Language**, **Material**, **Currency** and **MeasurementUnit**. The **Set** class is defined as an unordered group that is unique to Linked Art, as opposed to **E78 Curated Holding** in CIDOC-CRM, which emphasises physical and purposefully preserved sets.

Furthermore, **Activity** in Linked Art, especially **Provenance** and **Exhibition**, offer explicit connections between entities, setting Linked Art apart from other data models. Here, objects are distinct

²¹⁰ This figure is also available in an enlarged format in Figure B.5.

²¹¹ The CIDOC-CRM namespaces from classes and properties, their prefix numbers, inconsistencies (*is_*, *was_*, etc.), as well as most dashes are removed in Linked Art. **CamelCase** is also used as common practice for classes.

from works and provide a context, while people, groups, and places are entities in their own right [Sanderson 2023c].

Each of these patterns has an essential function in the structured representation and organisation of a wide range of data. This includes data about artworks, the artists who created them, the places associated with those artworks, digital representations, historical background, as well as information about collections, exhibitions and the rich metadata that supports this rich tapestry of CH [Raemy and Sanderson 2023, p. 12].

3.5.5.3 Linked Art API and Standards for Linked Art

In Linked Art, a clear distinction exists between the model and the API. The model is inherently flexible, allowing for a wide range of representations, while the API is tailored for software developers, providing a more defined and structured interaction mechanism. In this section I will look at the characteristics of the Linked Art API and other specifications relevant to Linked Art, such as JSON Hypertext Application Language (HAL) or AS.

Linked Art, drawing inspiration from IIF, espouses a set of design principles²¹² aimed at fostering usability and interoperability [Sanderson 2023b]. These principles emphasise the importance of defining scope through shared use cases and designing for internal utility, while adhering to the axiom of simplicity. They advocate for making easy tasks straightforward and complex tasks achievable, without over-reliance on specific technologies. Embracing REST and leveraging the benefits of web caching, these guidelines are grounded in RESTful architecture and the efficient use of network resources. As such, the Linked Art API revolves around the use of URIs as both identifiers and locators. It is also important not to infer information about the publisher based on the URI structure and to prefer HTTPS for security. Additionally, the principles prioritise designing for JSON-LD, employing LOD methodologies, and adhering to established standards and best practices wherever feasible. A key focus is on defining criteria for success to facilitate extensibility, thereby ensuring a robust and adaptable framework for Linked Art's implementation.

Further delving into the specifics of the Linked Art API in terms of requirements, four key areas are listed: trivial to implement, consistency across representations, division of information, and URI requirements [Raemy and Sanderson 2023, p. 13].

Linked Art advocates for trivial implementation. The framework is designed such that even hand-crafted files on disk are feasible for deployment, though automation may be preferred for managing larger data volumes. A key aspect is the consistency across representations, ensuring that each relationship is contained within a single document. This approach aids in maintaining clarity and coherence in data structuring. Moreover, Linked Art emphasises the division of information across different representations, adopting a structured approach that scales from many to few. For example, in the context of a book, the structure would flow from the page level to the book, and eventually to the collection, delineating a clear hierarchy of information. Additionally, the identity and URI requirements in Linked Art are strategically designed. One-to-one relationships are embedded directly and do not necessitate separate URIs. This simplifies the data model and enhances its accessibility. Furthermore, the URIs for records are devoid of any internal structure, adhering to a principle of simplicity and straightforward usability. This approach ensures that the model remains user-friendly and easily navigable, making it an efficient tool for representing and managing CH data.

²¹² Linked Art API Design Principles and Requirements: <https://linked.art/api/1.0/principles/>

The division of the graph in Linked Art ensures no duplication of definitions across records. Full URIs are used for references, simplifying client processing. Embedded structures, even if potentially identifiable, do not carry URIs to avoid unnecessary complexity. At the moment, there are eleven endpoints²¹³ in the Linked Art API, loosely based on the conceptual model presented previously:

1. **Concepts:** Types, Materials, Languages, and others, as full records rather than external references
2. **Digital Objects:** Images, services and other digital objects
3. **Events:** Events and other non-specific activities that are related but not part of other entities
4. **Groups:** Groups and Organisations
5. **People:** Individuals
6. **Physical Objects:** Physical things, including artworks, buildings or other architecture, books, parts of objects, and more
7. **Places:** Geographic places
8. **Provenance Activities:** The various events that take place during the history of a physical thing
9. **Sets:** Sets, including Collections and sets of objects used for exhibitions
10. **Textual Works:** Texts worthy of description as distinct entities, such as the content carried by a book or journal article
11. **Visual Works:** Image content worthy of description as distinct entities, such as the image shown by a painting or drawing

Each of these endpoints is accompanied by detailed documentation outlining the required and permitted patterns, complemented by a corresponding JSON schema²¹⁴.

Linked Art has investigated to use of both HAL, an IETF RFC, and AS for managing back-links and harvesting content. HAL links, which are media types for representing resources and their relations with hyperlinks [see Kelly 2023], are simple and reliable, with existing tooling including validation. It should play a fundamental role in addressing redundancy for Linked Art by mitigating response bloating between API responses. They are placed at the top level of JSON-LD documents, not treated as properties but as part of the API [Raemy and Sanderson 2023, p. 17]²¹⁵. AS, already implemented by the IIIF Change Discovery API, provides a common paging model across various standards, allowing for flexible data aggregation.

Overall, APIs for the Linked Art prioritise JSON for syntax, appealing to the software developer audience. By implementing Linked Art records in JSON-LD and using HAL links for back-links, along with HAL for search and aggregation, the specifications offer consistency, usability, and ease of implementation without the need for specialised technologies, yet retaining semantic richness. Essentially, implementing those specifications means adopting the model itself.

²¹³ Linked Art API Endpoints: <https://linked.art/api/1.0/endpoint/>

²¹⁴ Linked Art Schema Definitions: https://linked.art/api/1.0/schema_docs/

²¹⁵ Linked Art HAL Links: <https://linked.art/api/rels/1/>

3.5.5.4 Community and Implementations

The Linked Art community²¹⁶ boasts a diverse and international membership primarily from museums and universities in North America and Europe. Key institutions such as The National Gallery of Art (USA), The J. Paul Getty Trust, the Museum of Modern Art, the Frick, the Rijksmuseum, and the Victoria and Albert Museum, alongside universities and research centres like Oxford, Yale, the Foundation for Research & Technology – Hellas (FORTH), DHLab, and ETH Zurich, are among the earlier and active participating institution in the community. Other organisations such as Europeana and the American Numismatics Society have also been involved.

Linked Art has seen significant project-based support and investigation, notably from the University of Oxford e-Research Centre²¹⁷. This includes initiatives like the ‘Linked Art: Networking Digital Collections and Scholarship’²¹⁸ and ‘Linked Art II: Developing Community, Practice, and Scholarship’²¹⁹, both supported by AHRC funding. These projects exemplify the collaborative and research-focused ethos at the heart of the Linked Art initiative. In the area of collaborative projects, Linked Art lists partners such as Pre-Raphaelites Online²²⁰ and Linked Conservation Data²²¹ on its website. These partnerships highlight the diverse applications of Linked Art across different domains, from art history to conservation practices, underlining its adaptability and relevance to a wide range of CH endeavours.

From a library and archives perspective, the LD4 Art & Design Affinity Group in the USA²²², extending beyond museums to encompass libraries and archives, has listed Linked Art as a relevant initiative emphasising its cross-disciplinary appeal.

Several institutions have already adopted various implementations of Linked Art, each tailoring it to their specific needs and due to the release of Version 1.0 being delayed. Yale’s LUX Collection Discovery platform (see Chapter 8) stands out as a flagship implementation [Raemy 2023a, p. 26]. Additionally, digital platforms like The Art and Life of Georgia O’Keeffe, the Getty Museum Collection, the Getty Vocabularies, Van Gogh Worldwide, and the Rijksmuseum have either implemented or are planning to implement Linked Art.

However, the adoption of Linked Data in European museums and beyond remains limited, with few institutions displaying digital objects online. For instance, France and Forsberg [2021] argued that structuring data internally using the Linked Art model might be premature due to its ongoing development. This highlights the importance of establishing a stable implementation of Linked Art to facilitate broader adoption and unlock its full potential in the CH sector. Moreover, while Klic [2019, p. 23] mentions that Linked Art utilises the CIDOC-CRM ontology to its full expressiveness, there appears to be a lack of clarity or common understanding among readers regarding the precise nature and characteristics of Linked Art, which is a streamlined profile.

One potential driver for wider implementation is the integration of Linked Art with existing standards, akin to the Exhibition Object Data Exchange Model (EODEM)²²³, a framework that enables museum databases to export an import object data more easily. Establishing standardised mechanisms for data exchange could enhance the accessibility and applicability of Linked Art for more institutions.

²¹⁶ Linked Art Community: <https://linked.art/community/>

²¹⁷ Oxford e-Research Centre: <https://oerc.ox.ac.uk/>

²¹⁸ First AHRC grant associated with Linked Art: <https://linked.art/community/projects/researchnetwork/>

²¹⁹ Second AHRC grant associated with Linked Art: <https://linked.art/community/projects/linkedartii/>

²²⁰ Pre-Raphaelites Online: <https://preraphaelitesonline.org/>

²²¹ Linked Conservation Data: <https://www.ligatus.org.uk/project/linked-conservation-data>

²²² LD4 Art & Design Affinity Group: <https://ld4.github.io/art-design/>

²²³ EODEM: <https://cidoc.mini.icom.museum/working-groups/documentation-standards/eodem-home/>

Additionally, the Swiss Art Research Infrastructure (SARI) Reference Models²²⁴ and platforms like Arches²²⁵, developed by the Getty Conservation Institute²²⁶ and World Monuments Fund²²⁷, present promising opportunities for implementing Linked Art in various software and partnership settings.

Overall, while Linked Art does not aim for perfect semantic interoperability, it strives for practical generalisation, positioning itself as a potential benchmark for accessing and aggregating digital collections. Its development and implementation across a spectrum of CHIs underscore its role as a transformative tool in the field, as suggested by Raemy et al. [2023b]. As the community continues to grow and evolve, Linked Art holds the promise of shaping the future of digital CH, fostering greater connectivity and understanding across diverse collections and practices.

3.5.6 Implications and Opportunities for the Cultural Heritage Domain and the Humanities

In envisioning the ideal scenario for the CH domain and the humanities, the integration of LOUD technologies presents unprecedented opportunities. While IIF has laid a foundational path, the full spectrum of LOUD specifications potentially catalyses a more profound transformation in how we interact with and understand CH data. This transformation prompts us to ponder two critical questions:

- *How are LOUD technologies reshaping traditional humanities?* This question probes into the ways LOUD is influencing human interaction, societal communication, and the broader fabric of our cultural discourse. It challenges us to consider the evolving role of technology in mediating our understanding of history, culture, and the arts.
- *As new data becomes increasingly accessible and novel tools for data analysis emerge, how will LOUD enable the formulation of new research questions?* This inquiry looks at the potential of LOUD to unlock uncharted territories of knowledge, offering researchers ways to interrogate and interpret the vast expanse of human history and cultural expression.

The implications of LOUD extend beyond technological innovation; they invite a reexamination of our methodologies, perspectives, and frameworks. KR is inherently subjective, influenced by the context and intent behind data collection and modelling. While it is crucial to understand the epistemic underpinnings of knowledge, the LOUD principles advocate a balanced approach with usability at the forefront.

Does LOUD, and its precursor LOD, represent a paradigm shift in the CH domain and for DH at large, akin to the transformative changes described by Thomas Kuhn in his seminal work on the structure of scientific revolutions [Kuhn 1994]? This ecosystem of models suggests a substantial shift not only in technological capabilities but also in epistemological frameworks. LOUD serves as a unifying common denominator, simplifying the complex landscape of digital CH to hopefully making it more accessible for exploration and interconnection.

3.6 Characterising Community Practices and Semantic Interoperability

In this section, I will define the two main axes or perspectives of my thesis, chiefly ‘community practices’ in subsection 3.6.1 and ‘semantic interoperability’ in subsection 3.6.2. These two concepts, while distinct, are deeply interwoven, each playing a critical role in the advancement and application of CH.

²²⁴ SARI Documentation: <https://docs.swissartresearch.net/>

²²⁵ Arches: <https://www.archesproject.org/>

²²⁶ Getty Conservation Institute: <https://www.getty.edu/conservation/>

²²⁷ World Monuments Fund: <https://www.wmf.org/>

Community practices in CH are foundational to the collaborative and collective efforts that drive the field, ranging across various stakeholders, from DH practitioners, librarians, archivists, and curators to CHIs. This focus on community practices leads us to consider how these collaborative networks communicate and operate effectively on a technical level, which brings us to the concept of semantic interoperability.

Semantic interoperability is crucial in elevating community practices to a higher intellectual plane. It represents the technical backbone that enables these diverse communities to not only share data but also to understand and interpret data consistently and accurately across different systems. This seamless integration and interpretation of semantic data are what allow community practices to transcend beyond simple data exchange to more meaningful and impactful collaborations. By bridging these communal aspects of collaboration with the technical necessities of data exchange, semantic interoperability ensures that the collective efforts in CH are not only shared but also effectively leveraged and understood.

Thus, the transition from community practices to the intricacies of semantic interoperability reflects a natural progression from collaborative engagement to the technical enablers that make such sophisticated interactions possible.

3.6.1 Community Practices

This subsection explores the concepts of community, practices, and community practices, as well as explores the distinctions and overlaps between community practices and CoP. To fully grasp the nuances of community practices, it is essential first to understand what constitutes a ‘community’ and what is meant by ‘practices’. Following this, I will examine how these elements synergies within community practices. Lastly, this exploration will contrast these practices with CoP, highlighting both their characteristics and areas of intersection.

A community is a group of people who share common interests, values, goals or characteristics and come together to form a social unit. Communities can be based on geographical location, shared hobbies, cultural background or any other common factor. In an effort to capture the multifaceted nature of communities, I have selected three definitions from the [Oxford English Dictionary \[2023d\]](#), each highlighting a different aspect of what constitutes a community. These definitions reflect the diversity and complexity inherent in the concept of a community.

Community is:

- *‘A body of people who live in the same place, usually sharing a common cultural or ethnic identity’*
- *‘A group of people who share the same interests, pursuits, or occupation, esp. when distinct from those of the society in which they live.’*
- *‘An online facility, such as an electronic bulletin board, forum, or chat room, where users can share information or discuss topics of mutual interest.’*

The first definition emphasises the traditional, geographical-based community, often bound by a shared cultural or ethnic identity. This perspective, rooted in the physical coexistence of individuals within a specific locale, aligns with the observations by [Jewkes and Murcott \[1996\]](#) that ‘community’ can be variably interpreted, often constructed differently by its members and external observers. They argue that these varied interpretations, although they may not be consistent, are nonetheless meaningful, as each interpreter gives the concept of community a specific, albeit subjective, significance.

The second definition broadens the scope to include groups formed around shared interests, pursuits, or occupations. This acknowledges that communities are not solely confined to geographic boundaries but can be formed around any shared interest or activity, regardless of physical proximity. [McMillan and Chavis \[1986\]](#) expand on this by emphasising the psychological aspects of community, such as a sense of belonging, influence, and shared emotional connection. They highlight the critical role of boundaries, personal investment, and shared symbols in fostering a strong sense of community, illustrating how these elements contribute to the formation and sustenance of community ties.

Lastly, the third definition brings in the contemporary dimension of online communities. Here, [Hammond \[2017\]](#) points out that online communities are distinguished by their reliance on technology for interaction, yet share fundamental aspects with traditional communities, such as a sense of connection, trust, and reciprocal interaction. Hammond underscores that online communities vary in strength and depth, reflecting the evolving nature of community in the digital age. This perspective highlights the complexity of communities in today's world, where interactions transcend physical boundaries and are increasingly mediated by digital platforms.

Looking at practices, it becomes apparent that they encompass a wide range of activities, including routines, habits, skills, methods and regular activities undertaken by individuals. These elements are deeply influenced by a potent combination of cultural, social and personal factors that play a key role in shaping both individual behaviour and societal norms. The [Oxford English Dictionary \[2023f\]](#) provides insightful definitions of practice that illuminate its multifaceted nature.

Practice is:

- *'The actual application or use of an idea, belief, or method, as opposed to the theory or principles of it; performance, execution, achievement; working, operation.'*
- *'The habitual doing or carrying on of something; usual, customary, or constant action or performance; conduct.'*

Building upon the earlier discussion of practices, [Foucault \[1982, pp. 780-781\]](#)'s theory offers a critical lens through which to further examine their societal impact. Foucault's perspective suggests that practices are not simply personal or cultural habits but are deeply embedded within, and influenced by, broader societal power structures and knowledge systems [[Haugaard 2022, p. 345](#)]. In this light, the definitions provided previously by the [Oxford English Dictionary \[2023f\]](#) acquire additional layers of meaning. The first definition, which views practice as the practical application of ideas, beliefs, or methods, resonates with Foucault's idea that practices are where theoretical concepts materialise into actions that both reflect and reinforce existing power dynamics and knowledge paradigms. These practices are not merely acts of execution or operation; they are instrumental in the embodiment of power relations within everyday activities. The second definition, highlighting the habitual nature of practices, aligns with Foucault's view on the subtlety of power. Here, routine behaviours and customary actions are seen as vehicles through which power is exerted and normalised, subtly shaping individual behaviours and societal norms. This habitual aspect of practices contributes to the ongoing process of subject formation, subtly guiding the normalisation of certain discourses and power structures. Thus, from a Foucauldian perspective, practices bridge the theoretical and the habitual, playing a pivotal role in the dynamics of power and knowledge within society [see [Demenchonok 2018](#)].

As for community practices, they encompass a diverse array of shared activities, behaviours, and rituals that play a key role in strengthening the bonds among community members and defining their collective identity. These practices serve as the fabric that weaves individuals together within a particular community, shaping its character and reinforcing its sense of belonging. Moreover, community

practices are not static but rather dynamic, evolving over time to adapt to changing needs, values, and goals of the community. Communities formed around shared interests, pursuits, or occupations, such as hobbyist groups, professional associations, or activist collectives, engage in practices that are geared toward collaborative efforts and the advancement of common goals.

In line with Weil [1996], building community practices involves reinforcing the connections among individuals, groups, organisations, and communities in both geographic and functional terms. The author emphasises the need for activities and policies that foster positive relationships and support problem-solving resources within communities. This is particularly relevant in addressing the challenges posed by socio-economic disparities, racial tensions, and the disappearance of jobs for low-skilled workers. Weil's approach to community building advocates for proactive social work at grassroots and inter-organisational levels, focusing on empowerment and social development. These practices are aimed at enabling communities to not only cope with their problems but also to create sustainable solutions that align with the values of social justice and participatory democracy. Especially pertinent is the emphasis on economic development strategies that can provide communities, particularly those on the economic margins, with greater security and sustainability. This aligns with the broader social work goal of fostering inclusive, equitable, and non-discriminatory communities.

CoP represent a distinct form of community characterised by their members sharing a common professional or intellectual interest. In such communities, individuals collaborate with the goal of advancing their expertise within the specific domain they are passionate about. Key features of CoP include a shared body of knowledge, a cohesive community of people who actively engage in interaction and collective learning, and the presence of shared practices that organically evolve over time, as described by Wenger [2011]. Notably, the entire DH community has been likened to a CoP [see Siemens 2016] and specific initiatives, such as the Linked Infrastructure for Networked Cultural Scholarship (LINCS)²²⁸ recognised themselves as such [Brown et al. 2021, p. 1], highlighting Wenger [2011]'s description on people that *'share a concern or a passion for something they do and learn how to do it better as they interact regularly'*.

Evolution of CoP, as critiqued by Li et al. [2009], reveals a journey from a focus on situated learning and the novice-expert dynamic to broader applications in organisational development and competitiveness. Initially, CoP centred on the development of professional identity through interactions in social environments. This focus shifted to personal growth, trajectories of participation (peripheral versus core), and eventually, to CoP as tools for enhancing organisational effectiveness. The adaptability and evolving interpretation of CoP, however, present challenges in defining and maximising their effectiveness. This evolution suggests a need for strategies that balance personal growth with organisational goals and optimise the CoP characteristics, such as fostering interactions, knowledge sharing, and a sense of belonging. Interventions that enhance relationship building and promote knowledge exchange are essential for optimising the function of CoP in various settings.

According to Leigh Star et al. [2003], the convergence of information artefacts and communities of practice, especially in library science, is a dynamic and reciprocal process. Information artefacts not only shape but are also shaped by the practices and needs of these communities. The interaction between CoP and information systems, as discussed, is crucial for the development of effective and transparent information systems. This is similar to the collaborative practices seen in groups like IIF, where shared standards and open communication among diverse institutions enhance access and interaction with digital collections. For library science, understanding and supporting CoP and these

²²⁸ LINCS: <https://lincsproject.ca/>

collaborative practices is essential for adapting to technological advancements and evolving user needs. The socio-technical perspective offered highlights the importance of considering both social interactions and technical aspects in designing and managing information systems within libraries.

To provide a comprehensive understanding, let's examine the essence of the concepts of 'community practices' and CoP:

Community practices are shared activities and rituals that strengthen bonds among members, define identity, and adapt over time to serve common goals, shaping the character of the community.

CoP represent a specific type of community where individuals engage in collaborative learning and knowledge-sharing practices centred around a shared professional domain or interest. While CoP are a subset of community practices, community practices encompass a wider range of shared activities and rituals that can extend beyond the boundaries of professional or interest-based groups.

Within this context, it is essential to assess the mechanisms by which actors become entangled in consensus-making processes. Community practices often serve as a foundational framework for consensus-building, providing the common ground upon which diverse stakeholders converge. These practices not only define a community's identity but also shape the way decisions are made, knowledge is shared, and values are upheld.

The next subsection explores the second axis of this thesis, which is intricately intertwined with community practices. LOUD specifications, especially IIF and Linked Art standards, represent community-driven endeavours that hold the key to achieving semantic interoperability. These standards provide a shared foundation upon which organisations, individuals, and apparatuses can build, facilitating more effective communication, and reinforcing the consensus-making processes that has been discussed.

3.6.2 Semantic Interoperability

This subsection serves as a baseline exploration of the concepts of semantics, interoperability, and the powerful synergy they form in the space of information exchange known as semantic interoperability.

To begin, semantics can range from simple, general definitions of data elements to highly specific and complex definitions that fully describe the context, relationships, and constraints of the data within a particular domain or system. According to the [Oxford English Dictionary \[2023g\]](#), semantics in linguistics are the '*the meaning of signs; the interpretation or description of such meaning; (chiefly Semiotics) the study of the meaning of signs, and of the relationship of sign vehicles to referents*', as well as the branch of linguistics or philosophy concerned with meaning in language. Whereas in computing, it relates '*to the meaning of the strings in a programming language*'. It is essential to recognise that semantics provide the basis for shared understanding, both in language and technology.

In terms of a more formal definition, the standard definition of (semantic) information (SDI) can serve as a basis for what semantic information is. [Floridi \[2005, p. 367\]](#) describes semantic information as '*well-formed, meaningful, and truthful data.*' This definition can be formally represented using FOL as:

$$\forall x [W(x) \wedge M(x) \wedge T(x)] \rightarrow S(x) \quad (3.3)$$

Here is a breakdown of Equation 3.3 and the three criteria that data (x) must satisfy to be considered as representing semantic information $S(x)$:

1. x must be well-formed, denoted as $W(x)$.
2. x must be meaningful, denoted as $M(x)$.
3. x must be truthful, denoted as $T(x)$.

In simpler terms, ‘for all data x , if x is well-formed, meaningful, and truthful, then x represents semantic information’.

Transitioning to interoperability, it extends beyond basic compatibility, ranging from systems that can exchange data without errors [Fafalios et al. 2023] to more sophisticated levels of interoperability where systems can effectively use the data being exchanged to support a wide range of tasks and operations.

Shedding light on the perspective of the W3C [see Etemad and Rivoal 2023, 6.3.2.], Raemy and Sanderson [2023, p. 22] have taken a stance in positioning LOUD through the prism of interoperability: ‘*Interoperability is a state in which two or more tested, independently developed technological systems can interact successfully according to their scope through the implementation of agreed-upon standards*’.

To delve deeper into the nuances of interoperability, it is important to understand that it can take various forms, as outlined by Ouksel and Sheth [1999]. These forms manifest in different ways, and some examples and descriptions, drawing insights from various sources, are being explored here:

System interoperability which provides interactions between distributed objects and components.

Syntactic interoperability is accomplished when several systems of information discover and access information within transmitted objects. As examples of protocols to enable syntactic interoperability between institutional repositories that have been widely used in the CH field are OAI-PMH and OAI-ORE [see Freire et al. 2018b, Van de Sompel and Nelson 2015].

Structural interoperability involves using standardised data formats to enable seamless communication between systems, such as RDF in the Semantic Web.

Semantic interoperability which will be discussed hereafter.

Exploring the practical application of interoperability in the STEM, the Healthcare Information and Management Systems Society (HIMMS)²²⁹ - an American non-profit organisation founded in 1961 dedicated to transforming the global health ecosystem through information and technology – outlines four levels of interoperability [HIMSS 2020]:

Foundational (Level 1) This first level establishes the inter-connectivity requirements needed for one system or application to securely communicate data to and receive data from another.

Structural (Level 2) It defines the format, syntax and organisation of data exchange including at the data field level for interpretation.

Semantic (Level 3) It provides for common underlying models and codification of the data including the use of data elements with standardised definitions from publicly available value sets and coding vocabularies, providing shared understanding and meaning to the user.

²²⁹ HIMMS: <https://www.himss.org>

Organisational (Level 4) It includes governance, policy, social, legal and organisational considerations to facilitate the secure, seamless and timely communication and use of data both within and between organisations, entities and individuals. These components enable shared consent, trust and integrated end-user processes and workflows.

These levels illustrate the multifaceted nature of interoperability and its role in enhancing data exchange in common models.

Moreover, open standards play a pivotal role in fostering interoperability, aligning with the enrichments policy of Europeana [Manguinhas et al. 2023, p. 7]. Yet, the openness of a standard alone may not guarantee widespread adoption. As highlighted by Nelson and Van de Sompel [2020, 2022], successful interoperability often depends on having a significant platform, whether through commercial influence or active community engagement. In essence, a core ingredient for successful interoperability specifications is a substantial presence, either in the form of commercial influence or engaged communities. In their words:

‘Because of [the growing global adoption of open standards] by GLAM institutions, especially [IIIF specifications stand] as a testimony that rich interoperability for distributed resource collections is effectively achievable. But other promising specifications that aim for the same holy grail are struggling for adoption, and, many times, lack of resources is mentioned as a reason. While that undoubtedly plays a role, it did not stand in the way of rapid adoption of protocols that have emerged from large corporations, such as the Google-dominated schema.org. This consideration re-emphasizes that a core ingredient of a successful interoperability specification, and hence of achieving an interoperable global information web, is a large megaphone, either in the guise of commercial power or active community engagement.’
[Nelson and Van de Sompel 2022, pp. 8-9]

In the context of semantic interoperability, CIDOC-CRM defines it as *‘the capability of different information systems to communicate information consistent with the intended meaning’*. More specifically, the intended meaning comprises of the following elements [Bekiari et al. 2021, p. 18]:

1. the data structure elements involved;
2. the terminology appearing as data;
3. the identifiers used in the data for factual items such as places, people, objects, etc.

Furthermore, Sacramento et al. [2022, p. 124] provides a comprehensive definition of semantic or content interoperability, emphasising the simultaneous exchange of data and meaning. This is achieved through metadata and controlled, shared vocabularies.

‘Semantic (content) interoperability is the ability of computer systems to unambiguously exchange data using a shared meaning, independently of their hardware, software, or platforms. It is concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics). This is accomplished by adding data about the data (metadata) and linking each data element to a controlled, shared vocabulary (ontology).’

As such, by following Floridi [2005, 2011]’s logic of SDI, semantic interoperability can be defined as the seamless exchange of well-formed, meaningful, and truthful data between distinct systems.

The principle that governs the exchange of data between entities needs to be defined. It ensures that when two entities, represented by y and z , exhibit semantic interoperability denoted as $I(y)$ and $I(z)$ respectively, and there exists a seamless exchange mechanism E between them, then certain conditions must hold true for the exchanged data x . In FOL, it can be formally expressed as follows:

$$\forall y \forall z [I(y) \wedge I(z) \wedge E(y, z)] \quad (3.4)$$

Equation 3.4 asserts that for all entities y and z , if they exhibit semantic interoperability and there is a seamless exchange mechanism between them, certain criteria regarding the exchanged data must be met. Specifically, it implies that for any data x being exchanged between y and z , the three conditions of SDI presented in Equation 3.3 must also be satisfied.

The complete implication can be expressed as:

$$\forall y \forall z [I(y) \wedge I(z) \wedge E(y, z)] \rightarrow [\forall x (W(x) \wedge M(x) \wedge T(x)) \rightarrow E(y, z) \wedge S(x)] \quad (3.5)$$

Simply put, Equation 3.5 can be summarised in this way: ‘For all y , for all z , if both y and z exhibit Semantic Interoperability I and there is a seamless exchange E between y and z , then for any data x being exchanged between them, x must be well-formed, meaningful, and truthful.’

Moreover, building upon this definition of semantic interoperability, it is crucial to recognise that the exchange of well-formed or meaningful data hinges on specific constructs within information systems. Delving into this relationship, [Scharnhorst et al. \[2023, p. 1115\]](#) emphasise the close connection between the creation of controlled vocabularies and the paradigms emerging within research. They articulate that:

‘The creation of specific KOS is intrinsically connected to the innovativeness of new systems of thought in research and inherently non-interoperable. For service providers supporting humanities research and providing access to cultural heritage data the challenge lies in a balance between accommodating specific communities and their needs and fostering data exchange across communities. Part of this balance is realized via an emerging network of connected services. This problem is actually shared by all knowledge domains, and not specific to the humanities. What is needed to further foster the implementation of semantic interoperability is to explicitly discuss the context in which it is introduced, including its main target (findability, accessibility, or re-use), its function for research and beyond, and levels of maturity and sustainability when it comes to new technological solutions.’

This observation underscores the imperative nature of addressing these socio-technical requirements to enhance semantic interoperability. Achieving this capability enables different systems to communicate information consistently with the intended meaning, thereby promoting seamless and meaningful data exchange.

In conclusion, the concept of semantic interoperability, encompassing semantics and interoperability, forms a crucial axis of my research. The synergy between these elements and their profound influence on data exchange will be central to my exploration of LOUD through various lenses.

3.7 Summary and Preliminary Insights

This section provides a summary of what has been discussed in this literature review as well as some preliminary insights with regard to the LOUD ecosystem, chiefly the design principles, communities,

standards, and the implementations. It follows the flow of the present chapter and is organised into five subsequent parts. Finally, in subsection 3.7.7, I end with a few reflections on why we ought to care about CH data in the wider sense.

3.7.1 Cultural Heritage Data

CH data are unique and require different — or adjacent — methods of analysis than quantitative scientific data. The diversity of CH data, including tangible, intangible, and natural heritage, presents a challenge in preserving and promoting these resources effectively. Representing CH data in digital form can also be challenging, as it may lead to a loss of context and complexity. To address these challenges, a comprehensive understanding of the various types of heritage resources, their meanings, and values is necessary, along with an effective preservation and valorisation strategy that considers their heterogeneity. For this, I find that we need a vision that takes into account the diversity of human and non-human entities, entangled in community-based socio-technical activities, whether local, interdisciplinary or global. The following provides a concise summary of the key aspects explored in the first section, addressing the distinctive nature of CH data.

The dimensions of CH data I argue that three key dimensions need to be taken into consideration for CH data: heterogeneity, knowledge latency, and custodianship.

Representing and embodying CH data I discuss that the faithful recreation of materiality in digitised and digital-born resources appears to be a significant issue and that it needs to be treated as new types of affordances or surrogates. Moreover, while achieving a high degree of digital material fidelity may be challenging, it is important to acknowledge that accessibility remains a paramount concern in this context.

Collectives and apparatuses are entangled Actors, encompassing individuals, institutions, local and global communities, along with the (digital) infrastructures and their components, are deeply interwoven. These human and non-human entities collaboratively shape and navigate the complex networks of interactions and technologies that form society. Within this context, ANT emerges as a valuable theoretical framework for understanding these close relationships and interdependencies.

3.7.2 Cultural Heritage Metadata

A variety of different metadata standards have emerged, tailored to different functions and purposes, with some dating back to the 1960s, particularly in library settings. However, the majority of other traditional CH metadata standards have emerged between the 1990s and the 2000s. In terms of descriptive metadata standards and conceptual models, MARC, ISAD(G), CDWA, CIDOC-CRM are being supplemented — or gradually substituted — by models and profiles like LRM, RiC, and Linked Art, which — to some extent — align with Linked Data principles, enabling richer relationships and enhancing access to vast and diverse collections.

These evolving metadata standards support a more interconnected, open, and accessible information landscape, enabling researchers, practitioners, and the public that should hopefully navigate and explore CH resources more effectively²³⁰, especially when they are used in conjunction with controlled vocabularies, facilitating cross-domain reconciliation. As for aggregation of CH (meta)data, alternative web-based technologies to OAI-PMH appear to be viable options, using AS in particular.

²³⁰ It is worth noting that although archives and museums have standardised their metadata practices later than libraries, they seem to have identified core models that can be more easily implemented to meet Linked Data principles.

3.7.3 Trends, Movements, and Principles

CH data is intrinsically intertwined with broader trends including Big Data and AI, serving as both a reflection and a product of these influences. Yet it is vital to recognise that CHIs and DH practitioners are – or should be – instrumental in shaping and curating such data. They not only shape and curate them, but also inform and drive the very trends that permeate across our data landscape. Their values, actions and methodologies represent critical drivers in the process, guiding the collection, preservation, and interpretation of CH data.

The scientific movements and guiding principles of Open Science/Open Scholarship, Citizen Science/Citizen Humanities, FAIR, CARE, and Collections as Data collectively permeate and shape CH data. While the first two represent dynamic movements actively involving scholarly practices or the public, the other three are more concerned with the system in which data is situated. LOUD design principles and standards, which are more content-centric, bolster the effectiveness of these principles and movements, fostering open and accessible CH data through their real-world applicability.

3.7.4 Linked Open Data

The Open Web Platform and Linked Data are foundational to the evolution of scholarly research and CH practices, enabling the creation of federated datasets and KGs. The section explores these concepts and their applications in the CH domain.

The web architecture, underpinned by principles, protocols, and identifiers like URI, URL, and URN, facilitates the exchange of data and functionality across applications and systems. It emphasises architectural principles such as orthogonality and protocol-based interoperability and explores various web architectures, including the client-server model and SOA.

The Semantic Web, a vision for a web understood by machines, relies on standards like RDF, RDFS, OWL, and SPARQL. The limitations of RDF in complex KR can be mitigated to some extent with RDF reification and quoted triples. SHACL is also introduced for data graph validation.

Linked data principles promote the publication and interlinking of data on the web, creating a web of data that is navigable and usable. Challenges in Linked Data implementation include GUIs, application architectures, schema mapping, link maintenance, licensing, trust, quality, relevance, and privacy, enhancing the web's potential as an open, interconnected platform.

Deployment schemes like the Five-Star and Seven-Star models provide criteria for publishing open data. These models address the clarity, usability, and applicability of open data, emphasising schema documentation and data quality.

Finally, the application of LOD in the CH sector is explored through examples like Europeana. Despite its potential in improving data quality and visibility, challenges persist, including issues related to cataloguing, adoption of new standards, and the complexity of Linked Data terminology. The section underscores the need for collaboration and community-driven practices for effective LOD implementation.

3.7.5 LOUD: Design Principles, Communities, Standards

LOUD focuses on improving data accessibility primarily for software developers. It balances data completeness and practical considerations like scalability and usability. Its design principles are:

Right Abstraction for the Audience Tailoring data access to specific user needs.

Few Barriers to Entry Simplifying initial engagement with the data.

Comprehensibility by Introspection Ensuring data is largely self-explanatory.

Documentation with Working Examples Providing clear guidelines and practical use cases.

Consistent Patterns Over Exceptions Reducing complexity through uniform patterns.

The systematic review of LOUD in scholarly literature employed the weight of evidence framework. A Boolean query identified relevant papers, with findings showing 46 relevant references from 2018-2023, mainly in English. These papers were categorised into four main categories: mentions of LOUD, descriptions of LOUD, explanations of LOUD design principles, and comparative analyses where theLOUD principles have been reused in various applications.

LOUD integrates technologies, mostly community-driven, like IIF, WADM, and Linked Art. IIF facilitates the sharing of high-resolution images and audiovisual content through a series of specifications, WADM provides a standard for creating and sharing annotations across various platforms, and Linked Art provides a model and an API specification for semantically describing CH. Together, they demonstrate a transformative potential in how CH data is interacted with and understood, reshaping traditional humanities and opening new research opportunities.

3.7.6 Community Practices and Semantic Interoperability

In the exploration of community dynamics and the intricacies of data exchange, two axes or perspectives come into focus in my PhD: community practices and semantic interoperability. These axes represent not just isolated concepts but comprehensive frameworks that influence how communities function and thrive in an interconnected world. Community practices, as shared activities and rituals, weave the fabric of collective identity within communities, while semantic interoperability acts as the bridge for meaningful and truthful data exchange between different systems.

Both community practices and semantic interoperability will permeate the empirical parts of the thesis as I explore LOUD for CH through different prisms. These axes serve as critical lenses through which we can dissect and analyse the dynamics, challenges, and opportunities that arise within this landscape.

3.7.7 The Case for Cultural Heritage Data

One thing that has been partially touched upon but not strongly asserted in this chapter is '*why do we really care about CH?*'

The importance of CH data as primary and secondary sources for DH practitioners transcends the mere definition, interlinking, and preservation of these sources. The undertaking, notwithstanding technological assistance, can be inherently challenging, contingent upon numerous interdependencies.

Moreover, it is deeply rooted in our response to many pressing global challenges, including the far-reaching consequences of climate change largely caused and accelerated by human activities. In an era characterised by profound disruptions, exemplified by events such as heatwaves, fires, droughts, floods, rising sea levels, and the resultant migrations driven by these environmental changes, which affect not only humans but the entire biosphere, it becomes increasingly important to emphasise the societal responsibilities that accompany our engagement with CH as well as DH practices in the Anthropocene [see [Nowvieskie 2015](#)].

4

Exploring Relationships through an Actor-Network Theory Lens

‘As Jim Clifford taught me, we need stories (and theories) that are just big enough to gather up the complexities and keep the edges open and greedy for surprising new and old connections.’ [Haraway 2016b, p. 101]

This chapter serves as the theoretical framework of the dissertation, and its primary goals are to elucidate the theoretical underpinnings and provide a comprehensive toolbox for addressing the identified problem. In the preceding literature review chapter, I highlighted the issue that necessitates attention around interlinking CH. The theoretical framework, sometimes referred to as the ‘toolbox’, which can be likened to the ‘tools’ that will be employed to understand and address this problem.

Here, the primary purpose of this chapter is to offer an in-depth exploration of the tools – which comprises various theories, propositions, and concepts – delineating their characteristics, behaviours, historical applications, interrelationships, relevance to the study’s objectives, and potential limitations. Subsequently, the next chapter will elucidate how these tools will be operationalised in the research process.

The theoretical framework of this study is firmly rooted in Actor-Network Theory (ANT), which will be pursued systematically throughout the research. ANT is a constructivist approach that seeks to elucidate the fundamental dynamics of societies. Unlike traditional perspectives that restrict the concept of an ‘actor’ – or ‘actant’ – to individual humans, ANT expands this notion to encompass non-human and non-individual entities [see Callon 1999, 2001, Latour 1996, 2005].

ANT goes beyond the mere identification of actors and networks; it embodies a comprehensive methodology for exploring the intricate interplay of socio-technical systems. ANT distinguishes itself not only by recognising both human and non-human entities - from individuals and technological artefacts to organisations and standards - as actors (or actants), but also by examining their roles within heterogeneous networks of aligned interests. This approach facilitates a nuanced understanding of enrolment and translation processes, where diverse interests are aligned to form cohesive networks, and the concept of irreversibility, which describes the stabilisation of these networks over time. In addition, ANT introduces the concepts of black boxes and immutable mobiles, highlighting the persistent yet mobile nature of network elements such as software standards that transcend spatial and temporal boundaries [Walsham 1997, pp. 468-470]. These concepts are instrumental in dissecting the dynamics of IIF and Linked Art specifications, which can be considered either full-fledged actors or immutable mobiles, depending on the context of the network under consideration. This dual perspective underscores ANT’s role as both a theoretical lens and a methodological tool, providing a robust framework for dissecting the fabric of socio-technical assemblages and enriching our understanding of DH and CH interconnections.

Additionally, the theoretical framework is enriched by integrating complementary perspectives from Donna Haraway's Situated Knowledges (SK), Susan Leigh Star's Boundary Object (BO), and Luciano Floridi's Philosophy of Information (PI). Each of these frameworks contributes uniquely to our understanding of LOUD and its socio-technical landscape. Haraway's approach emphasises the contextually-embedded nature of knowledge, underscoring the importance of diverse perspectives in shaping our understanding of technological phenomena [see [Haraway 1988](#)]. The concept of BO provides a framework for examining the role of LOUD technologies as mediators among varied groups, highlighting the importance of flexibility and adaptability in technological systems [see [Star 2010](#), [Star and Griesemer 1989](#)]. Meanwhile, PI offers a foundational perspective, viewing information as an intrinsic part of the reality that shapes and is shaped by technologies [see [Floridi 2011](#)]. Collectively, these theories complement the ANT-based approach by providing a multi-faceted understanding of the complexities inherent in LOUD, its technologies, and the communities involved.

This ANT-grounded toolbox is composed of three elements:

1. Demonstrating how non-human entities exert agency.
2. Identifying the human and non-human actors involved in these processes.
3. Investigating the concept of translation and the process by which a network can be represented by a single entity.

For example, when considering the design of the PIA data model, a full-fledged actor in its own right, and more broadly for any KR system, pertinent questions arise about the influences exerted by the various groups of individuals involved in the process. These questions concern not only their interactions with each other but also their impact on the manifestation of the model and, consequently, how KRs can influence the various actors, both during its implementation and throughout its creation. In addition, the data model is composed of parts always bigger than the sum of their individual characteristics, as each part not only contributes to the overall functionality but also embodies a complex network of relationships and interactions. This viewpoint, inspired by [Latour et al. \[2012\]](#), asserts that in the realm of social connections, following Gabriel Tarde's monadological approach – i.e., viewing each individual or element as a self-contained universe or 'monad' with its own unique properties and relations [see [Tarde 2000](#)] – individual elements (such as stakeholders or data points) often carry more information and potential than what is apparent when they are viewed solely as components of a larger system. In this perspective, the complexity and richness of each individual element often surpasses the aggregate. Thus, the model, potentially acting as a boundary object when not aligned with standardised processes, serves as a site of negotiation and alignment among different stakeholders. As [Haraway \[2016b, p. 104\]](#) poetically puts it, software could also be defined as '*imploded entities, dense material semiotic "things"*', a notion that underscores the entanglement of information, technology, and materiality.

This chapter is organised into four sections, corresponding to the three aforementioned aspects and an additional section focusing on the revised epistemological foundations.

First, Section 4.1 explores the dissolution of rigid distinctions between human and non-human actors, emphasising the dynamic and interdependent nature of such relationships. This exploration is fundamental in understanding the broader implications of standards and community engagements in any field.

Then, Section 4.2 examines how collectives composed of differing actors can be assembled into a cohesive network where each entity's agency and influence are recognised. In this section, the concepts

of quasi-object of BO are introduced to elucidate the role of shared objects and concepts in mediating and facilitating interactions among diverse groups within a network.

Section 4.3 investigates the translation process where actors negotiate, modify, influence, and align their interests and identities in the formation and maintenance of networks. Additionally, PI, particularly the System-LoA-Model-Structure (SLMS) approach, is introduced here to provide a structured understanding of how information and knowledge are conceptualised, managed, and communicated within these networks, offering a deeper insight into the dynamics of standard adoption and community interaction in collaborative efforts.

Finally, Section 4.4 revises the epistemological foundations to address the nuanced inquiries presented later in the empirical chapters. In this section, SK is introduced, emphasising the importance of context-specific and perspective-driven knowledge in shaping our understanding of technological and cultural phenomena. This concept, developed by Donna Haraway, advocates for a more critical and reflexive view of knowledge, recognising that all knowledge is situated within specific cultural, historical, and personal contexts. This approach challenges the notion of objective or universal knowledge, asserting that all understanding is partial, located, and contingent. The incorporation of SK is crucial for comprehending how different actors' perspectives and experiences influence the implementation and interpretation of LOUD standards. It helps in examining how these standards and community participation in IIF and Linked Art are perceived and enacted differently across various social fabrics, particularly contrasting settings where these standards and communities are not engaged. Central to this discussion is the question, 'How does the implementation of LOUD standards for CH resources and participation in the IIF and Linked Art communities transform the social fabrics, in contrast to settings where these standards and communities are not engaged?' This question serves as a common thread, leading into the main research question: 'How to situate LOUD and to what extent has LOUD shaped or will shape the perception of Linked Data in the broader context of CH and DH?'. Throughout these sections, ANT forms the underlying common thread, with the other theories augmenting and enriching this comprehensive theoretical framework.

Overall, the theoretical framework will be drawn upon to explore what [Manovich \[2017, pp. 60-61\]](#) refers to as '*everything and everybody*'. Borrowing from Haraway's concept of Tentacular Thinking, this approach recognises the interconnectedness and interdependence of all elements within the research scope, from the minutiae of technical details to the broader societal implications [see [Haraway 2016a](#)]. This comprehensive view is essential for addressing the detailed nuances of technical implementations, yet it is also crucial for understanding their wider societal implications, and for considering the multi-layered complexities involved in the implementation and perception of LOUD.

4.1 Implosion of the Boundaries: Non-humans have Agency

The [Oxford English Dictionary \[2023a\]](#) assigns two primary sets of meanings to the term 'agency'. The first pertains to '*a person or organisation acting on behalf of another, or providing a particular service*'. The second, of greater relevance to this discussion, relates to an '*action, capacity to act or exert power*'. This second definition is further elaborated as an '*action or intervention producing a particular effect; means, instrumentality, mediation*'.

The concept of agency has been a central theme in various philosophical and sociological discourses. In ancient philosophy, both Plato and Aristotle contributed foundational ideas to the concept of agency, each offering distinct perspectives that have significantly influenced subsequent thought.

Plato, known for his theory of ideal forms, presented a dualistic view of reality, distinguishing between the world of forms (ideas) and the physical world. Within this framework, he saw agency as the soul's ability to recognise and conform to these ideal forms [Watson 1975, p. 209]. For Plato, true agency involved transcending the physical and sensory world and directing one's actions according to reason and intellect. This pursuit of knowledge and truth was seen as the highest form of agency, with actions aimed at realising eternal and immutable truths. Plato's vision of agency is closely tied to knowledge, virtue and the pursuit of the good, as seen in his portrayal of the philosopher-king in 'The Republic', who governs himself and the state with wisdom and insight [see Plato 360 B.C.E].

Aristotle, Plato's student, offered a more practical and empirical approach. He incorporated agency into his broader ethical framework, placing emphasis on the ability to act virtuously and to make decisions in accordance with a *telos* or purpose. Aristotle regarded every action and choice as directed towards an end, and this teleological approach is key to understanding his concept of agency. Agency in Aristotle's philosophy is deeply intertwined with the notions of potentiality and actuality, where potentiality represents inherent capabilities and actuality is their realisation through action. This perspective reinforces the importance of rational deliberation, moral virtue, and the realisation of potentiality in human life. In addition, Aristotle emphasised the role of choice (*prohairesis*) and practical wisdom (*phronesis*) in guiding deliberate, rational and virtuous action [see Charles 2017].

These ancient philosophical perspectives, with Plato's focus on reason and alignment with the ideal, and Aristotle's emphasis on practical wisdom and virtue, set the stage for later philosophical explorations of agency by modern thinkers such as David Hume (1711-1776), Immanuel Kant (1724-1804), and Georg Wilhelm Friedrich Hegel (1770-1831). In their perspectives, agency is understood as an individual's capacity to act in the world, based on intentionality and rationality [Pippin 1991]. Hume's empiricist approach saw agency as closely tied to the experiences and perceptions of the individual, underlining the role of personal choices and mental states [see Schier 1986]. Kant's critical philosophy stressed the importance of autonomy and moral law in agency, which he understood as the capacity of individuals to act according to universal moral principles derived from reason. Hegel offered a more dialectical approach, seeing agency as part of a broader historical and social process in which the actions of the individual are interwoven with the unfolding of rational will in history. These classical views of agency focus primarily on human agents and their conscious, intentional actions.

In contrast to these traditional perspectives, the advent of ANT and the works of Bruno Latour, John Law, Madeleine Akrich, and Michel Callon, mark a significant shift. These academics argue for a more inclusive understanding of agency, where non-human entities — ranging from technological artefacts to animals and even ideas — can also be agents that influence and shape the course of social events. This perspective is a departure from the anthropocentric view of agency and opens up new ways of understanding social dynamics and networks.

The concept of the 'actant', as introduced by Algirdas Julien Greimas (1917-1992), is pivotal in this context. In Greimas' semiotic theory, an actant can be any entity, human or non-human, that contributes to the progress of a narrative. This concept significantly expands the traditional narrative framework established by Vladimir Propp (1895-1970), which focused mainly on human characters and their roles in folk tales. Propp's analysis focused on the actions and roles of these character types, which he categorised into a standardised framework. Greimas' approach to narratives, as referenced by Boullier [2018], emphasises the potential of any entity to play a role in a story, thereby expanding the narrative scope.

'The agency's move is based on a well-known but seldom mentioned loan from Greimas' 1966 semiotics. The concept of "actant" allowed the potential arrangement of any entity that

populated the narratives to be aligned beyond Propp's tradition. While Greimas' formalism was certainly not preserved, the principle allowed for more open stories to be told and the concept of "allies" to be formalised, in particular, which extended the idea of "adjuvants" and "opponents" (without this being done from a strategic perspective, contrary to some interpretations).' [Boullier 2018]

This idea of actant resonates strongly within ANT, as it aligns with the theory's aim to dissolve the strict dichotomy between human and non-human actors. In ANT, actants are not limited to individuals or even sentient beings; they include any entity that can affect or be affected by the network. This redefinition of agency through the lens of ANT and the concept of the actant is a cornerstone in understanding the complex, interconnected networks that constitute social and technological realms. It allows for a broader and more nuanced understanding of how various elements within a network interact and influence one another, regardless of their traditional classification as human or non-human.

ANT provides a radical redefinition of agency, challenging the modernist and post-modernist interpretations. It proposes an 'amodern' viewpoint [Latour 1990], dissolving the dichotomy between human and non-human agency, and focusing on the network dynamics in a society where '*contemporary techno-science consist of intersections or "hybrids" of the human subject, language, and the external world of things, and these hybrids are as real as their constituents*' [Bolter and Grusin 1999, p. 57]. Agency, in ANT, includes not just intentional actions but the capacity of any entity to affect or be affected in a relational network [Latour 1996]. This expansive view of agency, influenced by the concept of the actant and informed by Greimas' semiotics, offers a more holistic understanding of actors within networks.

Adopting a Latourian approach, researchers such as anthropologists and sociologists are encouraged to observe the balance between human and non-human properties within networks. This balanced observation is crucial for a full understanding of the dynamics within these structures [Latour 1993, p. 96]. By acknowledging the agency of both human and non-human agents, and by recognising the blurred boundaries between subjects and objects, researchers can gain deeper insights into the complex interplay of forces that shape social realities.

4.2 Assembling the Collective

Following the exploration of agency in the previous section, the assembly of collectives is examined. This process involves identifying a myriad of actors, both human and non-human, and understanding how they coalesce into actor-networks. The assembly of such a collective, a differing cosmos of socio-technical agents, is predicated on the recognition of each actor's unique agency and the dynamic interplay of relationships that bind them together.

The transformation challenges of tools for DH and object knowledge, as discussed by Camus et al. [2013], highlight the complexities involved in integrating DH with traditional scholarly practices. The author emphasises the need for a nuanced approach to the digitisation and dissemination of CH resources, underscoring the pivotal role of collaborative efforts in bridging the gap between technology and humanities scholarship. This analysis aligns with the ANT perspective, which advocates for recognising the contributions of diverse actors within the wider DH ecosystem.

Within the LOUD ecosystem, a diverse set of actors comprises individual contributors, institutions, and several groups and committees, each with its own set of objectives. This ensemble also includes specifications and compatible software that facilitate interoperability, as well as end users. Interestingly,

the majority of these end users remain unaware — whether through seamless integration or simply because their interaction does not require conscious recognition — that their digital interactions are often mediated by or compliant with LOUD standards. This diversity underscores the importance of understanding how different actors, their objectives, and their contributions shape the development and adoption of LOUD specifications and practices.

Transitioning from the depiction of the LOUD ecosystem's varied participants, the concept put forth by [Gandon \[2019b\]](#) for an envisioned web architecture introduces a '*multi-agent system*' composed of diverse natural intelligences – such as humans, connected animals and plants – and artificial intelligences, including entities capable of reasoning and learning. This shift marks a deeper recognition of the layered interactions that form the backbone of digital platforms, and points to a future where technology adapts to embrace a wider range of intelligences within the structure of the web.

Classifying these diverse actors involves understanding their roles and interactions within the network. Quasi-objects and BOs provide frameworks for this classification, enabling a nuanced understanding of the socio-technical assemblage and facilitating communication among its varied components.

The foundation established by ANT, incorporating the concept of 'quasi-objects' from [Serres \[2014\]](#) challenges traditional categorisations by embodying characteristics of both subjects and objects. This conceptual framework is crucial for appreciating how non-human entities can exhibit agency and actively participate in social networks, thus broadening our understanding of actor-network dynamics.

Quasi-objects, existing in a state of flux and embodying characteristics of both subjects and objects, challenge our conventional understanding of agency. Concurrently, the concept of BOs is introduced, enriching the ANT-grounded toolbox by highlighting the role of shared objects and concepts in mediating and facilitating interactions among diverse groups within the network. Unlike quasi-objects, which symbolise a hybrid state between subjectivity and objectivity, BOs focus on interaction and communication. They are crucial in collaborative efforts, especially in diverse and interdisciplinary settings, by maintaining a common identity across various contexts while being interpreted differently in each. Understanding the distinction and the interplay between quasi-objects and BOs is vital for comprehending the dynamics of actor-networks. The introduction of BO in this section elucidates their role in mediating complex socio-technical interactions, highlighting the importance of BO in community-based initiatives and their broader impact. Star's reflection on BOs underscores their significance:

'Boundary objects are objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. They may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is key in developing and maintaining coherence across intersecting social worlds.'

[[Star 1999](#), p. 393]

The relevance of BO extends to the restructuring of residual categories through cycles of standardisation attempts that create said boundary objects as illustrated in Figure 4.1. This cycle emphasises the negotiation and alignment among different stakeholders, underscoring the adaptive and flexible nature of BO in managing the complexities of standardisation and the varied interpretations across social worlds [see [Star 2010](#)].

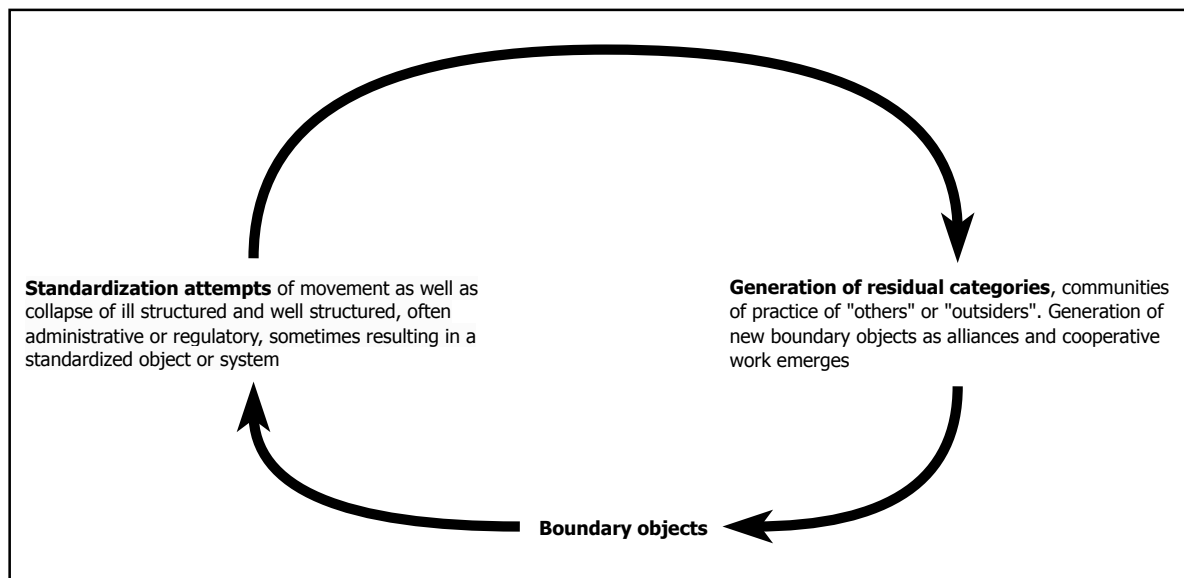


Figure 4.1: Relationships Between Residual Categories, BOs, and Standardisation Attempts. Adapted from [Star \[2010\]](#)

The integration of concepts such as quasi-objects and BOs within the ANT-based toolbox necessitates a reevaluation of existing ontological frameworks. This shift leads to a more interconnected understanding of societal dynamics, recognising the central role of diverse actors in forming networks.

The adoption of these concepts in ANT requires a re-evaluation of existing ontologies. This re-evaluation involves redefining our understanding of agency, action and influence to include a wide range of actors, both human and non-human. This approach leads to a more nuanced and interconnected understanding of society, where traditional boundaries between different types of actors are actively reimagined. In this way, ANT argues for a network-like ontology and social theory that can fully integrate the influence and interactions of disparate actors within society [[Latour 1996](#), p. 370].

In engaging with the complexities of assembling the collective, we encounter the necessity to cease replicating the ‘archontic principle of the archive’, as discussed by [Derrida \[2008\]](#). This principle, which dictates the preservation and accumulation of knowledge under a single authority or location, is challenged by the fluid and distributed nature of ANT’s networks. Derrida’s critique invites a rethinking of how knowledge and information are curated and disseminated, echoing the ANT perspective that stresses the distributed, multifaceted interactions of actors within a network. The dialogue with Derrida’s deconstruction of the archive complements the ANT approach by advocating a more open, inclusive understanding of how socio-technical collectives are formed and maintained.

For instance, the examination of BOs in the context of queer identities illustrates the potential of these concepts to challenge and redefine traditional typologies, offering new perspectives on identity and community formation within socio-technical networks [see [Junginger and Dörk 2021](#)].

Having assembled our collective and identified the diverse actor-networks, the focus shifts to exploring the relationships and communication mechanisms among them. This exploration, conducted in the subsequent section, is crucial for understanding how different actors – ranging from quasi-objects to boundary objects – contribute to and shape the collective narrative and functionality of a given network.

4.3 The Translation Process

The translation process within ANT refers to the dynamics of establishing associations and networks among diverse actors. Latour describes translation as a specialised relation that transforms mediators into coexistent entities without directly transferring causality. This concept underscores the complexity of interactions within networks, emphasising the creation and maintenance of associations that extend beyond mere causal relationships.

‘[T]he word “translation” now takes on a somewhat specialised meaning: a relation that does not transport causality but induces two mediators into coexisting’ [Latour 2005, p. 119]

Associations within ANT evolve continuously, demonstrating the ongoing and emergent nature of actor-networks. They materialise through the interactions and negotiations among actors. This perspective broadens the understanding of networks by stressing their fluid nature.

Understanding the dynamics of actor-networks requires an exploration of the ways in which actors influence each other. This is especially important in order to capture the nuanced roles and interactions within these networks. A key contribution in this area has been made by Latour, who suggests a compelling structure for disentangling these interactions as ‘mediation’. Latour’s conceptualisation is particularly insightful in that it breaks down the nature of influence into distinct yet interrelated processes. The essence of his argument can be summarised as follows:

‘This capacity of actors to influence each other was defined by Latour [2005] as mediation, further broken down into four types: interference, composition, black boxing, and delegation. Interference appears when one actor interferes with the goal of another. In composition, the actors influence the common goal of the network together. Black-boxing is when gradual complexification of actors (and their interrelations) reaches a point where treating the constellation as a single actor becomes more meaningful, and delegation is when meaning and expression is delegated to non-human objects.’ [Czahajda et al. 2022, p. 3]

In the context of the LOUD space, treating a constellation of actors necessitates a focus on delegation, a process pivotal for understanding how meanings and functions are assigned within networks. This emphasis on delegation underscores the necessity of expanding our epistemological horizons to encompass PI. Integrating PI offers a comprehensive framework for analysing how information not only mediates relationships within these networks but also influences the reality of digital ecosystems. Such an expanded perspective is essential for a thorough understanding of the dynamics within the LOUD ecosystem and its broader implications.

PI is conceived as a groundbreaking approach that examines the nature and dynamics of information. It explores how information fundamentally structures reality and thereby shapes the entities within it, termed ‘inforgs’ or informational organisms. These entities are embedded in information environments where they engage and interact in a vast information ecosystem. Floridi [2010]’s work illuminates the ways in which information underpins and transforms our understanding of reality, and suggests that living in the information age means recognising our role and identity as part of a complex, interconnected informational world. This perspective invites a deeper reflection on the implications of living in the midst of vast information networks, and urges a re-evaluation of how information influences human identity, society and our wider interaction with the digital and natural worlds.

Diving deeper into Floridi’s PI, the LoA concept emerges as a critical tool for dissecting the intricacies of complex computational systems, including LOUD-compliant ecosystems. This approach,

fundamental to PI [Ganascia 2015], aids in navigating the multifaceted layers and perspectives inherent in these systems [Angius et al. 2021], particularly those structured around client-server architectures. LoA provide a structured way to analyse and understand complex systems by breaking them down into different layers or perspectives. Each level focuses on specific aspects of the system, allowing for a clearer analysis of its components and their interactions. By advocating for a separation of concerns, the LoA framework equips us with a strategic method to manage and simplify complexity, enabling a focused examination of distinct abstraction levels within digital ecosystems [van Leeuwen 2014].

Building on the foundation laid by Floridi's LoA, Selbst et al. [2019] critique the fair ML field's reliance on abstraction and modular design for achieving fairness, identifying five abstraction traps that highlight the challenges of applying computational interventions to societal contexts without considering the interplay between social context and technical systems. This critique underscores the relevance of incorporating a socio-technical perspective in this thesis, emphasising the need to make use of STS methodologies in the design process to avoid these abstraction traps.

Considering the insights from Selbst et al. [2019], this thesis will explore four levels of LoA and one transversal dimension within LOUD ecosystems, acknowledging that each can act as its own actor-network or collectively form a singular network. These levels, from low to higher abstraction, include Algorithmic and Computational Processes, Infrastructure, Data Model, and Representation and Display. Societal implications, integrated across all levels, will address the broader cultural and social impacts.

Algorithmic and Computational Processes This level explores the specific algorithms, such as image conversion scripts or sorting algorithms, and computational frameworks leveraged within the LOUD ecosystem. For example, analysing how a recommendation algorithm influences the accessibility of digital resources.

Infrastructure It focuses on the servers, cloud-based services, and micro-services that support the ecosystem. This includes assessing the deployment of server architectures that facilitate scalable data storage and retrieval, like using Amazon Web service or GitHub Pages for hosting LOUD data.

Data Model It covers the organisation and structuring of data, including metadata standards and LOUD specifications. An example would be examining how CIDOC-CRM is used in Linked Art.

Representation and Display This level not only encompasses the JSON-LD outputs for effective data sharing and interlinking but also focuses on how users see and interact with LOUD representations across different platforms – browsers, viewers, and players. It examines the GUI that make digital collections accessible and engaging. For example, how a virtual gallery allows users to explore a digital exhibit.

In the exploration of LOUD ecosystems, the concepts of 'immediacy' and 'hypermediacy', as delineated by Bolter and Grusin [1999], provide additional insightful perspectives on the role of interfaces as LoA. Immediacy refers to the design of interfaces that aim to create a seamless, transparent UX, making the technology invisible and allowing for direct interaction with the content. This is evident in interfaces such as OSD, which displays high-resolution images and strives to provide a smooth and immersive viewing experience by minimising the perceptibility of the API-compliant resource. On the other hand, hypermediacy emphasises the presence and visibility of the medium, drawing attention to the various forms of mediation. Interfaces embodying hypermediacy offer a multi-layered, heterogeneous presentation, making users aware of the different media elements and their interactions, e.g. Exhibit for storytelling purposes. This duality enriches the user's digital encounter, underscoring the need

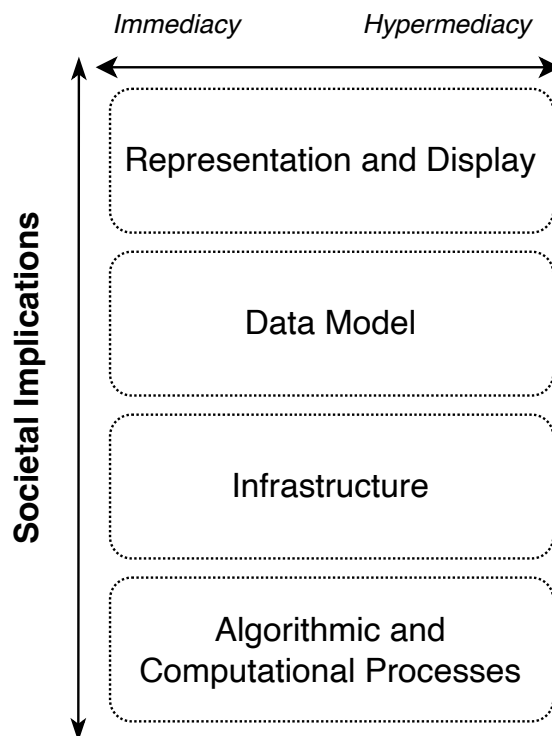


Figure 4.2: Exploring LoA in LOUD Ecosystems: Integrating Societal Implications with the Concepts of Immediacy and Hypermediacy

for LOUD-compliant tools and services to mediate these experiences effectively. By embracing these concepts, the framework can strategically leverage interfaces to either conceal or reveal the intricacies of the digital medium, facilitating a nuanced engagement with information ecosystems. Figure 4.2 presents a comprehensive view of the LOUD ecosystem's LoA, enriched by the inclusion of societal implications as a cross-cutting dimension and the incorporation of immediacy and hypermediacy as critical concepts at the representation and display level for understanding user interaction and interface design.

The System-LoA-Model-Structure (SLMS) scheme, which includes a framework for each identified LoA, equips this research with a comprehensive lens for analysing computational systems, revealing how they can be effectively combined with ANT. This methodology, as depicted in Figure 4.3, enriches the exploration of digital ecosystems. This combination offers a unique perspective on understanding the interplay between computational entities and the broader networks they inhabit. The SLMS scheme can be summarised as follows:

System This refers to any entity or collection of entities that can be studied. It could be anything from a physical object, a biological entity, to a social or computational system.

LoA Floridi emphasises the importance of levels of abstraction in understanding and analysing systems. A level of abstraction is a way of observing a system, focusing on certain aspects while ignoring others. It's a conceptual framework or lens through which we can understand complex systems.

Model At each level of abstraction, we create models of the system. These models are simplifications or representations of the system that highlight certain features while omitting others, based on the chosen level of abstraction.

Structure This refers to the organisation or arrangement of the components within the system, as understood or represented at a given level of abstraction.

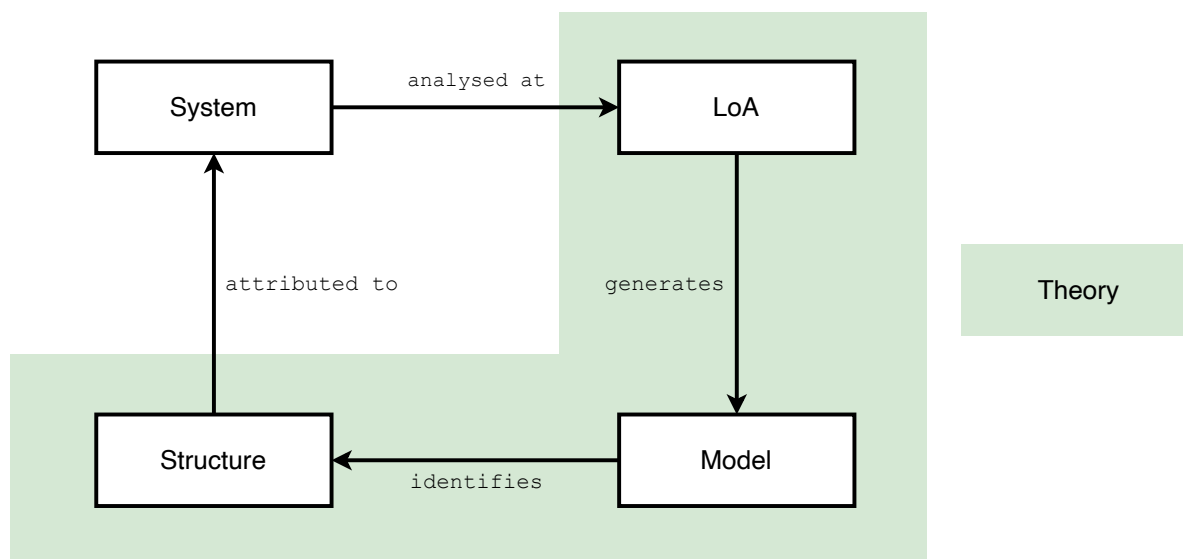


Figure 4.3: The System-LoA-Model-Structure (SLMS) According to Floridi [2008]

Gobbo and Benini [2016] expand on the SLMS scheme by highlighting the challenges and intricacies of quantifying and qualifying computational information. They advocate for a comprehensive methodology that appreciates both the physical and conceptual dimensions of data, facilitating a deeper understanding of programmable artefacts and their informational content. This perspective not only complements the analytical capabilities of ANT but also opens new avenues for investigating the dynamics of information and technology.

As I venture to revise the epistemological foundations and introduce Haraway’s concept of SK, it becomes increasingly manifest that the integration of ANT with Floridi’s PI and the insights of computational information theory provides a robust framework for exploring the complexities of digital and networked environments. This interdisciplinary approach lays the groundwork for a comprehensive exploration of the digital world, emphasising the importance of situated, contextual knowledge in understanding and navigating the digital landscape.

4.4 Epistemological Foundations

This section establishes the epistemological foundations, presenting ANT, BO, and PI, alongside Donna Haraway’s SK. Rather than synthesising these theories, this chapter places an emphasis on situating LOUD within a feminist perspective to construct a new materialistic foundation reminiscent of Haraway’s ‘alignment in tentacular worlding’ [Haraway 2016b, p. 42]. This assemblage seeks to navigate the controversies and mappings within LOUD-like communities, applying a Tardian approach to trace the spreadability of ideas Latour et al. [2012].

To analyse the relevant actor-networks effectively, especially being part of both the IIIF and Linked Art communities, a particular lens is required. ANT, while expansive, has faced criticism for its perceived flatness in analysing networks. Here, Haraway’s SK becomes instrumental, providing a stance that enriches the ANT-grounded theoretical framework with a comprehensive lens that prioritises context in shaping knowledge. SK emphasises that knowledge is always situated, partial, and contextually produced, offering a critical perspective on determining relevance within networks. SK complements ANT by adding depth to the analysis of actor-networks. It highlights the significance of context – both human and non-human – in the production of knowledge, thereby enriching the theoretical framework with a nuanced lens for exploring the dynamics within the LOUD space.

SK, as articulated by Haraway [1988], emphasises the contextual nature of knowledge and challenges the pursuit of an objective, universal truth divorced from the position of the knower. Haraway’s framework, which integrates standpoint theory, concedes that knowledge is inherently shaped by its social, cultural and historical context, and supports an understanding of knowledge as partial and situated. This approach, noting the influence of epistemological privilege and intersectionality, argues against universalism by stressing the importance of being conscious of the specific perspectives and biases that inform one’s understanding. The relevance of SK to ANT lies in its complementary perspective of acknowledging the diverse, context-specific factors that influence knowledge production within networks, enriching ANT’s analysis of actor-networks by incorporating a critical, reflexive lens on the situatedness of knowledge.

In forging this theoretical framework, I seek to transcend the notion of merely ‘cobbling together’ disparate ideas. Instead, I aim to weave their contributions into a coherent web of thought, ensuring a seamless and comprehensive framework that embodies the essence of their respective insights, one that transcends disciplinary boundaries. The theoretical framework can be synthesised as follows, elegantly interweaving distinct yet complementary perspectives to enrich our understanding of socio-technical ecosystems:

- SK, by advocating for an understanding of knowledge as inherently partial and situated, complements ANT by adding depth to the analysis of actor-networks. It enriches the theoretical framework with a nuanced lens for exploring the dynamics within the LOUD space.
- BO further refine this framework by offering a means to characterise actors and mediate interactions within the network, facilitating connectivity and translation among diverse groups.
- PI enriches ANT by framing information as a fundamental element in actor-network interactions. It offers insights into the different LoA, or different actor-networks, each with a different perspective or granularity, and how information is created, shared, and used, and how these processes influence the relationships and dynamics within networks.

Figure 4.4 illustrates how these theories intertwine to form the epistemological foundation of this research, demonstrating the synergistic potential of combining ANT with SK, BO, and PI to navigate the complexities of digital ecosystems in the CH field.

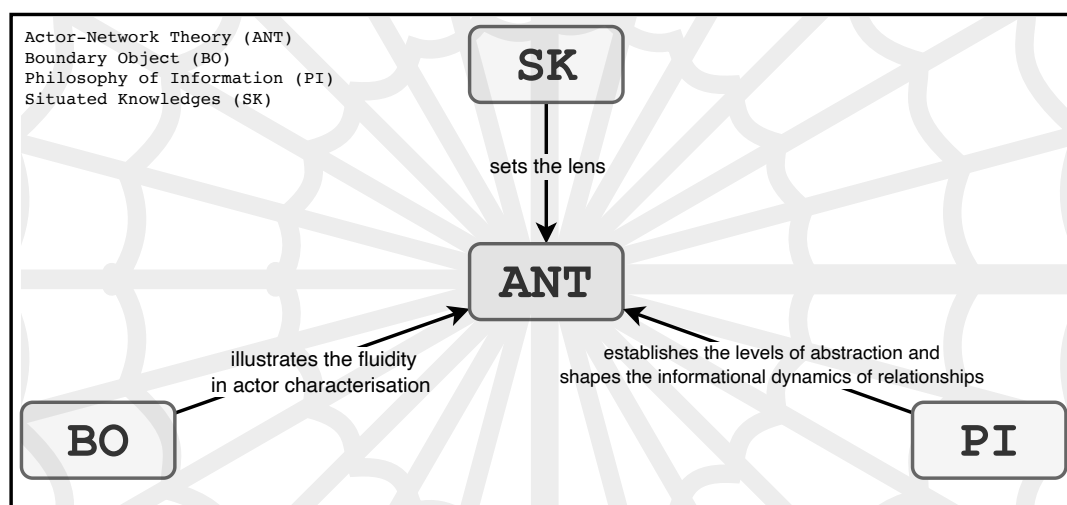


Figure 4.4: A Sympoiesis of Theories: ANT Entangled with SK, BO, and PI

Concluding this chapter, the developed toolbox lays a coherent foundation for empirical research, poised to explore the dynamics within actor-networks. This exploration is not entirely novel in the context of CH, as exemplified by [Guillem et al. \[2023a\]](#)'s use of ANT to spotlight the keystones that were destroyed by the fire at Notre-Dame de Paris. As this narrative unfolds into Chapter 5, the ANT-based toolbox, enriched with a modern and feminist perspectives, will be instrumental in navigating the forthcoming empirical landscapes.

5

Research Scope and Methodology

This chapter delineates the Research Scope and Methodology, laying the groundwork for the empirical exploration within this thesis.

Structured into five key sections, this chapter outlines the methodological framework adopted to explore the purposes of the research and specifies the systematic approach used to consider the outcomes. It initiates with the presentation of research questions in Section 5.1 and hypotheses in Section 5.2 that are crucial for evaluating LOUD within the CH context. Following this, Section 5.3 details the objectives aimed to be achieved based on the research findings. The methodology, elaborated in Section 5.4, encompasses the research methods, data creation, collection, curation, modelling, analysis, alongside considerations of limitations and ethics. The final section, Section 5.5, defines some of the scholarly implications of the thesis, highlighting its substantial contributions to DH and the broader field of the humanities. This discourse encapsulates the implications of the research, affirming its value and positioning within the scholarly community.

Throughout this thesis, the two pivotal perspectives, *Community Practices* and *Semantic Interoperability*, serve as threads interwoven across chapters. While certain research questions, hypotheses, and objectives are specifically aligned with one perspective, others bridge both, showcasing their interconnected relevance.

5.1 Research Questions

Within this thesis, the primary research question seeks to critically evaluate the role and impact of LOUD, exploring how it influences and integrates within the fields of CH and DH. This research seeks, in part, to uncover the depth of LOUD's contribution to these disciplines, and to explore its potential to reshape the perception and use of Linked Data. Framing the central inquiry of this investigation, the primary research question is articulated as follows:

How can the concept of LOUD be situated within the broader framework of CH and DH, and to what extent has it influenced – or is likely to influence – the perception and use of Linked Data in these fields?

In order to explore this overarching question, three specific research questions have been formulated to probe deeper into the socio-technical dynamics of LOUD:

- RQ01** What are the socio-technical requirements for developing specifications in line with the LOUD design principles?
- RQ02** How does the implementation of LOUD standards within the PIA research project contribute to understanding their potential benefits and limitations in terms of facilitating data reuse and broader participation?

RQ03 What are the implementation strategies, challenges and outcomes of Yale’s large-scale deployment of LOUD standards, and how does ensuring the consistency of Linked Art and IIIF resources within the LUX platform contribute to the CH domain?

These questions collectively aim to illuminate the multifaceted implications of LOUD, paving the way for a nuanced understanding of its benefits and challenges. The subsequent section will transition into the hypotheses derived from these research inquiries, further delineating the expected insights and contributions of this investigation.

5.2 Hypotheses

Transitioning from the research questions, this section posits a main assumption based on the expectation that LOUD design principles, being grassroots-driven and based on consensus and transparency, will foster greater acceptance within CH and DH fields. Despite the nascent recognition of LOUD as a distinct entity and its perception as a technical utility rather than part of a broader socio-technical ecosystem, these hypotheses aim to explore the multifaceted impacts of LOUD standards on these disciplines.

The hypotheses outlined below seek to elucidate the dynamics and potential outcomes identified in the research questions:

HYP01 Grassroots initiatives to establish functional communities take significant commitment from all sides and are frequently developed slowly since the majority of those participating or interested can only be extensively committed for brief periods of time. However, the efforts undertaken within these initiatives, such as the development of LOUD-compatible standards, are generally successful in the medium to long term.

HYP02 The PIA research project’s implementation of LOUD standards, leveraging shared APIs such as the IIIF and Linked Art specifications, will shed light on the potential benefits and limitations of these standards in facilitating data reuse and wider participation in the context of CH and DH. We anticipate that digital infrastructures using LOUD standards should be able to improve the long-term accessibility and re-usability of CH data, thereby encouraging active participation and engagement by diverse user communities.

HYP03 The research aims to demonstrate that Yale’s large-scale implementation of LOUD standards involves a variety of practices. It is expected that while LOUD standards such as the IIIF APIs have become more accessible for implementation, challenges remain due to specific skills and resource constraints within CHIs. It also aims to showcase how overcoming these challenges and ensuring the consistency of Linked Art and IIIF resources within the LUX platform contributes significantly to the advancement of shared processes and accessibility within the CH domain, ultimately facilitating more widespread and effective use of CH data.

In summary, these hypotheses underpin the aim of the thesis to examine the influence and operationalisation of LOUD, setting the stage for a detailed exploration of the objectives in the following section.

5.3 Objectives

This section aims to define the concrete objectives that guide the empirical exploration of LOUD. Through leveraging ANT epistemological foundations, these objectives are crafted to dissect various aspects of LOUD’s integration, adoption, and impact. Each objective targets a specific dimension, from

community consensus and participation to the practical implementation within PIA and evaluation of LOUD standards in Yale. More specifically, five objectives have been identified. They do not maintain a 1:1 relationship with what was being presented earlier as two objectives are linked for the first and the third hypothesis and research question.

OBJ01.1 Analysing consensus building and advocacy initiatives across the IIF and Linked Art communities

OBJ01.2 Examining community participation, the evolutionary adoption of LOUD standards and their potential with regard to scientific movements and principles

OBJ02 Implementing and evaluating LOUD standards in an explorative setting

OBJ03.1 Reporting how Yale has rolled out LOUD standards on a large scale

OBJ03.2 Verifying Linked Art and IIF consistency across the LUX platform

These objectives are explored extensively in three empirical chapters, offering a detailed insight into the operationalisation and impact of LOUD standards. In the following section, the methodology used to achieve these objectives is discussed in greater detail.

5.4 Methodology

This section is designed to rigorously address several pivotal questions that underpin the research approach, reflecting on the epistemological and ontological stances guiding this study. It explores the selection and application of data collection tools, detailing their relevance and the implications of their use, alongside the timing and demographic details of data collection. This section also looks at the analytical tools used, discussing their significance, implications and any ethical considerations that arose.

The methodology of this research is streamlined into three parts. This consolidation provides a holistic overview, starting with an outline of the research methods presented in each of the empirical chapters in subsection 5.4.1, through a combined exploration of data creation, collection, curation and modelling, to an examination of data analysis. The section concludes by addressing the overall limitations and ethical considerations of the research in subsection 5.4.2 and subsection 5.4.3, ensuring a careful and ethically sound approach to the study of LOUD.

5.4.1 Outline of the Research Methods

This subsection clarifies the multifaceted approach to data collection, creation, curation, modelling and analysis that is central to the empirical chapters that follow. By interweaving quantitative and qualitative methods, this research not only navigates the landscapes of cultural processes, but also bridges the micro-macro divide in the study of cultural constellations. Each empirical pursuit is dissected in detail, spotlighting the ways and means used to unveil the underlying dynamics at play.

At its core, this thesis seeks to unravel the intricate web of cultural processes and constellations through a lens that is both micro in perspective and rich in empirical evidence. It emphasises an actor and practice-centred inquiry, as articulated by [Wietschorke \[2014, p. 160\]](#). This study aligns with such a perspective, aiming to dissect and understand cultural phenomena through a detailed examination of specific cosmologies, like those of LOUD-driven communities, the PIA research project, and Yale's LUX.

The attempt to understand a particular cosmology requires immersion in the practices and dynamics of the communities at its heart. Inspired by [Latour and Woolgar \[1986\]](#), this research situates itself within the vibrant ecosystems of the IIF and Linked Art communities. As an active participant, the research reflects the intricacies of a laboratory where the fusion of practices unfolds, providing a rich tapestry of insights into the research objectives.

Embracing the role of a ‘praxiographer’²³¹, in the vein of [Mol \[2002\]](#), this exploration goes beyond mere observation. It addresses the ‘doing’ of research, emphasising the situated, embodied practices that shape knowledge production. This praxiographic lens foregrounds the complex, often tacit, processes through which the communities’ cosmologies are enacted and understood. It acknowledges that these practices are not just observed but are actively participated in, thereby crafting a nuanced narrative of the research landscape. This approach aligns with Mol’s advocacy for a more embodied and engaged form of inquiry, one that recognises the researcher as an integral component of the knowledge ecosystem, contributing to and being shaped by the evolving dynamics of community practices.

Recognising the transformative potential of digital methods, this study echoes [Venturini and Latour \[2009\]](#)’s call as well for a paradigm shift. It emphasises the need to fuse digital tools and data with traditional approaches, thereby enriching micro-analyses of interactions with insights into broader macro-structures. This not only challenges existing methodologies, but also invites a reappraisal of the dichotomies that have long governed the social sciences and humanities at large, and argues for a deeper engagement with the complexities of the digital sphere.

As outlined in Table 5.1, the empirical chapters are meticulously aligned with the research’s overarching axes, offering a panoramic view of the questions, hypotheses, and objectives steering this inquiry. This segmentation underscores the nuanced exploration of community practices, semantic interoperability, and their intersections, paving the way for a comprehensive understanding of the cultural and technological fabrics under investigation.

Table 5.1: Empirical Chapters: Research Axes and Scope Overview

Chapter	Empirical Research	Community Practices	Semantic Interoperability	Research Scope
Chapter 6	Assessment of Practices in IIF and Linked Art Communities	YES	To some extent	RQ01/HYP01 – OBJ01.1, OBJ01.2
Chapter 7	Deployment of LOUD within PIA	To some extent	YES	RQ02/HYP02 – OBJ02
Chapter 8	Review of a Large-Scale LOUD Implementation and Data Consistency within LUX	YES	YES	RQ03/HYP03 – OBJ03.1, OBJ03.2

Embarking on this research journey requires a versatile approach that encompasses data creation, collection, curation, modelling and analysis. This approach not only ensures a robust empirical foundation, but also fosters a dynamic interaction with the data, allowing new insights and perspectives to emerge. The following narrative will explore these dimensions, highlighting the central role of each in shaping the contours of this scholarly enterprise. Through this lens, the research not only contributes to academic discourse, but also navigates the challenges and opportunities presented by LOUD in the CH field.

Data Creation and Collection This includes an in-depth look at the types of data collected from different sources and the strategies adopted to create either new datasets or workflows. The process combines the collection of existing data relevant to the analysis of LOUD ecosystems and

²³¹ ‘Praxiography is a concept that describes the particular methodology of practice theory-driven research.’ [[Bueger 2019](#)]

the creation of new datasets to explore under-researched areas, thereby ensuring a strong empirical base for the study.

Data Curation and Modelling Data curation is essential for ensuring data integrity and usability, which are essential for reliable analysis and improved user satisfaction. This process involves the rigorous management of data, including essential components such as data cleansing, matching and conversion, to ensure that data remains accurate, accessible and meaningful for its intended use. Furthermore, data modelling is characterised as a rich process involving the definition of domain ontologies, the schema mapping, and the transformation of data into semantic graphs. This process is crucial for the effective integration and interpretation of data [Ciula and Eide 2016, Fafalios et al. 2023]. Nasarek [2020, p. 132] emphasises the crucial role of data modelling in facilitating insightful interpretations of reality and extending datasets for more in-depth analysis. In addition, Meunier [2017, pp. 40-42] highlights the widespread application of (conceptual) modelling within DH, stressing its necessity, stemming from the semiotic complexity of the field's diverse subjects, which conceptual modelling skilfully navigates, thus enabling a deeper understanding and exploration of interdisciplinary domains.

Data Analysis Drawing on Wietschorke [2014, p. 163], it emphasises the necessity of critical source evaluation through consistent criticism and heuristic classification, transforming historical fragments into meaningful sources by posing insightful inquiries. This foundational approach underscores the importance of evaluating sources within their historical and heuristic contexts, ensuring that the data analysis remains grounded in a comprehensive understanding of the source material. Incorporating the reflective process of considering the researcher's own position, as advised by historical ethnography, enriches the data analysis further. This methodology serves a dual purpose: it acts as an epistemological guide for achieving a dense description of actors, practices, and situations, and it aids in the reflection of the research process itself by acknowledging the researcher's situated knowledge [Wietschorke 2014, p. 164-165]. Such an approach not only deepens the cognitive objectives of the research but also enhances the transparency and reflexivity of the analytical process.

Research Outcomes This research aims to produce a range of outputs, from raw data to highly structured analyses. These outcomes are intended to serve not only as a testament to the empirical rigour applied, but also as a beacon for further scholarly inquiry. Raw outputs provide unmediated insights into the data collection and generation processes, while more structured outputs will provide deep dives into the nuances of LOUD ecosystems, underpinned by robust data curation, modelling and comprehensive analysis.

In the following three parts (sub-subsections 5.4.1.1, 5.4.1.2, and 5.4.1.3), dealing with the empirical research, a diagram is provided to illustrate the process from data collection or generation to the results that will be explored in later chapters. Figure 5.1 provides a visual template detailing the specific shapes and colours used to represent different stages and components of the research process. This schematic representation helps to fully understand the progression from initial data gathering to the analytical findings that form the basis of subsequent discussions.

5.4.1.1 Assessment of Practices in IIIF and Linked Art Communities

Chapter 6 engages with the social fabrics of IIIF and Linked Art, aiming to achieve two objectives through a mixed-methods approach. The research process and steps undertaken are detailed in Figure 5.2.

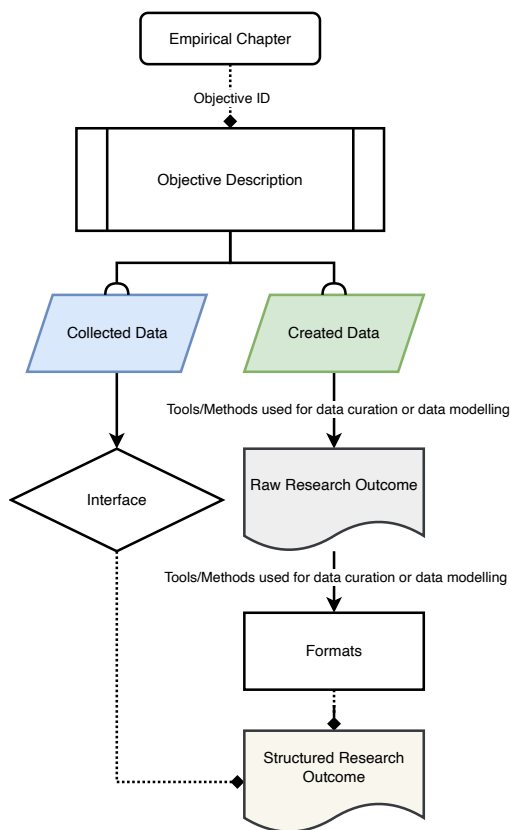


Figure 5.1: Research Outline Visual Template

For the first objective, which is to analyse capacity-building and advocacy initiatives within the IIF and Linked Art communities, data was collected from - generically referred to here as - guidelines on each initiative’s website, such as cookbook recipes to help create resources and principles, as well as meeting minutes produced in Google Docs and GitHub issues. The meeting minutes from Google Docs were extracted with the help of a tool and a series of Python scripts. As for the guidelines, a matrix was prepared for the qualitative data, while the more quantitative data was extracted using the GitHub API. A first analysis was performed by aggregating and structuring all the data in CSV files according to the levels of abstraction discussed in the previous chapter, here in terms of category (e.g. viewer, type of support, JSON similarity, etc.). Further analysis was carried out using NetworkX²³², a sophisticated Python library designed for the analysis of networks between social entities and their information exchange [Hagberg et al. 2008]. This tool allows the generation of visual representations that aim to achieve the goals of Social Network Analysis (SNA). SNA is understood here more as an analytical goal realised through visual means than as a methodological tool [see Akhtar 2014]. It complements the theoretical framework based on ANT, enriching it with a quantitative and visual approach to studying the interconnectedness of actors and the influence of these connections on the distribution of information and resources. By incorporating SNA, this research seeks to uncover the rich network of relationships and patterns between actors within LOUD ecosystems, with visualisations that can be effectively realised through tools such as Gephi²³³, a network analysis and visualisation software [Bastian et al. 2009]. For GitHub issues, a case study was conducted to show how an issue going through the IIF review process works.

²³² NetworkX: <https://networkx.org/>

²³³ Gephi: <https://gephi.org/>

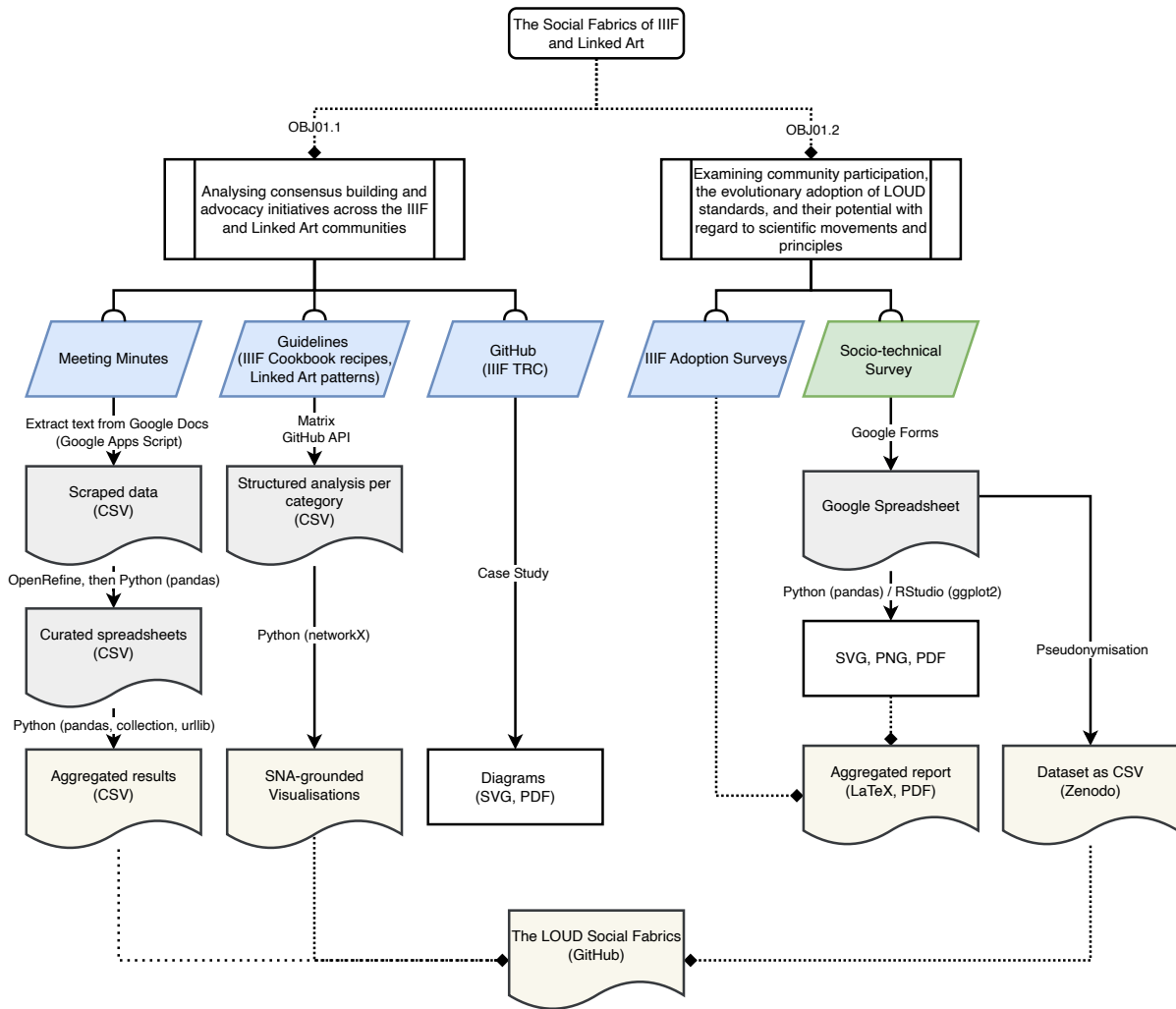


Figure 5.2: Research Outline for Assessing the IIF and Linked Art Practices

As for the second objective, it consists of analysing the adoption surveys that have been carried out within the IIF community over the years on the adoption of its specifications and creating my own survey on LOUD’s socio-technical characteristics. For the latter, Google Forms was used and the raw data has been made available after a pseudonymisation stage. Graphs and statistics were created using Python, in particular using its pandas library²³⁴, as well as employing RStudio²³⁵ to produce various files for publication in a final report.

Finally, the main scripts and outcomes of this empirical research were made available through a GitHub repository²³⁶.

5.4.1.2 Deployment of LOUD within PIA

The deployment of LOUD standards and the underlying tools and software required to implement them within the PIA project is covered in Chapter 7. The implementation and analysis process has a single objective, captured in Figure 5.3.

²³⁴ pandas – Python Data Analysis Library: <https://pandas.pydata.org/>

²³⁵ RStudio: <https://posit.co/download/rstudio-desktop/>

²³⁶ The LOUD Social Fabrics: <https://github.com/julsraemy/loud-socialfabrics>

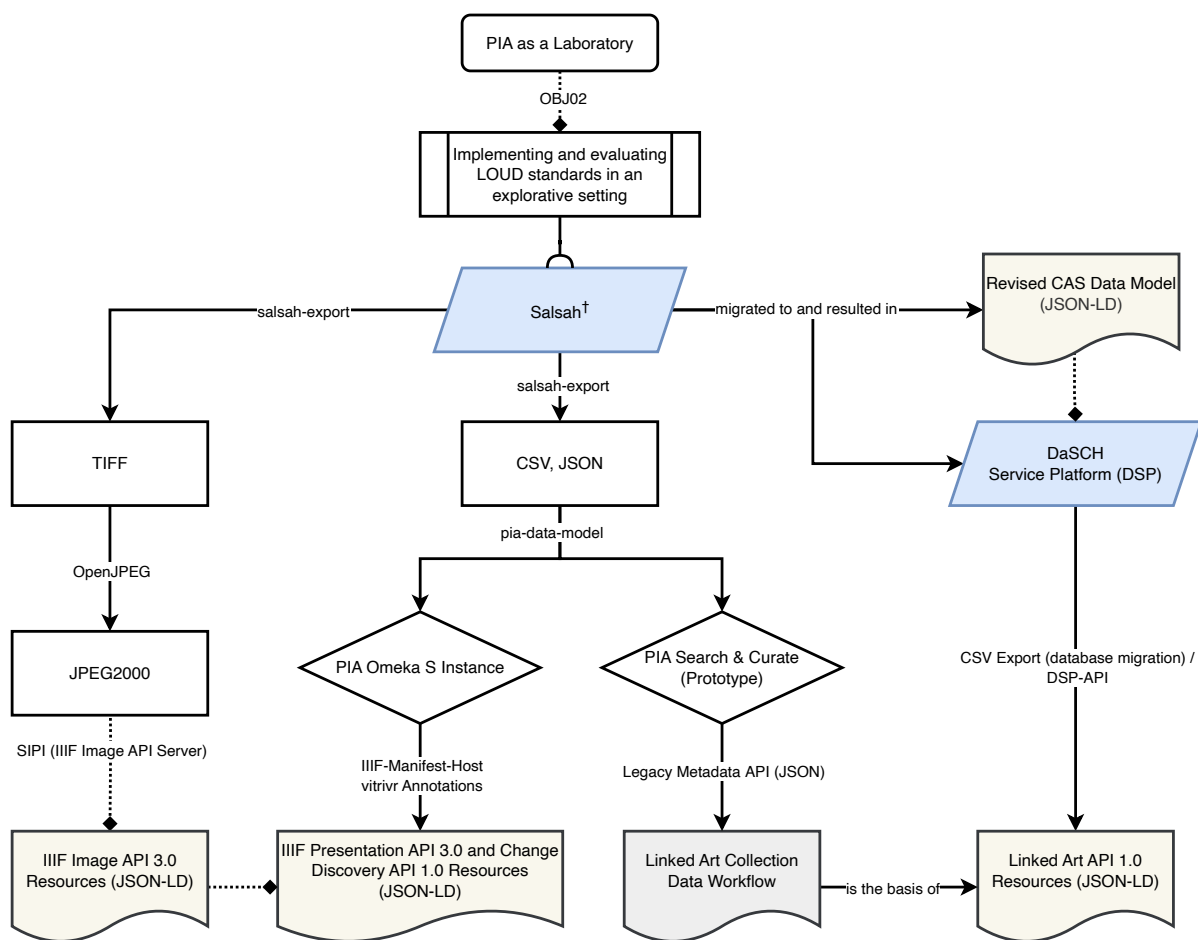


Figure 5.3: Research Outline for Deploying LOUD Standards in the PIA Project

In keeping with the chapter’s title, it describes a laboratory effort focused on creating and transforming data into resources that conform to LOUD standards. Initially, data was extracted from Salsah²³⁷, a digital platform developed by DaSCH that stores images and metadata from the CAS photographic archives. This involved using a Python script to extract CSV, JSON and TIFF files. The process required specific image conversion steps – TIFF to JPEG2000 – to facilitate their distribution via SIPI, a IIIF Image API server developed by the DHLab. In addition, the IIIF Presentation API 3.0 and Change Discovery API 1.0 resources were generated by an early-stage micro service. In parallel with the development of Linked Art data, a workflow prototype²³⁸ was created in collaboration with members of the University of Oxford, based on a preliminary JSON API designed for the PIA project. At the same time, a revised data model was developed to ensure compatibility with the requirements of DSP, setting the stage for the migration of (meta)data from Salsah. This foundational work enabled the subsequent updating of resources to conform to the Linked Art standard, taking advantage of the established workflow.

5.4.1.3 Review of a Large-Scale LOUD Implementation and Data Consistency within LUX

The aims of Chapter 8 are to analyse how the Yale University has rolled out the LOUD specifications as part of LUX, Yale Collections Discovery, and also to assess the consistency of Linked Art and IIIF resources within the LUX platform. These two objectives are depicted in Figure 5.4.

²³⁷ Salsah: <https://web.archive.org/web/20240229093041/https://www.salsah.org/>

²³⁸ Linked Art Collection Data Workflow: <https://github.com/tgra/Linked-Art-Collection-Data-Workflow>

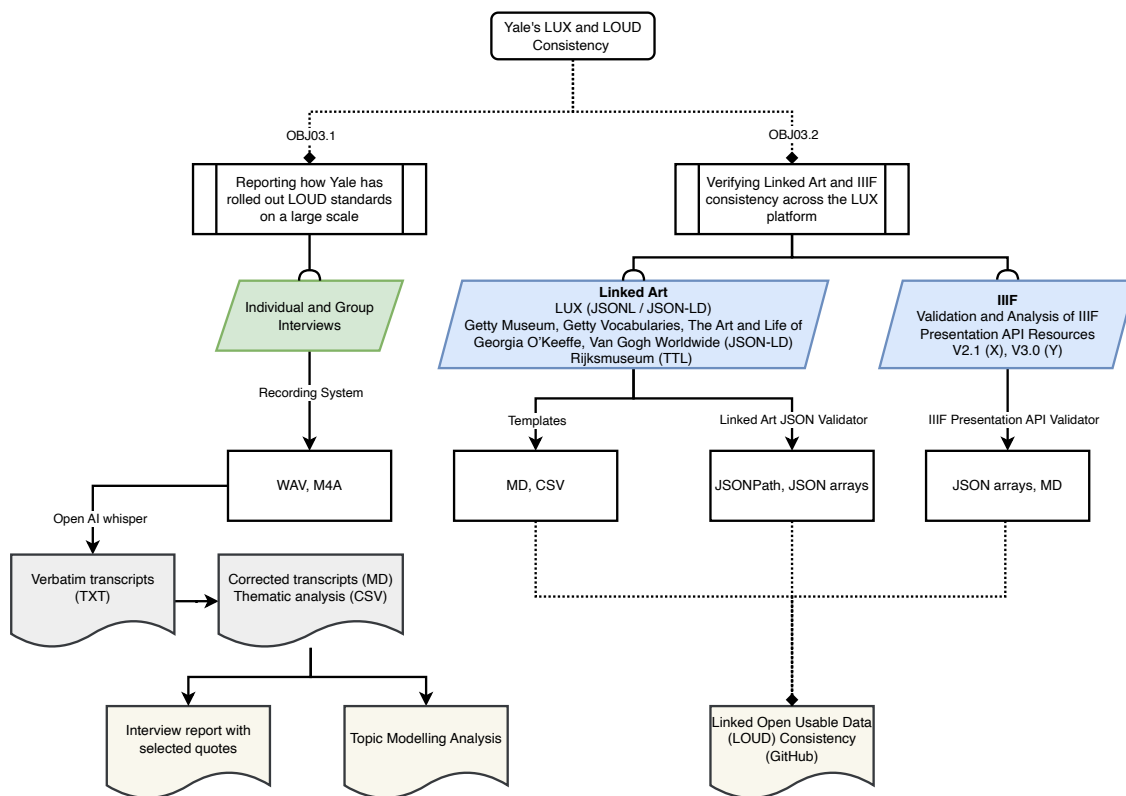


Figure 5.4: Research Outline for Reviewing LOUD Implementation and Data Consistency within LUX

The first objective was to document the collaborative efforts of Yale employees involved in the development of LUX. To achieve this, I conducted both group and one-on-one interviews and recorded these discussions. These recordings were first transcribed using Whisper²³⁹, an advanced speech recognition tool developed by OpenAI. To ensure accuracy, I meticulously checked and amended the transcriptions. I then performed thematic analysis on the transcripts, now formatted as CSV files, and consolidated the revised verbatim transcripts into a Markdown document. This extensive process provided the basis for writing a report comprising selected quotes, as well as an analysis using simple topic modelling to extract key themes [see Benabdelkrim et al. 2020], which provides distant reading insights.

The second objective was to verify the consistency of Linked Art and IIF resources. Once the data had been collected, an effort was made templating the identified patterns by selecting various resources qualitatively and validating them quantitatively using a JSON validator. For IIF resources, the Presentation API Validator²⁴⁰ designed by IIF-C was leveraged. All of this data, along with the scripts, was compiled and analysed onto a GitHub repository²⁴¹.

5.4.2 Overall Limitations

The overall limitations of this thesis cover several key areas, reflecting the constraints encountered in the study of LOUD.

The extent of initiatives and infrastructure supporting LOUD standards is another critical limitation. The variability in support and implementation across platforms and projects may affect the

²³⁹ Whisper: <https://github.com/openai/whisper>

²⁴⁰ IIF Presentation API Validator: <https://presentation-validator.iif.io/>

²⁴¹ LOUD Consistency: <https://github.com/julsraemy/loud-consistency>

generalisation of the study's findings, suggesting the need for broader engagement and standardisation efforts amongst CH practitioners. Moreover, the influence of platforms' formatting power on the dissemination and reception of LOUD standards cannot be overlooked [Boullier 2018], highlighting the need for a nuanced understanding of how platform-specific affordances shape scholarly communication and knowledge exchange. Moreover, according to McCarty [2023], at '*a low-tech end of scholarship*', the integration of LOUD within DH faces challenges despite its seemingly low-tech orientation in practice and tool adoption. This situation echoes McCarty's discussion of the need for a '*combinatorial habit of mind and machine*', suggesting that while some of the tools developed by the wider IIF community are accessible, they still require a significant level of technical expertise. It emphasises the importance of harmonising traditional scholarly methods with digital tools to enrich humanistic research, and urges a critical evaluation of the limitations and potential of digital tools to ensure that they complement rather than constrain scholarly inquiry.

Obtaining data from individuals and their CHIs through generic investigation echoes Bourdieu [1988]'s theory of fields²⁴² and therefore may not capture the full socio-technical landscape of LOUD ecosystems. This approach may overlook the wider range of actors and their interactions that are crucial for a holistic understanding of the impact of LOUD.

The implementation of LOUD standards and practices within the PIA project served as an experimental ground for testing the semantic model and data reuse in a participatory manner, revealing the gap between theoretical constructs and their practical implementation. This discrepancy underscores the challenges of deploying layered data models in real-world interfaces, and highlights potential limitations in empirical validation and UX insights.

Digital methods, while instrumental, may not fully capture the nuanced diffusion of ideas, a central tenet in ANT where non-humans are given agency. This gap points to the need for methodological innovations that can better account for the dynamics of idea diffusion within and across networks. In addition, the '*zooming effect*' introduces methodological constraints, demonstrating that the depth of analysis is often limited by the computational resources available. This limitation affects the scale at which data can be analysed, potentially distorting the understanding of network dynamics and interaction patterns [see Boullier 2018].

In summary, while these identified areas represent the primary limitations of this research, it is important to acknowledge that the range of limitations may not be exhaustive. The intrinsic multilayered nature of LOUD necessitates ongoing review and negotiation across modes of research that bridges theoretical frameworks with empirical realities.

5.4.3 Ethical Considerations

Moving on from exploring general limitations, it is equally essential to address the ethical considerations inherent in undertaking such research. The LOUD community, predominantly composed of members from the Global North and characterised by a lack of diversity in terms of gender and ethnicity, presents specific ethical challenges. This composition raises concerns about inclusivity and representation, and points to the need for conscious efforts to ensure that the development and application of LOUD standards do not inadvertently perpetuate existing inequalities or biases.

²⁴² Bourdieu's theory of fields conceptualises social life as a series of distinct but interrelated arenas or 'fields', each with its own rules, structures and forms of capital. Actors within these fields compete for dominance, status and resources, guided by both their 'habitus' and the specific capital they possess.

Furthermore, the dominance of privileged circles within the LOUD community underscores the importance of adopting ethical research practices that actively seek to amplify underrepresented voices. Engaging with a wider range of stakeholders, including those from marginalised communities, is crucial to fostering a more inclusive and equitable DH and CH landscape. This approach not only enriches research with diverse perspectives, but also aligns with broader ethical commitments to social justice and equality.

Ultimately, ethical considerations extend to the methods used to analyse and interpret data within LOUD ecosystems. Ensuring transparency, respecting privacy and acknowledging the limitations of one's own research stance are fundamental to ethical scholarship. By addressing these ethical dimensions, the thesis aims to contribute thoughtfully and responsibly to the ongoing discourse around LOUD in the heritage domain and to advocate for practices that promote diversity, inclusivity and equity.

While this thesis navigates the realities of LOUD for CH and addresses overarching limitations and ethical considerations, it's paramount to acknowledge the inherent privilege of conducting such research as PhD Candidate in DH in Switzerland. While this work aims to highlight and mitigate issues of inclusivity and representation, the ability to address all facets of marginalisation is limited. This recognition serves as a basis for advocating for a more inclusive and equitable approach in future research endeavours.

5.5 Contribution to the (Digital) Humanities

My thesis aims to contribute to the humanities, increasingly digital, by situating itself at the intersection of DH and STS. Addressing the disillusionment with digital methods identified by [Venturini et al. \[2015, 2018\]](#), my research addresses issues such as the overly narrow yet ambitious notions of digital traces, the fluctuating levels of awareness and confidence in their production, the misconception of equating digital methods with mere automation, and the misplaced hope of easily applying traditional quantitative methods to digital traces.

CH resources, documented and preserved by cross-border memory institutions, are increasingly interconnected through digital platforms. The role of web architecture, in particular LOD technologies and henceforth standards conforming to LOUD design principles, is central in this context. The incorporation of shared methodologies, as seen in the IIF community, exemplifies the socio-technical interactions that are central to the scholarly and heritage domains. These community-driven technologies facilitate not only the description and contextualisation of CH records, but also their integration into scholarly discourse, thus underlining the vital role of community practices and semantic interoperability in improving accessibility, interoperability and engagement within the humanities.

In an effort to transcend the conventional boundaries of academic dissemination, I have decided to produce both a PDF version of the thesis intended for the university's repository and an HTML version of the thesis²⁴³, and I plan to make the structured data of the dissertation available via Linked Art API endpoints, time permitting. This endeavour not only embraces the ethos of open scholarship and self-publishing as discussed by [Capadisli \[2020\]](#), but also ventures into experimental territories of knowledge representation. By considering diverse and niche use cases, this work seeks to extend the methodological framework of DH and encourage a broader and more nuanced engagement with LOUD.

²⁴³ LOUD for Cultural Heritage – PhD Thesis by Julien A. Raemy: <https://phd.julsraemy.ch>

6

The Social Fabrics of IIF and Linked Art

This chapter unfolds as a foray into the social fabrics of the IIF and Linked Art communities, and represents the first of three empirical chapters within this doctoral thesis. In this endeavour, I seek to disentangle the many threads that make up these communities, with the aim of identifying the essential strands that contribute to their collective weave. Drawing on [Haraway 2016b, p. 116], this investigation explores the extent to which their fabric encompasses an entanglement of ‘*people, critters and apparatuses*’ [Haraway 2016b, pp. 128-129], examining dependencies as highlighted by [Latour 2022, p. 71]. In doing so, I attempt to shed light on the interplay of relationships that form the backbone of IIF and Linked Art, unveiling the complex tapestry of collaboration and interconnectedness that defines them.

With respect to scope, this chapter considers the socio-technical dynamics required to produce specifications that are aligned with the LOUD design principles. The central hypothesis is that although grassroots efforts to forge active communities require substantial and sustained commitment, these efforts typically succeed in the medium to long term. The analysis focuses on how consensus building and advocacy within the IIF and Linked Art communities promote the adoption of LOUD. It also examines the role of community participation in the evolutionary adoption of these standards.

Most of the (technical) documentation related to this empirical investigation, such as Python scripts, raw data, and diagrams, is made accessible through GitHub²⁴⁴. This open repository serves not only as a resource for replicating and verifying the findings but also exemplifies a commitment to transparency, in keeping with the ethos of the IIF and Linked Art communities.

The chapter begins in Section 6.1, where I delve into some of the foundations of the IIF and Linked Art communities. The first objective, concerning consensus building and advocacy within these digital communities, is addressed in more detail in Section 6.2. This section unpacks the collaborative strategies and initiatives that are central to promoting standards adoption and community cohesion. Following this, Section 6.3 focuses on the second objective by analysing community participation and adoption of LOUD standards.

In Section 6.4, the chapter synthesises the findings from the previous analyses, weaving together the threads of the research to provide a comprehensive overview of the terrain explored. The chapter concludes in Section 6.5, where reflections and forward-looking perspectives consider the wider implications of the study’s findings for the future of the IIF and Linked Art communities, anticipating the challenges and opportunities on the horizon.

²⁴⁴ The LOUD Social Fabrics: <https://github.com/julsraemy/loud-socialfabrics/>

6.1 Exploring the Foundations

The core objective of the IIIF and Linked Art communities is the creation and maintenance of specifications. Through these specifications, which act as mediators, there is the potential for a harmonious symphony of technology and purpose. At the heart of this ecosystem is the API, mediated by the JSON-LD's API `@context`, which provides clarity and cohesion to the standards being developed and implemented. Validators ensure compliance with the specifications, with software conforming to these standards and encapsulating the layered abstractions that define the technological landscape.

For this potential to be fully realised, however, it is imperative that those within these communities possess resources that not only comply with these shared APIs, but also embody them. This embodiment is facilitated by an underlying infrastructure that includes servers, scripts, micro-services and other technological components, as well as viewers, players and browsers that render compliant (meta)data.

In this respect, while specifications serve as the foundational goal, the broader vision of achieving harmony through technology and purpose requires the active participation of community members. They are tasked with fostering an ecosystem where specifications are not only developed and maintained, but brought to life through practical implementation. This requires a commitment to true collaboration within and across community boundaries, and a concerted effort to implement best practices, especially in user interface design. Such a commitment ensures that the integration of vision and execution is not just a theoretical ideal, but a lived reality that reflects the shared aspirations of both IIIF and Linked Art and their respective constituencies.

The complexity of the IIIF community provides a rich area for exploration, in contrast to Linked Art, where community interactions are more streamlined and occur primarily through fortnightly calls²⁴⁵. This complexity is vividly illustrated in Figure 6.1²⁴⁶, which describes the range of entities within the framework. The decision to focus on IIIF is motivated by its rich constellation of committees, working groups and supporting institutions, providing a deeper dive into the development of digital standards in a robust community setting.

In the diagram, a colour scheme helps to identify the different components of the IIIF ecosystem: entities in purple represent people, resources and institutions that are part of the IIIF-C. Light pink highlights the committees associated with the consortium, reflecting the structural elements that support the governance and operational dynamics of the consortium. Soft blue denotes the community committees, including the CoCo and the Editorial Committee, which are central to guiding community initiatives and maintaining specifications. Green identifies the community and TSGs, highlighting the collaborative spirit that drives the development of IIIF. Finally, yellow identifies essential resources and tools, such as the specifications or cookbook recipes, highlighting the practical outputs that enable the adoption and implementation of the APIs.

This visual representation serves not only as a guide to the organisational structure of the IIIF community, but also as a representation of the vibrant interaction between disparate entities. It highlights the integral role of both the formal IIIF-C bodies and the wider community in fostering a culture around interoperability, facilitating collaboration and sharing best practice. This comprehensive approach to community engagement and standards development underscores the concerted effort needed to move the digital scholarship and CH fields forward. The interconnectedness depicted underscores

²⁴⁵ In the earlier days of the IIIF community, there was much more fluidity in terms of roles and responsibilities, especially within the Editorial Board, something Linked Art is still experiencing.

²⁴⁶ This figure is also available in an enlarged format in Figure B.6.

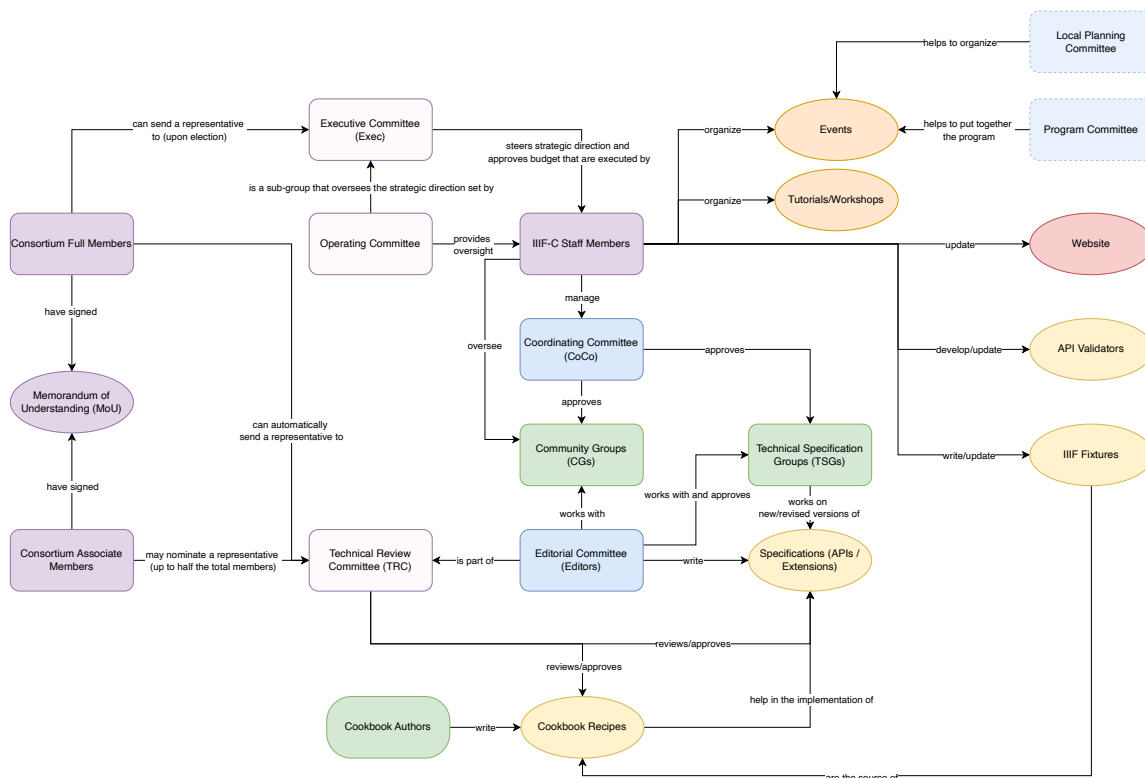


Figure 6.1: Overview of the IIIF Community

the role of essential platforms, facilitating their evolution. For instance, both the IIIF and Linked Art communities maintain accessible Google Drives, where meeting minutes on Google Docs, often linking to GitHub issues, underscore the technical discourse within these initiatives. In addition, the use of dedicated Slack workspaces for continuous conversations further reinforces the importance of regular engagement.

The following sections explore in more depth the processes of consensus building and the advocacy initiatives as well as the engagement of the community that not only promote adherence to specifications, but also encourage best practices. Such endeavours are foundational to fostering a LOUD ecosystem characterised by community-driven commitment to a few standards, yet embracing a multitude of implementations. This rich tapestry of connected knowledge and practices ensures the digital heritage domain is both universally accessible and deeply engaging.

6.2 Consensus Building and Advocacy Initiatives

The analysis examines the mechanisms of capacity building and advocacy within the IIIF and Linked Art communities, focusing on three key areas.

Initial research centres on subsection 6.2.1, exploring the different types of meetings that play a crucial role in both community engagement and the wider collaborative process, with illustrations from the IIIF Discovery TSG and Linked Art meeting minutes. Attention then shifts to subsection 6.2.2, highlighting the role of GitHub in facilitating code and use case documentation, particularly within TRC, as a testament to the community’s collaborative spirit and commitment to transparency. Finally, subsection 6.2.3 examines community guidelines and best practice efforts, discussing how each community provides resources that transcend the standards, such as the IIIF cookbook recipes and Linked Art patterns, to guide and encourage further community contributions and foster interoperability.

6.2.1 Types of Meetings and Engagement Opportunities

Meetings within the IIF and Linked Art communities are essential for collaboration, sharing insights and steering the direction of projects and standards. These predominantly virtual meetings are orchestrated by designated chairs and follow structured agendas to maximise productivity and inclusivity. IIF meetings vary in frequency - fortnightly, monthly or quarterly - reflecting the dynamic needs of different parties. Linked Art, which maintains a consistent rhythm, meets fortnightly. Each meeting begins with a roll call, which allows participants to register their presence or send their regrets in advance, a gesture that underlines their interest even if they are absent. To further enhance the collaborative nature of these discussions, one or more participants are asked to take the role of note-taker for each meeting.

A foundational element of these meetings is the ongoing, mostly asynchronous, collection of use cases, primarily facilitated through GitHub. Although the initial sessions heavily focus on gathering these use cases, the practice continues across all stages of development and discussion. Each meeting is structured around a predefined agenda, ensuring focused and productive sessions that typically span one hour. Additionally, meetings are announced at least a day in advance. This process ensures that discussions about the implications of use cases and their potential integration into documents – such as specifications, early drafts or recipes – are grounded in a broad and inclusive collection of community inputs, pending subsequent approval or implementation.

In addition to virtual sessions, face-to-face meetings provide invaluable opportunities for deeper engagement. Within IIF, the annual conference acts as a catalyst for stakeholders to meet – whether as part of the official programme or in parallel sessions. In addition, ad-hoc in-person meetings have also been instrumental in driving significant progress, such as the inclusion of audiovisual resources in the Presentation API 3.0 and the ongoing efforts to include 3D materials in the forthcoming version of this specification.

The Linked Art initiative, although newer, has held five major face-to-face meetings, three before the COVID-19 pandemic and two in 2023. The fourth meeting, held in New Haven, CT, USA, focused on resolving outstanding issues to finalise the first version of the standard. The practical implications of developing the Linked Art API, alongside the deployment of the LUX platform, illustrate the tangible outcomes of these meetings. Many discussed issues were long-standing, exacerbated by the pandemic. For instance, discussions around rights statements²⁴⁷ and API Displacements and/or Transclusions²⁴⁸ have highlighted the evolving requirements and challenges faced by the community. It also became clear that discussions on identity – encompassing nationality, gender, and ethnicity – had matured, enabling new insights and considerations. These discussions are delicate, often challenged by inherent biases and assumptions in metadata assignment, highlighting the need for a more inclusive and diverse approach to identity representation within CH.

Furthermore, while discussions about extending the CIDOC-CRM hierarchy to include new properties or classes - such as the **Transfer** class and the inclusion of non-human actors - take place within the broader CRM-SIG, Linked Art approaches these considerations from a different perspective²⁴⁹. Unlike CRM-SIG, where ontological purity is a major concern, Linked Art prioritises more practical applicability and flexibility in modelling CH data. This leads to decisions where Linked Art has, for pragmatic reasons, created its own classes and leveraged existing classes designed for human entities to encompass non-human actors. This approach, while departing from strict ontological standards,

²⁴⁷ Do we have a documented example for expressing image rights?: <https://github.com/linked-art/linked.art/issues/311>

²⁴⁸ API Displacements and/or Transclusions: <https://github.com/linked-art/linked.art/issues/411>

²⁴⁹ Inject classes/properties into the CRM hierarchy?: <https://github.com/linked-art/linked.art/issues/479>

aims to address immediate modelling needs within the community, recognising that it may occasionally compromise semantic interoperability for the sake of broader and more inclusive data representation practices.

Transitioning from these theoretical and practical considerations, I delve into detailed case studies of community interactions as this section explores specific series of meetings by looking at the notes taken during them – in subsection 6.2.1.1 meetings from the IIIF Discovery TSG²⁵⁰, a completed IIIF group, and in subsection 6.2.1.2 the ongoing fortnightly meetings of the Linked Art initiative.

This exploration provides a lens into the number of meetings, the people involved and the outcomes that shape the evolving landscapes of both communities. Meeting minutes from both groups were extracted using a tool developed by Bainbridge [2021] to streamline the process of extracting texts from Google Docs into organised spreadsheets. This data then underwent a cleaning and curation process to ensure accuracy and consistency for analysis and visualisation²⁵¹.

6.2.1.1 IIIF Discovery TSG Meetings

The IIIF Discovery TSG formally ended its work in summer 2022, but continued to meet several times until March 2023. The purpose of the Discovery TSG was to address the critical need for interoperable resources to be easily discoverable. Recognising the importance of standardised discovery processes, the group aimed to develop specifications to improve the findability and reusability of IIIF resources. This initiative focused on adopting existing techniques and tools, facilitating their widespread use within the community, and supporting the creation of infrastructure such as registries. Key deliverables from the group comprise the Change Discovery API and the Content State API.

In addition to these APIs, the TSG focused its efforts on producing essential documentation and support materials to facilitate the wider adoption and implementation of these standards. This includes the creation of a Registry of (Metadata) Profiles²⁵², which serve as a mechanism for discovery-related metadata profiles to leverage through the `seeAlso` property in IIIF Manifests. A note on Search Engine Optimization (SEO) guidelines was also produced²⁵³, with a view to improving search engine ranking of web pages containing IIIF resources.

Figure 6.2²⁵⁴ shows that 84 people, represented as individual horizontal bars, attended a total of 101 Discovery TSG meetings. Each person attended an average of 10.49 meetings, but the median attendance was 2. This illustrates a significant imbalance in participation: a small core group was highly active and contributed to the majority of discussions, while a larger number of participants attended less frequently.

A closer look at meeting attendance over the years reveals a gradual decline in participation, providing insights into the operational dynamics of the Discovery TSG and engagement trends. Initially, in 2017, there were 20 meetings with 42 unique participants and an average attendance of 12.35 per meeting. In 2018, while the number of meetings increased to 22, the number of unique participants decreased to 35 and the average attendance decreased to 8.73.

²⁵⁰ IIIF Discovery TSG: <https://iiif.io/community/groups/discovery/>

²⁵¹ Cf. The LOUD Social Fabrics – Meeting Minutes: https://github.com/julsraemy/loud-socialfabrics/tree/main/01_consensus-advocacy/meeting-minutes

²⁵² Registry of Profiles: <https://iiif.io/api/registry/profiles/>

²⁵³ Thinking about IIIF and SEO: <https://guides.iiif.io/IIIF-and-SEO/>

²⁵⁴ This figure is also available in an enlarged format in Figure B.7.

In 2019, the number of meetings decreased to 18 with 26 unique participants, and the average attendance further decreased to 7.78. The trend continued in 2020 with 15 meetings, 20 unique participants and an average attendance of 8.2. In 2021, there were 17 meetings with a slightly higher pool of 23 unique participants and an average attendance of 7.94.

The most significant decline occurred in 2022, when only 6 meetings were held, with 13 unique participants and an average attendance of 5. This decline continued in 2023, with only 3 meetings, 7 unique participants and an average attendance of 4.67. This significant reduction in engagement can be explained by the fact that the majority of the group's objectives and outputs had been achieved by the summer of 2022, leading to a natural reduction in the need and frequency of meetings as the project neared completion.

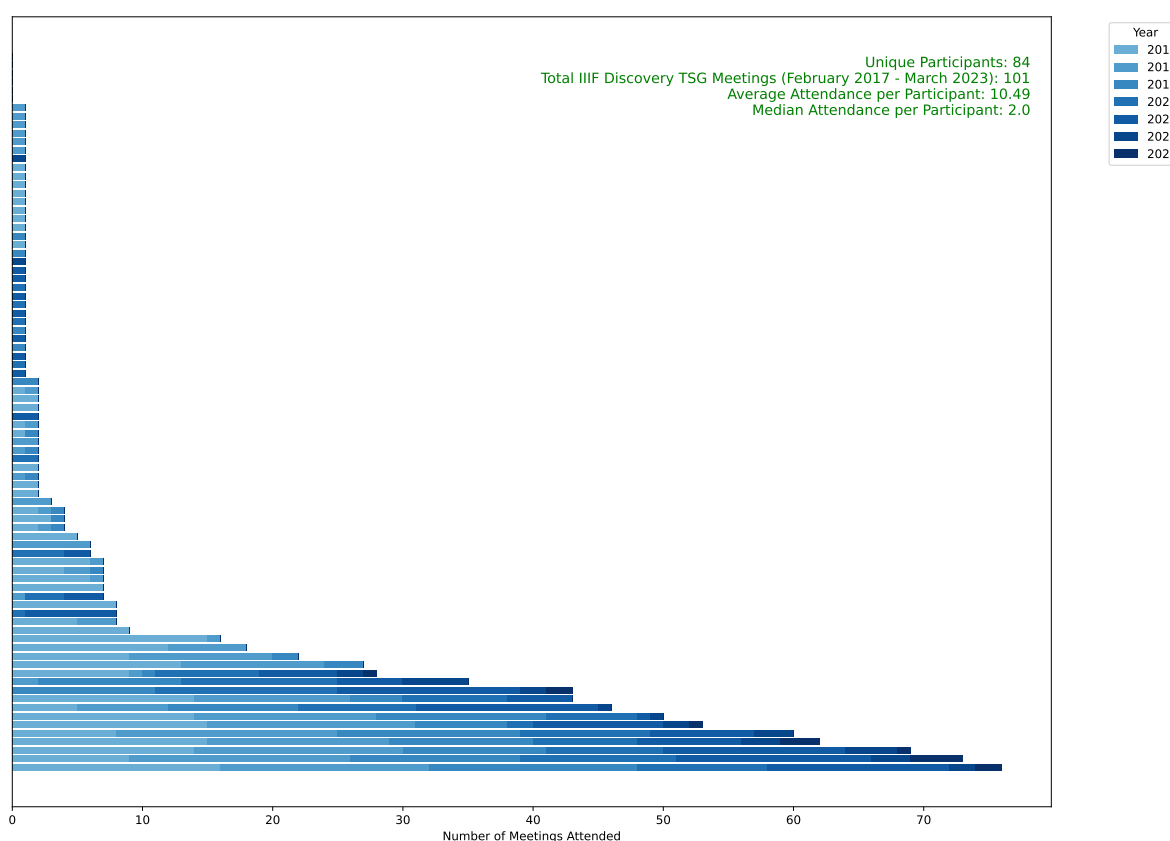


Figure 6.2: Stacked Bar Chart Showing IIIF Discovery TSG Meeting Attendance

The data also highlights the centrality of `github.com` in discussions and collaborations, as evidenced by over 400 hyperlinks to GitHub issues, pull requests and other GitHub-related content mentioned during meetings. In addition to GitHub, other frequently referenced domains include `iiif.io` with 169 mentions and `docs.google.com` with 139 mentions, reflecting the use of official documentation and shared documents for referencing standards. In addition, `www.w3.org` was mentioned 60 times, demonstrating its relevance for W3C standards aligned with IIIF practices. This frequency underscores GitHub's essential role as a central platform for facilitating communication and project development within the IIIF Discovery TSG, alongside these other critical resources.

As the Discovery TSG wrapped up its active phase, its contributions to the IIIF community proved significant, setting critical standards for interoperability. However, the full impact of its work remains to be seen. To date, only a limited number of institutions have implemented the Change Discovery API,

and while some viewers support the basics of the Content State API, the full intent of the standard has been realised predominantly through demo clients. This scenario underscores the ongoing challenge within the community: not only to develop these capabilities, but also to deploy and use them widely.

6.2.1.2 Linked Art Meetings

Now transitioning into the Linked Art meeting minutes, some parallels with the IIF Discovery TSG meetings can be observed, yet with distinct dynamics reflective of its ongoing activity. This analysis provides a snapshot of the Linked Art meetings from January 2019 to March 2024. The group convenes fortnightly to deliberate on refining the model and the API, with a strong emphasis on releasing the first stable version and integrating the patterns and specification into various platforms.

Figure 6.3²⁵⁵ indicates participation trends over this period, showing that 130 different individuals attended the 115 meetings held, predominantly in a virtual format, with five of these meetings being face-to-face²⁵⁶. Each participant attended an average of 13.57 of sessions, but the median attendance was only 2, indicating a very long tail of participation. This suggests a core group of highly engaged members who contribute consistently, while a larger number of participants engage more sporadically. This distribution highlights the challenges and successes of sustaining engagement over time, particularly in a volunteer-driven context where continuous engagement can be difficult to maintain.

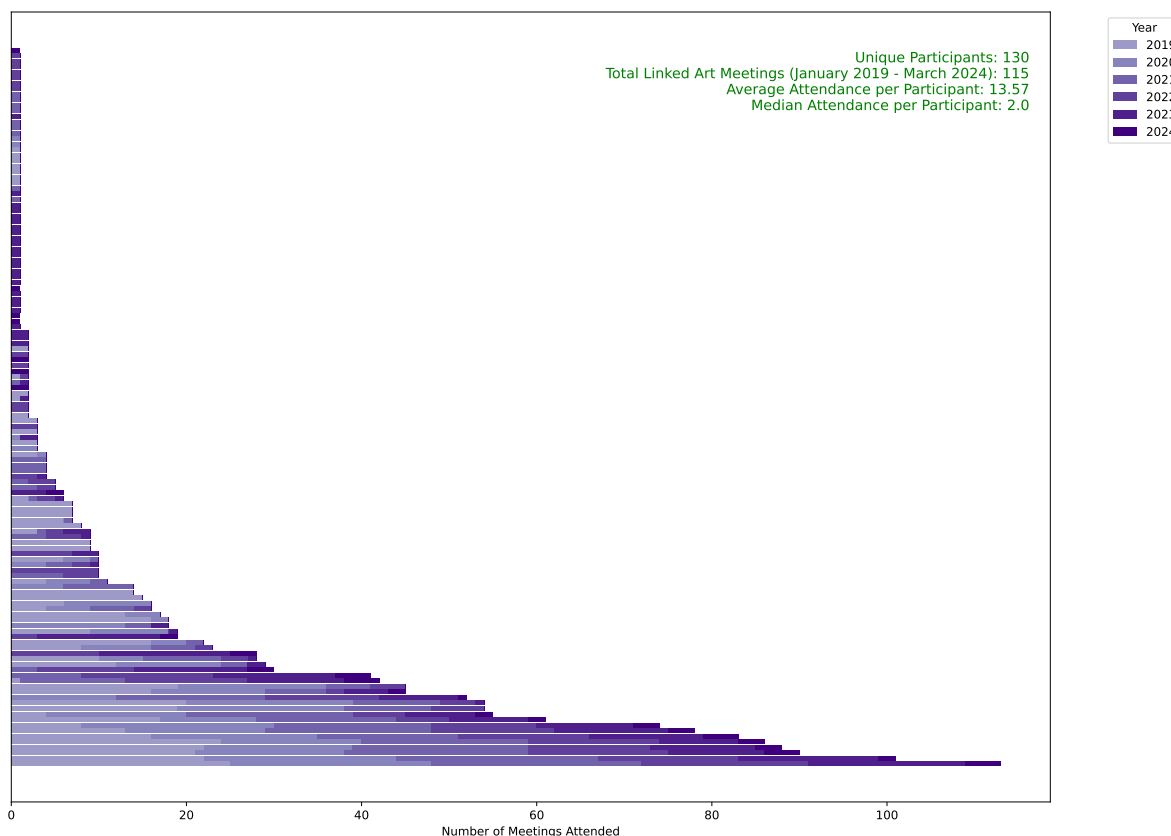


Figure 6.3: Stacked Bar Chart Showing Linked Art Meeting Attendance

²⁵⁵ This figure is also available in an enlarged format in Figure B.8.

²⁵⁶ For one of these face-to-face meetings, documentation was produced daily over four days, resulting in four distinct meeting records. Therefore, while only five face-to-face gatherings occurred, they account for eight entries in the total of 115 meetings documented.

A detailed examination of meeting attendance over the years for the Linked Art initiative reveals an active yet fluctuating participant engagement. In 2019, the year began with 26 meetings and included 53 unique participants, each attending an average of 17.88 sessions. This high level of involvement suggests robust initial interest and active participation in the foundational phases of the initiative.

However, subsequent years showed a slight decrease in engagement: 2020 hosted 23 meetings with 34 unique participants and an average attendance dropping to 15.48. In 2021, the meetings slightly increased to 25 with 48 unique participants, averaging 13.92 attendees per session. The year 2022 saw 19 meetings with 46 participants and an average of 14.74, showing a stabilisation in engagement levels despite fewer meetings.

The year 2023 continued with 18 meetings, and an increase in unique participants to 57, suggesting a widening of the community base, with an average attendance of 14.11 per meeting. The early data for 2024 shows 4 meetings with 25 unique participants and a relatively high attendance rate of 15.25, indicating sustained, albeit selective, engagement.

The data also underscores the significance of `github.com` as a pivotal platform for discussion and collaboration within the Linked Art meetings, with over 431 links pointing to GitHub highlighting its use for project management and technical discussions. Other frequently mentioned domains include `docs.google.com` with 35 references and `linked.art` with 44 mentions, reflecting their roles in documentation and project-specific discussions, respectively. Domains such as `www.getty.edu` and `vocab.getty.edu` were also significant, cited 10 and 9 times respectively, indicating the integration of controlled vocabularies in standardising terminologies within the Linked Art context.

While the Linked Art group originally had optimistic goals of releasing the first stable version of the specifications by 2020 and then 2021, the actual release is now expected to occur in 2024. This extended timeline, while longer than originally planned, has had an unforeseen benefit: it has allowed various platforms to experiment more extensively with implementing the evolving model, even in the absence of a finalised standard. Such experimental implementations have not only helped to refine the specifications through practical application, but have also fostered a richer understanding among actors of how the model can be adapted and used. Although Linked Art focuses primarily on museum artworks and their relationships, treating them as first-class citizen resources within the digital representation sphere, there is an observable shift towards embracing the broader GLAM domain. This broadening of scope suggests significant potential for Linked Art to impact a wider range of CHIs, improving cross-domain semantic interoperability while maintaining a hands-on orientation.

6.2.2 Collaborative Development through GitHub Documentation

Within IIF and Linked Art, GitHub has solidified its position as a central hub for collaboration across both communities, further illustrating the pragmatic approach adopted towards research and development [see Scroggie et al. 2023]. This preference for GitHub, complemented by the use of universally accessible tools like Google Drive for documentation, illustrates a pragmatic shift towards platforms that facilitate broad participation and collaboration. Such practices are integral to the collaborative process, enhancing transparency, and efficiency in addressing and resolving issues [Escamilla et al. 2023].

Figure 6.4 illustrates the journey from a conceptual use case to its technical realisation within both the IIF and Linked Art communities, highlighting the shared methodologies despite the different governance structures. The diagram begins with the community's identification and documentation of a use case, which leads to the creation of a corresponding GitHub issue. This issue acts as a nexus for collaborative discussion, leading to the creation of a GitHub pull request proposing specific changes

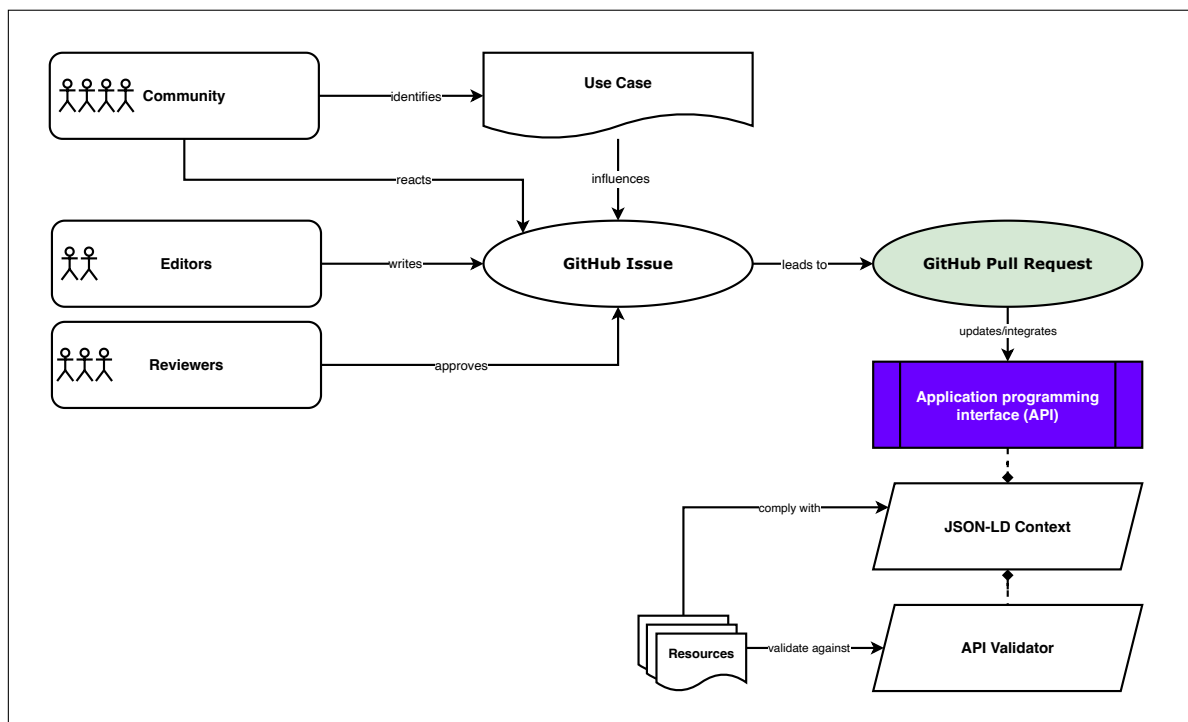


Figure 6.4: Illustrative Diagram of the GitHub-Based Process for Approving and Integrating Use Cases

or additions to the specifications. The figure highlights the critical role that community editors and reviewers play in this sequence, and the collaborative vetting process that ensures that proposals are consistent with the technical and philosophical ethos of the communities. For IIF, this process extends to formal review mechanisms involving the TRC, which ensures adherence to established standards and eventually validation against a validator if needed. In contrast, Linked Art, while using a similar collaborative and open platform for discussion and documentation, manages this process without a formal TRC, relying on broader community engagement and consensus, especially from its Editorial Board.

Figure 6.5 details the process undertaken by the IIF community to review and integrate a specific GitHub issue²⁵⁷, the validation of the Image and Presentation APIs 3.0 that happened in May 2020²⁵⁸. This detailed illustration shows the relationships and interactions between the various actors, and highlights the structured approach to consensus building, specification development and implementation by compatible software. The process begins with a human collective, such as the IIF Editorial Committee, identifying and proposing changes, which are then carefully discussed and voted on by TRC members using platforms such as GitHub and Zoom. The involvement of the TRC is critical, providing the formal review and approval necessary for the adoption of changes, marked by key milestones such as the release of the API 3.0 Release Candidate. This figure emphasises the essential role of compliance and compatibility checks against the JSON-LD API @context and the relevant validators, ensuring that updates are seamlessly integrated into the wider IIF ecosystem, supporting a wide range of compliant clients.

The approval of both specifications by 30 members of the TRC marks a significant milestone, reflecting a substantial level of consensus for such a key initiative. While this level of participation is

²⁵⁷ Release Image API and Presentation API 3.0: <https://github.com/IIF/trc/issues/37>

²⁵⁸ Further details can be found in the official announcement of the final release of the third version of the core IIF specifications, available at <https://iif.io/news/2020/06/04/IIF-C-Announces-Final-Release-of-3.0-Specifications/>

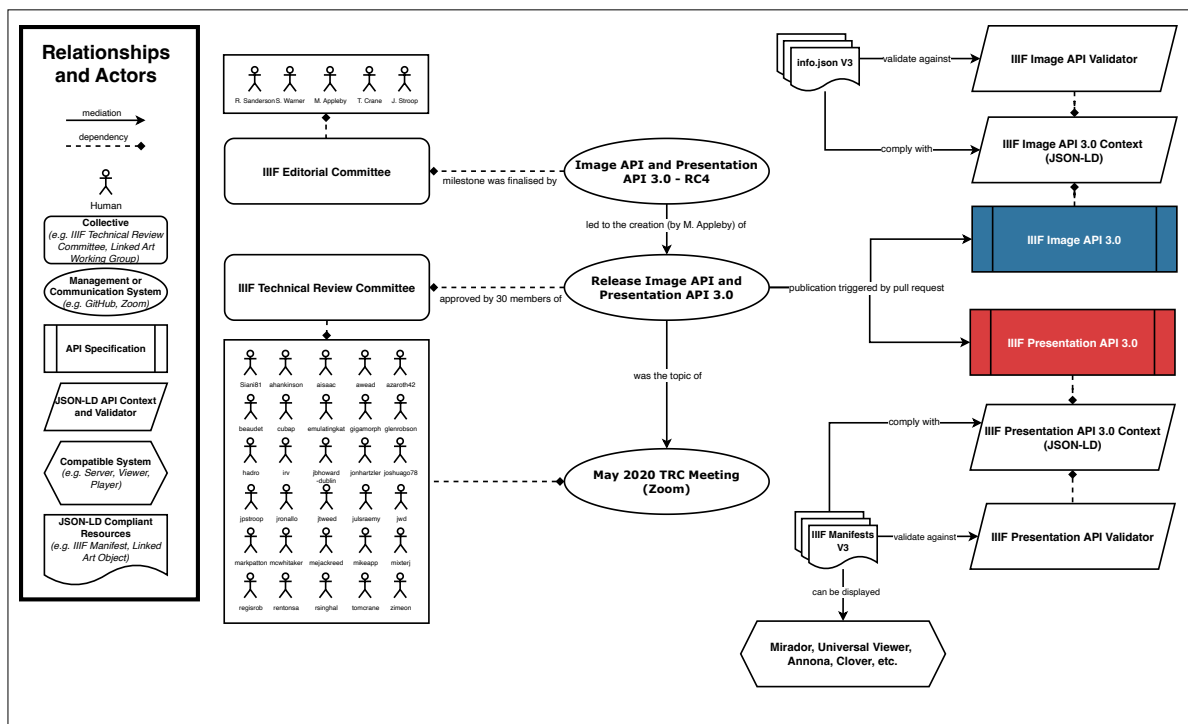


Figure 6.5: Detailed Illustration of the GitHub-Enabled Approval Process by the TRC for the Image and Presentation APIs 3.0

remarkable in the context of other use cases, it also demonstrates the difficulty of achieving widespread participation. The fact that not all TRC members were present or voted on these crucial specifications, and that not every institution within the IIIF-C consistently sends a representative, suggests areas for improvement in engagement and representation within the decision-making processes of the community²⁵⁹.

Despite these challenges, there’s a great deal of trust within the community in those who manage the development and implementation of the specifications. Many institutions and developers are diligently adapting their content and software to meet or upgrade to the latest specifications. However, without a concentrated authority to enforce compliance, deviations occur. These range from instances where cross-institutional interoperability remains achievable despite unorthodox approaches, to scenarios where outright anti-patterns are deployed, underscoring the diverse adherence to the APIs²⁶⁰.

While there is no ‘policing’ of implementations within the IIIF or Linked Art community, understanding and addressing these deviations is crucial for the ecosystem’s health. Moreover, as both communities maintain open and shared specifications, the responsibility falls on the community itself to oversee and assess how the standards are leveraged across different settings.

²⁵⁹ The TRC includes a distinct group of *ex officio* members alongside IIIF-C representatives and up to five community representatives. As of May 2020, IIIF-C comprised 58 member institutions. Despite the significance of the issue at hand, less than half of the eligible voting members, totalling 30, participated in this decision-making process, which can take place asynchronously and as late as two weeks after the meeting.

²⁶⁰ An example of deviation from the IIIF specifications can be seen in the *Digitaler Lesesaal* of the archives of Basel-Stadt and St. Gallen. Despite their efforts to implement IIIF within their systems [see [Kansy and Lüthi 2022](#)], there are technical challenges that could hinder the accessibility of IIIF Manifests beyond their portals. Key issues include non-compliant IIIF Manifests, problematic CORS header configurations, and the way in which URIs are minted, which prevent their use in external IIIF clients. Such practices not only challenge the core tenets of interoperability, but also suggest the need for greater adherence to the IIIF design principles and the exploration of alternatives that would still ensure compliance.

6.2.3 Community Guidelines and Best Practice Efforts

As highlighted in the literature review, IIF and to some extent Linked Art publish guidelines, primarily on their website. These resources are designed to facilitate the discovery of standards-compliant resources across different platforms and to assist individuals, projects and organisations in implementing these standards. This dissemination of knowledge plays a key role in enabling a wide range of users to effectively engage with and use IIF and Linked Art specifications, in many cases mediated by external interfaces.

These best practices are critical as they demonstrate the full potential of the specifications and, in the case of Linked Art, demonstrate the full promise of the model. However, when moving from one client to another, one often encounters BOs, as parts of a resource may be supported differently or not at all [see Raemy 2022b]. In the context of Linked Art, starting with the model before considering the implementation of API endpoints can lead to complex situations. This complexity arises when considering cases of transclusion, as the rigidity of the API requires careful consideration of how different aspects of the model are represented and accessed across endpoints, necessitating a nuanced approach to integrating the model with API capability.

To further understand the nuances of these guidelines and best practices, in March 2024 I undertook an in-depth exploration of the IIF cookbook recipes in subsection 6.2.3.1 and Linked Art patterns in subsection 6.2.3.2. This exploration was done to be reasonably grounded in SNA using NetworkX and Gephi Lite, allowing commonalities within this set of guidelines to be discovered and visualised.

6.2.3.1 IIF Cookbook Recipes

Figure 6.6²⁶¹ illustrates a network derived from the IIF cookbook recipes and their support across various viewer clients, including UV, Mirador, Annona, Ramp, Clover, and the navPlace Viewer²⁶². In this graph, nodes represent either specific IIF Cookbook Recipes, identified by unique slugs and labelled with descriptive titles, or viewer clients, serving as the primary agents interacting with these recipes. The edges between nodes are colour-coded to signify the degree of support – ‘yes’ for full support or ‘partial’ – for each recipe by the respective client.

The allocation of node colours, predicated on their modularity class, categorises the recipes, elucidating their thematic relevance and grouping within the IIF ecosystem. This classification sheds light on the recipes’ roles and their significance to the community, from simple manifest creation to more complex implementations like audio, video, and multi-volume work representations. The size of each recipe node quantitatively reflects the extent of client support, visually emphasising recipes that have garnered widespread adoption or those that may require further advocacy. In a more analytical lens, the graph unveils the pivotal roles of Mirador and UV within the framework, underscoring their centrality and indispensability in the broad spectrum of implementing the core APIs. Concurrently, it delineates specific trajectories, such as Annona and Mirador’s inclination towards annotations, and Ramp’s association with audiovisual content.

In terms of JSON Path analysis, a thorough examination was conducted to discern patterns and insights regarding the structure and use of properties across the IIF Cookbook Recipes. Although this

²⁶¹ This figure is also available in an enlarged format in Figure B.9.

²⁶² A few months after this investigation, Aviary, the Glycerine Viewer, Theseus, and the Curation Viewer have been added to the support matrix.

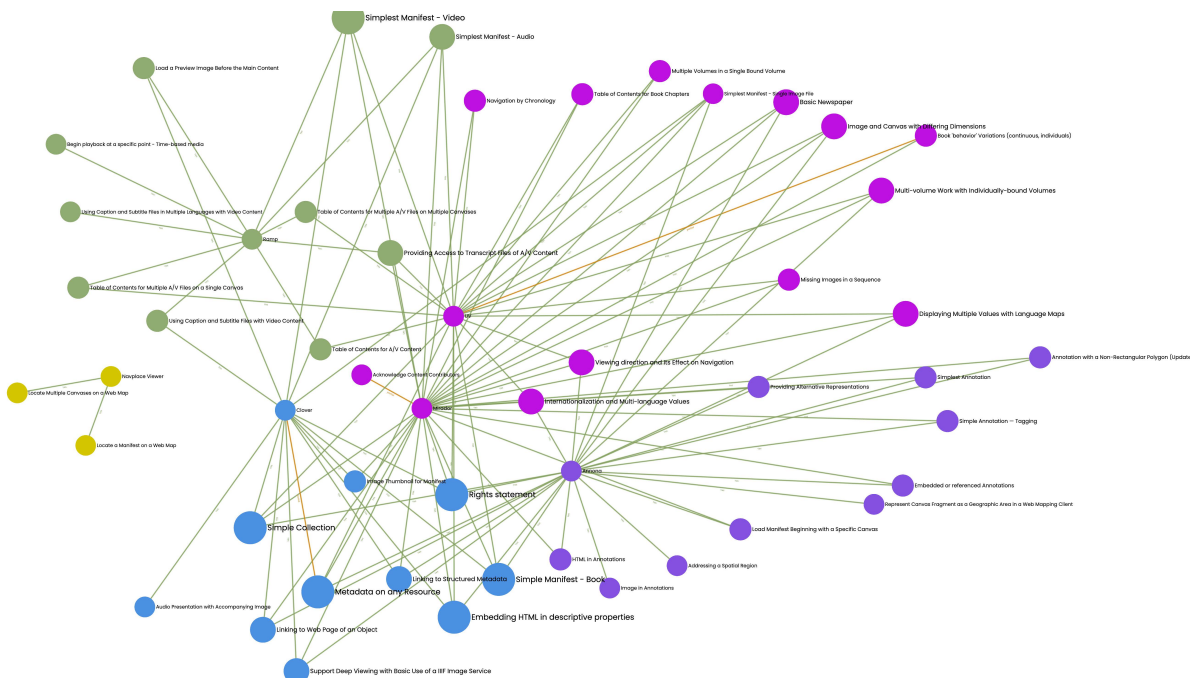


Figure 6.6: Exploring IIIF Cookbook Recipes by Support with Gephi Lite

analysis was meticulously performed, it did not yield any particularly striking results or unexpected insights²⁶³.

For content similarity, leveraging the cosine similarity scores from the analysis, two pairs of recipes emerged as notably identical, showcasing the closest thematic and content-based connections within the dataset. The recipes ‘Embedding HTML in descriptive properties’²⁶⁴ and ‘Rights statement’²⁶⁵ shared the highest similarity score of 0.672318, indicating a significant overlap in their content and approach. This high degree of similarity suggests that these recipes, while distinct in their focus, employ comparable structures or narratives. Another pair, ‘Support Deep Viewing with Basic Use of a IIIF Image API Service’²⁶⁶ and ‘Image in Annotations’²⁶⁷, also demonstrated a notable similarity score of 0.541640.

Additionally, the examination sheds light on the significant re-use of fixtures or content links across the recipes, which underscores the foundational role of these materials within the IIIF ecosystem. The repeated utilisation of specific resources, such as the image captured during the 2019 IIIF Conference in Göttingen, Germany²⁶⁸, which appears 41 times across various guides, exemplifies how readily available and central these resources have become in demonstrating IIIF capabilities. Similarly, audio and video fixtures from Indiana, such as the Mahler Symphony Audio²⁶⁹ and the Lunchroom Manners Video²⁷⁰, highlight the framework’s adaptability and richness in managing a diverse range of multimedia content. However, it also highlights a certain tendency towards using readily available or well-known resources, often those directly provided by the IIIF community or known to be in the public domain. This practice,

²⁶³ More information can be found at https://github.com/julsraemy/loud-socialfabrics/tree/main/01_consensus-advocacy/guidelines.

²⁶⁴ Embedding HTML in descriptive properties: <https://iiif.io/api/cookbook/recipe/0007-string-formats/>

²⁶⁵ Rights statement: <https://iiif.io/api/cookbook/recipe/0008-rights/>

²⁶⁶ Support Deep Viewing with Basic Use of a IIIF Image API Service: <https://iiif.io/api/cookbook/recipe/0005-image-service/>

²⁶⁷ Image in Annotations: <https://iiif.io/api/cookbook/recipe/0377-image-in-annotation/>

²⁶⁸ Image of the Gänslesiel in Göttingen taken at the 2019 IIIF Conference: <https://iiif.io/api/image/3.0/example/reference/918ecd18c2592080851777620de9bc5-gottingen/full/max/0/default.jpg>

²⁶⁹ Mahler Symphony Audio: <https://fixtures.iiif.io/audio/indiana/mahler-symphony-3/CD1/medium/128Kbps.mp4>

²⁷⁰ Lunchroom Manners Video: <https://fixtures.iiif.io/video/indiana/lunchroom-manners/high/lunchroom-manners-1024kb.mp4>

while practical, sometimes leads to a lack of diversity in the resources used, underscoring an opportunity for the IIF Cookbook authors to explore and integrate a wider array of materials.

6.2.3.2 Linked Art Patterns

Turning the analytical gaze to Linked Art, I extend my exploration to the patterns that underpin this vibrant ecosystem. This progression makes it possible to discern the web of relationships that animate the Linked Art data model, mirroring here the methodological approach through graphs with the IIF cookbook recipes.

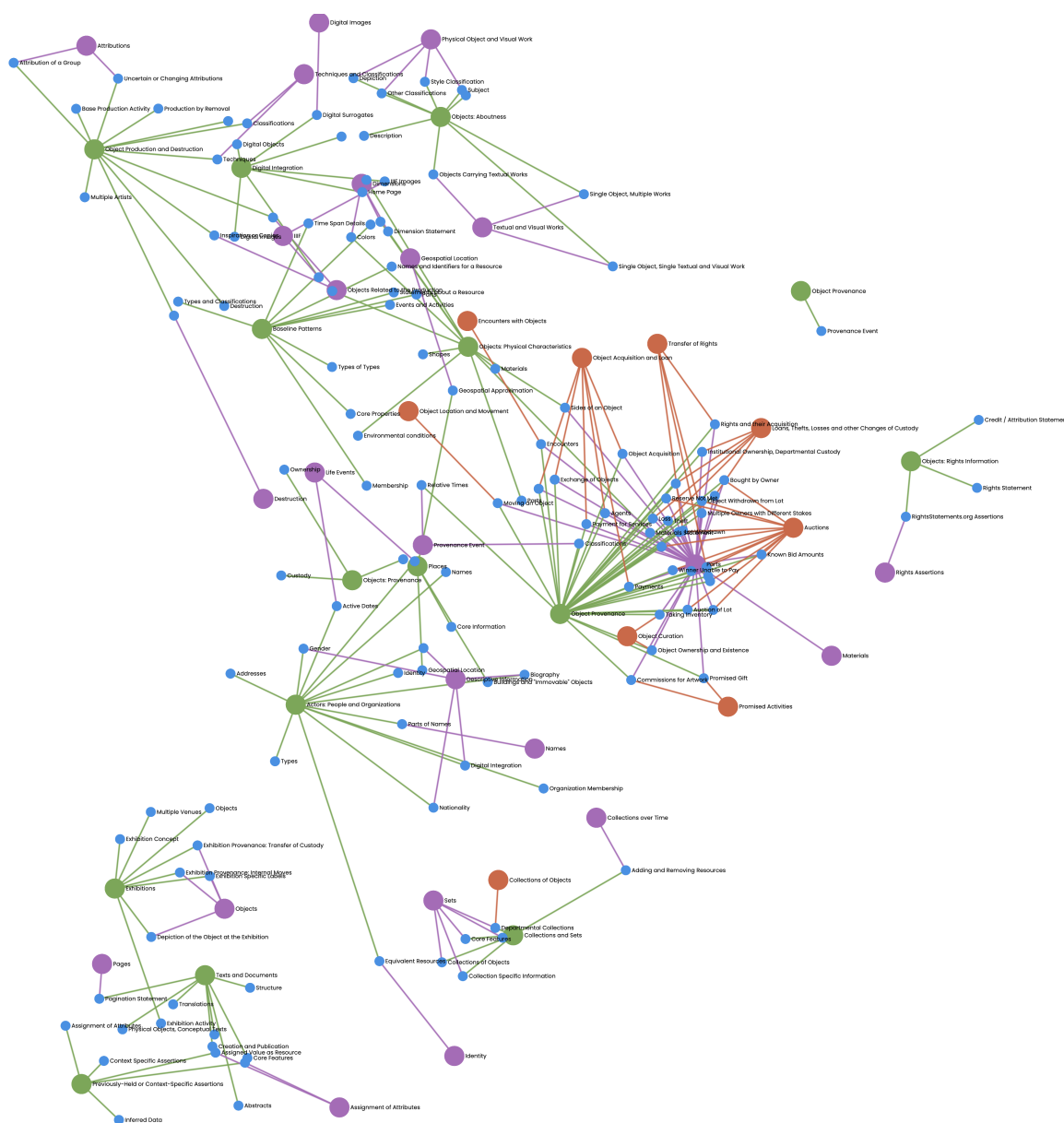


Figure 6.7: Exploring Linked Art Patterns by Category with Gephi Lite

Figure 6.7 presents a graph where nodes symbolise Linked Art patterns²⁷¹. The edges in this visualisation represent the relationships derived from the patterns' data, notably in terms of model components and their (sub)category. The colour of each node is determined by the pattern's category, signifying the thematic or functional group to which it belongs. Meanwhile, the edge colour is informed by the source nodes, providing insight into the directionality and origin of each relationship.

A diverse array of 136 patterns emerges, categorised under themes such as 'Baseline Patterns'²⁷², which serve as the foundation for resource description and interaction. The analysis identifies a comprehensive list of patterns, each addressing specific aspects of CH data modelling. Notably, some patterns function as sub-patterns, inheriting their parent name as one of their categories. Exploration of pattern similarity within the Linked Art model reveals noteworthy connections²⁷³, reflecting the integrated and comprehensive nature of the model. While these findings underscore the versatility and depth of the framework, its true value extends to practical applications. The insights gained from these similarities provide a rich resource for optimising Linked Art implementations in real-world scenarios, suggesting a broad potential for reuse in various CH contexts.

It's important to recognise that, like the IIF cookbook recipes, this graph represents a snapshot in time, reflecting the entities and relationships documented up to a certain point. While this visualisation provides a window into the modelling efforts within Linked Art, other examples using different entities could be explored, recognising the dynamic and evolving nature of the initiative.

6.3 Community Participation and LOUD Standards: Adoption and Impact

This section addresses the objective of examining community participation, alongside the evolutionary adoption of LOUD standards and their potential impact with respect to scientific movements and principles. In pursuit of this objective, data were collected from adoption surveys conducted within the IIF community, enriched further by insights drawn specifically for this study on the socio-technical characteristics of both IIF and the Linked Art community. This comprehensive dataset [see Raemy 2023e] serves as a foundation for our exploration, offering a unique lens through which the dynamics of community engagement, technological adoption, and the overarching influence of LOUD standards.

Following the publication of a detailed survey report on the characteristics of IIF and Linked Art [see Raemy 2023a], this section presents a comprehensive summary of its findings. The survey has provided insightful revelations into the dynamics of community engagement, preferred practices and tools, barriers to wider participation and the resulting implications for the CH sector. These findings underscore the central role of open standards and collaborative efforts in shaping practices, and provide a nuanced understanding of how these communities contribute to the development and adoption of LOUD standards.

Conducted between 24th March and 7th May 2023²⁷⁴, the online survey attracted 79 participants from 20 countries²⁷⁵. It was targeted to individuals involved in IIF and Linked Art, but also people

²⁷¹ It must be said that each time a new pattern is created or modified on the Linked Art website, it is likely that existing URIs of JSON-LD serialisation may also change as they have been dynamically minted.

²⁷² Baseline Patterns: <https://linked.art/model/base/>

²⁷³ More information can be found at https://github.com/julsraemy/loud-socialfabrics/blob/main/01_consensus-advocacy/guidelines/patterns/linked_art_patterns_similarity.csv

²⁷⁴ The survey was being promoted primarily on the Slack workspace of both communities, on their respective mailing list, as well as on Mastodon [see Raemy 2023c].

²⁷⁵ Amongst the survey respondents, 16 individuals reported involvement in both the IIF and Linked Art communities, demonstrating a notable cross-community engagement. In addition, 38 participants were active in either IIF or Linked Art, with a significant majority, 36 respondents, aligning themselves with the IIF community. In contrast, a much smaller group, only 2 respondents, indicated that they were exclusively involved with Linked Art. Interestingly, a

that have heard of the communities or the specifications. The survey’s structure included sections tailored to gauge involvement in either or both communities, with branching logic to ensure relevance to participants’ experiences. It captured data on socio-demographics, organisational affiliation, and the extent of participants’ involvement in community events, activities, and calls, offering a window into the communities’ composition and engagement patterns.

Results highlighted a global distribution of participants, with a noticeable concentration in the Global North, and indicated a wide range of engagement levels, from passive awareness to active participation in community events and activities. The findings underscore the communities’ commitment to fostering global collaboration, despite challenges such as geographical and institutional representation biases. The survey serves as a foundational effort to map the landscape of these communities, identifying areas for growth and increased inclusivity.

The survey findings reveal a core group of individuals actively involved in both the IIIF and Linked Art communities, underscoring the depth of engagement among key members. This involvement reflects a collaborative spirit and a shared dedication to advancing the digital cultural heritage landscape.

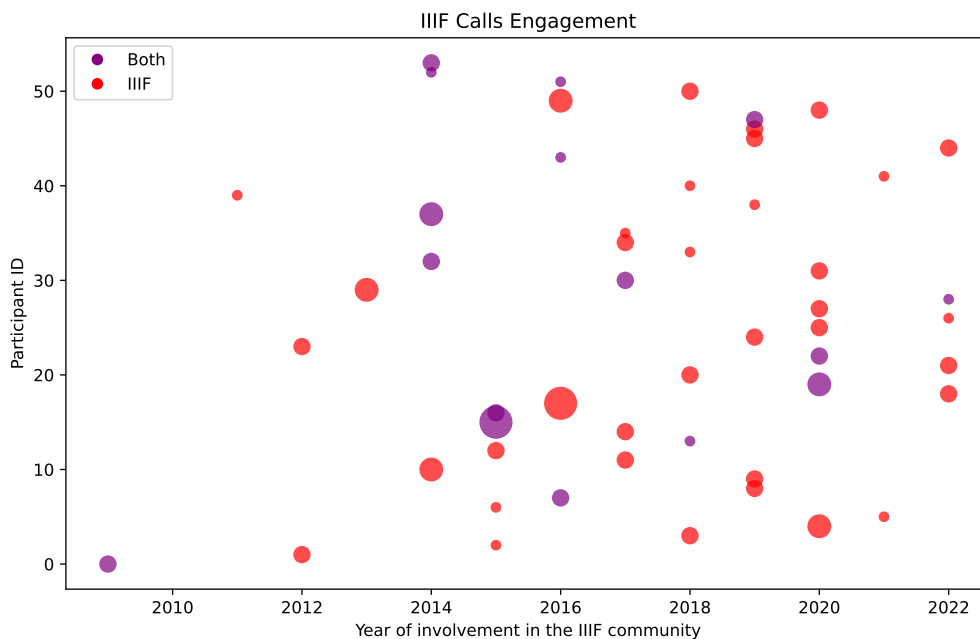


Figure 6.8: Year of Involvement in the IIIF Community [Raemy 2023a, p. 20]

Figure 6.8 illustrates the dynamics of this engagement within the IIIF community, mapping out the timeline of involvement and highlighting the active participation in community calls. It shows the year of IIIF involvement along the x-axis against participant IDs on the y-axis, representing engagement through point size for IIIF call participation over the past year. It highlights a pattern of greater participation from members involved since 2016 or earlier, with one instance of engagement traced back to 2009, demonstrating the impact of long-term involvement and hinting at early contributions to the

significant proportion, 25 individuals, indicated no affiliation with either community, highlighting the diverse levels of engagement and interest within the CH field.

IIIF community’s evolution. Participants who are only involved in the Linked Art community are not included in this plot.

Following on from the visualisation of IIIF’s community engagement, the responses of 52 participants are analysed in Figure 6.9, which situates IIIF’s contributions within broader scientific and data governance movements. The data reveals a strong recognition of IIIF’s pivotal role in several areas: a substantial majority affirm its importance to Open Science and Citizen Science initiatives, illustrating its far-reaching benefits. Participants also highlighted IIIF’s substantial contribution to the FAIR principles, particularly in the areas of interoperability and reusability, demonstrating its effectiveness in promoting a more connected and accessible digital research environment. However, further research is needed to fully assess IIIF’s adherence to the CARE Principles for Indigenous Data Governance, suggesting a potential area for future developmental focus.

Situating IIIF to movements and principles (n=52)

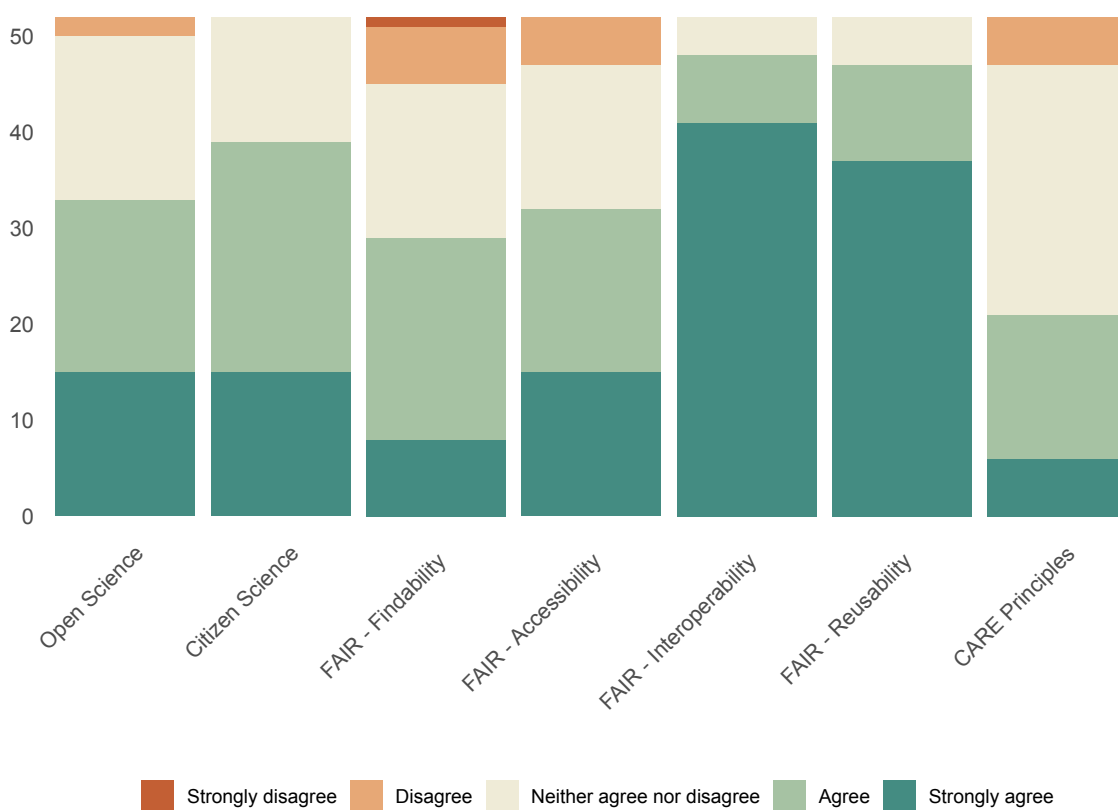


Figure 6.9: Situating IIIF to Movements and Principles [Raemy 2023a, p. 21]

Shifting the emphasis to Linked Art, analysis of the engagement of 16 participants in Figure 6.10 provides a nuanced view of the role of Linked Art in relation to the same scientific and governance movements. Although the pool of participants is smaller, resulting in a more limited data set, the responses still provide useful insights. Participants recognise Linked Art’s potential contributions to Open Science and Citizen Science, albeit to a lesser extent than IIIF. In addition, the feedback highlights Linked Art’s strong capabilities in interoperability, in line with the FAIR principles, particularly in the areas of discoverability, accessibility and reusability. These findings highlight the need for further empirical studies to validate and refine Linked Art’s alignment with CARE, and provide a clear direction for future research and community engagement strategies.

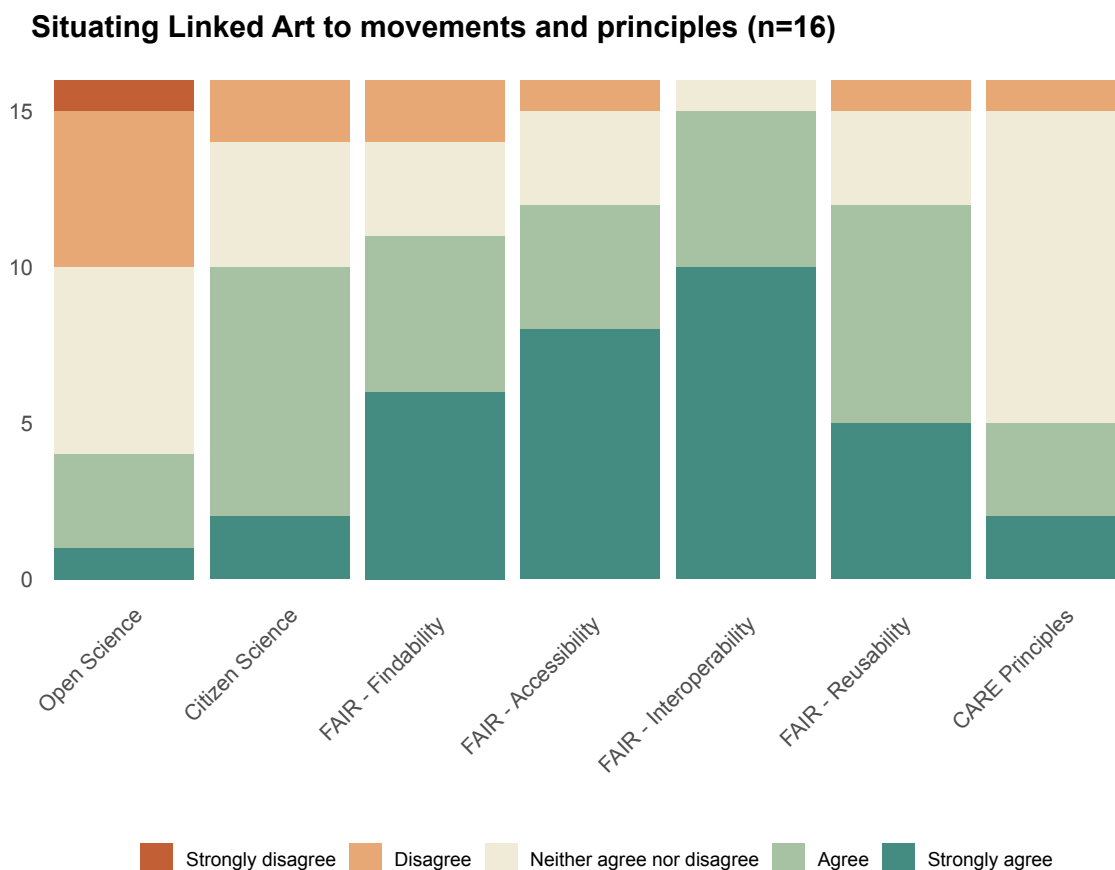


Figure 6.10: Situating Linked Art to Movements and Principles [Raemy 2023a, p. 22]

Following the survey's broad findings on participant distribution and engagement, we transition to a focused exploration of the IIF and Linked Art communities' characteristics. Through this research, I seek to gain a better understanding of the contributions and challenges faced by these communities.

Community Engagement The survey revealed a vibrant ecosystem of practitioners committed to the ideals of IIF and Linked Art. A significant proportion of respondents actively contribute to both initiatives, embodying a spirit of cross-community collaboration. This strong commitment is indicative of a shared vision for the future of digital cultural heritage, where collective effort and knowledge sharing are fundamental pillars.

Practices and tools Looking at the communities' technological preferences, there is a strong trend towards open standards and interoperable tools. This trend is not only a testament to the technical rigour of the communities, but also aligns with the FAIR and CARE principles. Such adherence ensures that digital artefacts and the frameworks used to manage them are accessible, interoperable, reusable and ethically managed, fostering a digital ecosystem that is inclusive and equitable.

Barriers to Involvement Despite the thriving community and progressive adoption of open standards, the survey identifies critical barriers to wider participation. The most prominent of these is the need for more inclusive outreach and comprehensive support for newcomers. This challenge highlights an area of opportunity for community growth, requiring initiatives aimed at lowering barriers to entry and fostering an environment that welcomes diverse perspectives and expertise.

Impact on CH The contribution of IIF and Linked Art to the heritage domain is profound, particularly in enhancing accessibility, interoperability, and reusability. The survey highlights the central

role that IIF and Linked Art play in democratising access to heritage data, enabling richer, more connected narratives that span time and geography. IIF has expanded into various domains, while Linked Art remains focused on art-related objects for a smaller audience.

6.4 Synthesis and Insights

This section synthesises the present chapter and provides some and insights that can be learned from the social fabrics of IIF and Linked Art.

Behind the visible activities like meetings lies a preponderance of preparatory work managed by co-chairs, editorial boards, or driven by community-generated use cases. This foundational work often predetermines the direction and outcomes of formal gatherings. Meeting patterns indicate initial high attendance, which peaks at certain project milestones before declining. Scheduling meetings to accommodate broader global participation, particularly from Asia and Oceania, should be considered to enhance inclusivity, although the primary interaction remains asynchronous.

GitHub serves as a critical hub for community engagement, with a core group of active contributors who often consistently attend meetings. This platform lowers the barrier to decision-making within the community, although it also reflects biases similar to those found in Free/Libre and Open Source Software (FLOSS) communities in terms of inclusiveness. The assertion that *'standards are made by the people who show up, and not nearly enough people are showing up'* as posited by Cramer [2016], underscores the critical role of active participants in shaping these standards. However, when individuals from privileged backgrounds wield disproportionate influence in these participatory roles, the issue of over-representation comes into sharp focus.

This demographic homogeneity can give rise to several challenges, including the perpetuation of biases and the neglect of pivotal issues of relevance to underrepresented or marginalised groups. Such imbalances can lead to standards that inadequately reflect the comprehensive spectrum of societal needs and realities. Furthermore, participation in these standardisation processes is itself a privilege, a point underscored by Capadisli [2023], suggesting the need for greater inclusivity in these influential roles. Moreover, the assumption that internet access and digital devices are universally available is critically examined in works such as Czahajda et al. [2022], which uses ANT to reveal key actors in the digital landscape. This echoes the issues of the IIF community, where the generation of IIF resources presupposes means that may not be accessible to all, particularly outside the Global North and among non-English speakers.

The IIF Cookbook recipes and Linked Art patterns demonstrate the tension between the creation of advanced specifications and their practical implementation by platforms. While new recipes and patterns continue to emerge, actual implementation in software and tools lags behind, reflecting a gap between ideation and practical application. For instance, the IIF community is actively developing customised tools, particularly for annotations, aimed at catering to users with advanced needs. These tools are designed to enhance usability and functionality but have yet to achieve widespread deployment. This ongoing development indicates that the community is still navigating the optimal routes to broad adoption and interoperability, striving to balance sophisticated requirements with general usability. As another example, the question of how digital frameworks represent non-human entities remains a point of discussion within Linked Art, reflecting wider societal challenges in ensuring that diverse perspectives are included in digital and cultural narratives.

Survey findings also underscore the need for ongoing efforts to develop LOUD standards that foster an inclusive, dynamic digital ecosystem. Future strategies should include creating educational resources and frameworks that support interdisciplinary collaboration and reduce barriers to participation.

In summary, as the IIIF and Linked Art communities progress, their continued relevance will increasingly depend on their ability to embrace diverse perspectives and adapt to changing technological and cultural landscapes. The concept of *sympoietic collaboration*, inspired by Donna Haraway’s ideas in Haraway [2016b, pp. 86-87], illustrates this path forward. ‘Sympoiesis’, or making-with, in contrast to ‘autopoiesis’, or self-making, portrays these communities not just as isolated entities but as interdependent networks. This interdependence must extend to incorporating new voices into ideas, meetings, and implementations, fostering vibrant ecosystems where diverse contributions enrich the shared narrative. By prioritising mutual collaboration over individual autonomy, sympoietic collaboration aims to enrich the field, ensuring that digital heritage remains dynamic and inclusive. Pragmatically, this discussion will be continued in the following section, where I will explore specific perspectives and suggest potential pathways forward.

6.5 Perspectives

This section reflects on the empirical findings of this chapter and outlines prospective avenues for both the IIIF and Linked Art communities, emphasising strategic steps forward in community engagement, technical innovation, and culminating in discussing integration practices.

6.5.1 Enhancing Community Engagement

Both the IIIF and Linked Art communities face challenges in transforming increased membership into active participation. Despite a tenfold increase in members on platforms like IIIF Slack workspace over the past 7-8 years, this has not translated into proportional growth in contributions. Advocacy efforts need to focus on making these platforms more welcoming and integrating passive members into active roles. One of these efforts in the IIIF community is the Ambassadors programme²⁷⁶, which could be expanded. For Linked Art, it is essential to increase engagement through clear role definitions and regular interactive sessions that go beyond the technicalities of the model, including workshops and webinars. These efforts should aim to demystify the participation process and encourage more diverse community involvement.

While the primary tools for collaboration in the IIIF and Linked Art communities are hosted by Google, Apple, Facebook, Amazon, Microsoft (GAFAM) – namely GitHub and Google Drive – there are growing calls to consider alternatives that may better align with open-source values and data sovereignty concerns. Codeberg²⁷⁷, as an alternative to GitHub, represents one such possibility. However, transitioning to new platforms could be a significant undertaking, requiring considerable effort and adaptation. Exploring these options could help mitigate the risks associated with reliance on major tech conglomerates and foster a more decentralised approach to digital resource management, though the change would be quite tremendous and involve significant logistics.

6.5.2 Technical Development

The technical backbone of both the IIIF and Linked Art communities relies on the continuous development and maintenance of essential tools.

²⁷⁶ IIIF Ambassadors: <https://iiif.io/community/ambassadors/>

²⁷⁷ Codeberg: <https://codeberg.org/>

For IIIF, this includes servers and clients such as Cantaloupe, Mirador, and UV. These tools not only need regular updates to stay compliant with evolving standards but also require careful management of dependencies, such as upgrading to the latest versions of underlying technologies like React for Mirador, to avoid security vulnerabilities [see [White 2023](#)]. Similarly, Linked Art's integration efforts must focus on enhancing interoperability with broader standards like RiC and Schema.org, which will improve its usability across various DAMs. This effort should be supported by the development of targeted guidelines that cater to different technical levels and user groups, and the creation or availability of comprehensive workflows and tools essential for the uptake of the standard.

These measures are crucial for ensuring the broader adoption and sustainability of IIIF and Linked Art. By keeping technological infrastructure up-to-date and secure, and by enhancing interoperability with other standards, the communities can accommodate a wider range of platforms and users. Additionally, providing clear guidelines and accessible tools is essential for lowering entry barriers, thereby facilitating a more widespread and effective implementation of these standards across diverse environments.

6.5.3 Integration Practices

Robust governance is essential for sustaining the technical infrastructure and community cohesion in both IIIF and Linked Art. Establishing diverse expert governance bodies can guide strategic developments and ensure adherence to best practices. These bodies should also address the preservation of institutional knowledge to maintain continuity amidst changes. The role of initiatives like *iiif-commons*²⁷⁸ is critical here, as they lower barriers to entry and support the development of both general-purpose and specialised tools, accommodating a wide range of community needs²⁷⁹.

LOUD entities, in the context of this analysis, can be perceived as '*[a construction of] new properties to perceive in the world*' [p.251 [McCarty 2023](#), citing [Anderson \[2014\]](#)], suggesting that these entities do more than just exist; they actively shape our understanding and engagement with CH. By rearranging data and metadata in unprecedented ways, LOUD instances allow us to see connections and interactions that were previously obscured. This conceptual framework not only enriches our interaction with digital archives, but also challenges us to rethink how these technologies can be used to foster a more inclusive and dynamically interconnected cultural landscape.

Ultimately, the synthesis of insights from the IIIF and Linked Art initiatives provides a robust framework for future growth. By fostering an environment that encourages sympoietic collaboration and integrates diverse perspectives, these communities can continue to thrive and lead in the creation of open ecosystems. Integrating new voices and adapting to the shifting technological landscape will be crucial for the continued relevance of the IIIF and Linked Art communities.

The following two chapters look at the practical implementation of their respective APIs in a research project and on larger systems, illustrating their versatility and implications in a number of different settings.

²⁷⁸ IIIF Commons: <https://github.com/IIIF-Commons>

²⁷⁹ The contrasting models of support between these Mirador and UV illustrate the diversity of needs and approaches within the IIIF ecosystem. The goal is to balance the development of versatile, multi-functional tools with the creation of specialised, task-driven applications.

7

PIA as a Laboratory

‘[D]igital environments are, of course, themselves socio-technical assemblages, with agency, affective and material qualities.’ [Edwards 2011, p. 48]

This chapter focuses on an empirical exploration of the implementation and evaluation of LOUD standards within the PIA project, which runs from February 2021 to January 2025. Specifically, the discussion covers the refinement of the data model, the creation of bespoke IIF-compliant resources, Linked Art templating and the technological framework to undertake the digitisation, cataloguing and indexing of the associated CAS photographic collections (SGV_05, SGV_10 and SGV_12).

This is an attempt to consider how the PIA project’s implementation of LOUD standards enhances our understanding of their potential to facilitate data reuse and wider participation. I argue that the implementation of shared API allows for a prompt understanding of the benefits and limitations of these standards, drawing on the expertise and contributions of the extended communities. Expected outcomes include improved accessibility and reusability of heritage data, as well as wider engagement with diverse user communities. The aim is to implement and evaluate these standards in an exploratory setting to assess their impact on long-term engagement. The choice of PIA as a testing ground is suited both to the fluidity of research environments and to the possibility of experimenting with different modes of practice. Research agendas are inherently transient and flexible, making PIA a dynamic space for the uptake and evaluation of LOUD standards. This approach not only deepens the understanding of these technologies, but also increases the flexibility to adapt to emerging capabilities.

In Section 7.1, I begin by providing an overview of the infrastructure baseline that was established at the start of the project and the data that was aggregated to set the stage for a thorough assessment of the research. The application and impact of LOUD standards within the research project will be explored in Section 7.2, a section that is crucial to deriving actionable insights and understanding the wider impact of LOUD in a real-world and relatively short-term context. The synthesis of these findings in Section 7.3 provides a comprehensive overview of the impact and potential benefits of LOUD standards on participatory knowledge practices.

The chapter concludes with Section 7.4, which offers insights into the future implications of this work for research and practice, for what will be done until the end of the PIA project or beyond.

7.1 Data and Infrastructure Baseline at the Project Onset

The journey of the analogue archives of CAS began at the end of the 19th century with the creation of the first archives by the SSFS in 1896, now known as CAS. They began collecting photographs in the 1930s and have approximately 300,000 objects spread across 30 collections. This first phase, which continued well into the era of digitisation, consisted of maintaining and expanding analogue archives. Over time, as social and cultural contexts evolved, so did archival practices and technologies. Early

photographic collections were central to shaping national identities and cultural origins during periods of significant industrial and social change [see [Caraffa and Serena 2015](#)]. These collections were not simply repositories of existing cultural elements, but were instrumental in constructing the very notions of culture they purported to document through the systematic collection, archiving and interpretation of images and materials.

In the early 2010s, the second stage began, involving the digitisation of approximately 105,000 photographs, in collaboration with the DHLab and the restoration atelier of Anklin & Assen [[Graf et al. 2019](#)]. This phase included the development of a database for the digitised images and the addition of metadata to these images. The move to digital archives offered many benefits, including easier classification, access from anywhere, and increased security through redundant storage [see [Rosenthaler and Fornaro 2016](#)]. Moreover, as the physical life of analogue originals diminishes and reproduction becomes more difficult, digitisation has become an essential strategy not only for preservation but also for redefining collections as integral components of CH in today's knowledge society.

The digital transformation of the CAS archives has involved a network of interconnected actors, primarily centred around the University of Basel. The analogue CAS archives are stored on one of the university's premises, underlining the collaborative ties and shared resources between the Institute of Cultural Anthropology and European Ethnology and the broader university framework [[Eggmann and Bischoff 2010](#), [Leimgruber et al. 2011](#)]. The digitisation initiative, overseen by the DHLab, reinforces the university's key role in the digital preservation and management of CH [[Fornaro and Chiquet 2020](#)]. After its inception as a project within the DHLab, DaSCH became an independent entity in 2021, operating as a national research data infrastructure with substantial support from the SNSF, with the University of Basel continuing as the hosting institution [[Gautschy 2022](#), [Rosenthaler et al. 2015](#)]. This transition underlines the integral role of the University of Basel in the management of digital archives. In addition, the PIA project, involving similar stakeholders such as the Institute and the DHLab, and now also the HKB as design experts, with DaSCH overseeing the long-term preservation of the (meta)data [[Cornut et al. 2023](#)].

Among the practices for cataloguing and indexing the photographic archives of CAS was, and still is, spreadsheet-based processing of metadata according to a specific data model agreed upon with the DHLab, and uploading this along with the associated data, i.e. the digitisations, onto the Salsah platform, which was the first virtual research environment developed for DaSCH. The metadata uploads could be corrected at any time using an online mass editor on Salsah, and sometimes the digitised objects were uploaded first, sometimes in parallel with the metadata, depending on the needs and resources available.

Figure 7.1²⁸⁰ shows a rather elementary data model of SSFS, which was still in use at the beginning of the PIA project. It is primarily centred on photographic objects, through `sgv:Image`, `sgv:Album` or `sgv:Tonbildschau`, although it already had links to other types of entities, whether people or subjects, or links to Geonames²⁸¹ for geographical referencing. While this model aligns with the spirit of Linked Data, its integration was quite basic; apart from using Dublin Core for titles (`dc:title`), there were no other subproperties or mappings being actively implemented²⁸². Additionally, the naming convention for properties did not fully adhere to camel case²⁸³, which combines words or phrases by capitalising all

²⁸⁰ This figure is also available in an enlarged format in Figure B.11.

²⁸¹ Geonames: <https://www.geonames.org/>

²⁸² The legacy data model in JSON-LD can be viewed at https://raw.githubusercontent.com/Participatory-Image-Archives/pia-data-model/main/ontology/SGV-EKWS/salsah/sgv_old_ontology.json.

²⁸³ Cf. https://en.wikipedia.org/wiki/Camel_case

words following the first word and removing spaces (e.g., `CamelCase`), often used to improve readability and compatibility in coding environments. The eight classes of the SSFS data model are as follows:

sgv:Collection This class is representative of one of the archive's photographic collections. While the material of 17 different collections was partially or fully available through Salsah, 20 were already described at this level.

sgv:Image This is the most widely used class of the data model representing digitised images, totalling around 105,000. This class also includes the largest number of properties, whether for a title, date or location, or to specify the format or type of object, or links representing the subject(s) covered by the image, or to identify the people or institutions depicted in it.

sgv:Album This is an intermediate class between `sgv:Image` and `sgv:Collection` for objects belonging to a photographic album. This class was conceived for visualisation purposes by integrating the notion of sequence number, but was only realised conceptually.

sgv:Tonbildschau A series of diapositives that were originally shown through a carousel slide projector. Like `sgv:Album`, this is also an intermediate layer, and 26 resources from the `SGV_13` collection were assigned to it.

sgv:Film The photographic archives also hold a number of audiovisual materials, such as filmed anthropological interviews. 34 films, all from the `SGV_01` collection, were uploaded onto Salsah.

sgv:Sequence The sequences are oral transcriptions of a `sgv:Film`. For each film, a series of sequences has been associated, generally under the time constraint and therefore includes a property to mark an interval. 290 resources belong to this class.

sgv:Persons They may be individuals, groups of individuals or legal entities. 2515 resources were identified using this class.

sgv:Subject These are monitored topics specifying recurring keywords, but which do not necessarily come from controlled vocabularies. 3005 entries were identified as subjects.

Furthermore, a series of lists completes this legacy data model, notably on the accepted formats (through `has:Format`), medium (via `hasModel`), and types of photographic objects (via `hasObjectType`), in which a hierarchy was created and is maintained by CAS.

A dedicated website has also been created to present the digitised photographic archives, aggregating (meta)data from the Salsah API. On this interface, and still at the time of writing in April 2024, visitors can access around the images digitally preserved by DaSCH, although some of these have not yet been catalogued and/or digitised.

This phenomenon can be clearly illustrated by the three corresponding CAS collections studied over the course of the PIA research project. In the `SGV_05` collection, which comprises approximately 90,000 objects, none had been catalogued or digitised prior to the initiation of PIA, with these processes beginning concurrently once the project commenced. However, not all items in this collection are slated for digitisation or fine cataloguing. In the case of the `SGV_10` collection, although over 12,000 photographs were digitised and catalogued, the level of detail in this cataloguing remains minimal. Regarding the Kreis Family collection, approximately fifty photo albums had yet to be digitised. For the `SGV_12` collection, while more than 47,000 black and white negatives were successfully digitised and catalogued, there remained over 10,000 prints yet to be processed. These prints, a curated selection by the photographer Brunner, are organised within boxes according to a nomenclature established by

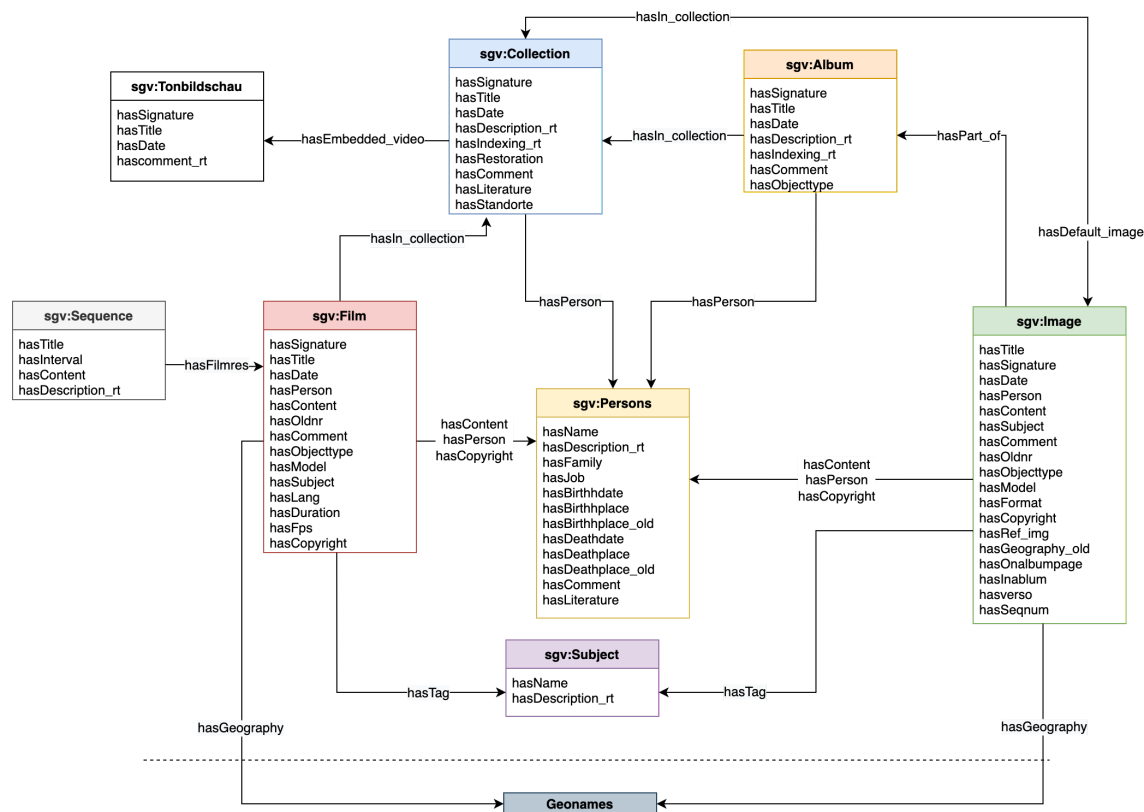


Figure 7.1: Overview of the Legacy SSFS Data Model

the photographer himself. Additionally, some of these objects from these three collections needed to be restored prior to digitisation²⁸⁴.

The following consideration regarding the assignment of identifiers within the CAS photographic archives is intended to provide additional contextual information. Understanding the structure of these identifiers is important, particularly for images without extensive associated metadata, as the identifiers themselves provide valuable clues as to the original format of the items and the collection to which they belong. This systematic approach to naming helps to infer details that are not always explicitly documented elsewhere within the metadata.

Collection Prefix All identifiers start with *SGV_*, followed by a two-digit number which represents the specific collection within the CAS archives. For example, *SGV_10* represent the Kreis Family collection.

Item Type After the collection number, a letter or letters denote the type of item. Each type has a designated letter:

- A for Albums (e.g., *SGV_10A_00001*)
- D for Dias (slides) (e.g., *SGV_10D_00001*)
- N for Negatives (e.g., *SGV_10N_00001*)
- P for Positives (photographic prints) (e.g., *SGV_10P_00001*)
- S for Schriftliches (written documents) (e.g., *SGV_10S_00001*)

²⁸⁴ Within PIA, a dedicated tool was created so that the conservator-restorers could write notes about their work on objects from the three collections: <https://pia-restoration.dhlab.unibas.ch/>.

- G for Gegenstand (objects) (e.g., `SGV_10G_00001`)
- F for Film (e.g., `SGV_10F_00001`)
- T for Ton (sound recordings) (e.g., `SGV_10T_00001`)

Sequential Number Following the item type, there is an underscore followed by a sequence number formatted with leading zeros. This number uniquely identifies each item within its type and collection. For example, the first album in the `SGV_10` collection would be `SGV_10A_00001`.

Special Identifiers for Carousel Items Some collections, such as those involving carousels, have additional components in their identifiers. After the type and sequential number, additional identifiers may include a carousel number and position within that carousel, separated by underscores. For instance, `SGV_13D_22_00001_a` indicates the first item in the 22nd carousel of the `SGV_13D` collection, position `a`.

Initially, the database migration from Salsah to DSP was scheduled to be completed by the start of the project in 2021. However, delays meant that this task extended nearly to the project's conclusion, necessitating some extraction and handling of (meta)data, especially at the beginning of PIA. This ongoing requirement significantly influenced the project workflow and data management strategies, underscoring the dynamic nature of managing large-scale digital archives.

The migration to DSP also provided an opportunity to redesign the data model to better meet the needs of both CAS and PIA. A balance had to be struck between minting new PIA-specific properties and dealing with the extension of the model on the research project so as not to overload the CAS data model. The development of DSP started as a replacement for Salsah and consists of three main components: DSP-APP, DSP-API and DSP-TOOLS. DSP-APP is the GUI that allows users to view and edit data directly in their browser, improving accessibility and interaction. DSP-API acts as the core of the software stack, an RDF database that provides API access to data, enabling data interactions. Finally, DSP-TOOLS is a command-line tool for uploading data models and large datasets to a DSP server, streamlining the data integration process.

One of the challenges was to integrate LOUD standards within these processes and to coordinate this implementation alongside the myriad satellite activities of the research project. In the following section, I will delve into the specifics of these efforts and discuss how collaboration and technological adaptation were managed concurrently with the main objective tied to my thesis within PIA.

7.2 LOUD Standards in an Explorative Setting

PIA has envisioned the development of processes, tools and interfaces to generate and make visible knowledge in a participatory manner. The purpose is to enable intuitive access and exploration, based on the example of three collections from the CAS photographic archives. To this end, we have used a variety of face-to-face, virtual and hybrid formats: plenary meetings every six weeks, group and thematic meetings at different intervals, meetings between PhD students, sprint meetings among developers or designers, project-related courses, as well as internal and external workshops. Additionally, we had to navigate the creation of a technical infrastructure that would allow us to begin our research activities independently or in tandem with the ongoing database migration, restoration, and digitisation efforts. This infrastructure setup was critical to ensure that our research could proceed without too much delay, even as fundamental records management tasks were being completed.

With a view to achieving semantic interoperability affordances beyond or in parallel with the development of GUI-based tools, I, with the help of my colleagues, have approached the deployment

and establishment of LOUD specifications and practices within the project incrementally, as a way of addressing the needs of different audiences, whether human or machine, but also as an attempt to extend participation beyond the set of tools being developed specifically for PIA. As such, the two main IIIF APIs were first implemented to streamline access to image-based resources and to facilitate user annotation of resources in accordance with WADM. Then, an endpoint compatible with the IIIF Change Discovery API to support ongoing updates and changes was deployed. On the Linked Art front, a conceptual model mapping the current classes and properties, and boilerplate documents were created, followed by the roll-out of a preliminary version to coincide with the ongoing development of the initiative's first stable specification release. Not only does Linked Art serve as a benchmark for accessing CH semantic metadata, but also as an alternative endpoint that expands the scope and accessibility of data interactions.

Figure 7.2²⁸⁵ presents a synoptic - and asynchronously processed - overview of the PIA technical infrastructure, providing a high-level perspective focusing on APIs and (meta)data integration rather than the various GUI-based prototypes. This view helps to outline the phased implementation of the infrastructure, which will be explored hereafter, starting with (meta)data extraction, followed by an explanation of the updated CAS data model. I will then look at the IIIF Image API and the IIIF Presentation API along with WADM, moving on to IIIF Change Discovery and concludes with an overview of the metadata APIs, including Linked Art. Each of these components plays an important role in the project's architecture. The systematic deployment of these processes and standards has been strategically designed to build layers of semantic interoperability incrementally.

7.2.1 Metadata Extraction and Integration

For the extraction of (meta)data, we used and extended a Python-based workflow originally developed by DaSCH²⁸⁶. The primary script helped us to export data in XML and images in various formats. Further, the script's ability to filter and batch-download data according to specific parameters, such as project identifier or resource type, streamlined the management of large datasets. This process of exporting images from the Kreis Family and Ernst Brunner collections, along with all available metadata, was repeated in the early stages of the research project in 2021 to populate our database and our own

²⁸⁵ This figure is also available in an enlarged format in Figure B.12.

²⁸⁶ salsah-export: <https://codeberg.org/PIA/salsah-export>

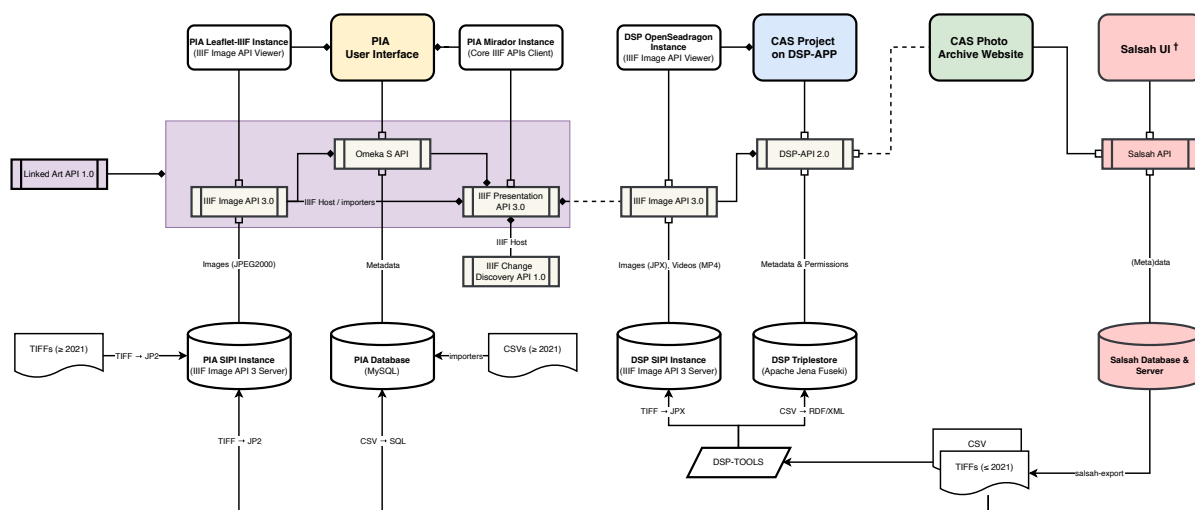


Figure 7.2: Synoptic View of the PIA Infrastructure: Showcasing its Connection to DSP and the CAS Photo Archive Website

SIPI instance to serve IIIF-compliant images.

As shown in Table 7.1, approximately 40% of the images lack a specific date, and further analysis revealed that 17% of the items (10,423 objects) have no metadata beyond their identifier. This scenario is relatively typical for an archive that often lacks detailed individual records. However, the meticulous item-level cataloguing and indexing undertaken by CAS and more specifically for this project places us in an environment more akin to a museum than a traditional paper archive. Typically, the latter describe collections at a file or series level, but our approach ensures a richer, more detailed context for each item, improving the accessibility of the content.

Table 7.1: Exported Images Along with Their Respective Time-Based Metadata from Salsah (SGV_10 and SGV_12)

Description	Count	Percentage
Images without a date	23,993	40.06%
Images with accuracy to the day	4,344	7.25%
Images with accuracy to the month	11,882	19.84%
Images with accuracy to the year	19,674	32.85%
Images with date ranges	2,377	3.97%
Total images count	59,893	100.00%

All images and associated metadata have been hosted on two Ubuntu virtual machine servers at the University of Basel, where we have also installed code repositories. We then purchased for a few years the domain `participatory-archives.ch` from Infomaniak, a Geneva-based web hosting company. Post-2021 digitised data was also temporarily stored on these servers pending migration to DSP. Although PIA has primarily served as an interim stage for image hosting, it is envisaged that the project’s micro-services, together with the front-end prototypes and functionalities, should persist and possibly be developed beyond the end of the project. However, these developments may not necessarily be maintained within the PIA ecosystem, suggesting a possible transition to other platforms or frameworks to ensure their long-term sustainability and integration.

7.2.2 CAS Data Model

Initial discussions on database migration began in November 2021, and the migration process itself was completed on the production server in March 2024. Throughout this period, the extraction and upload of (meta)data to DSP was performed repeatedly, primarily facilitated by DSP-TOOLS and DSP-INGEST²⁸⁷, a service which streamline and optimise the process of ingesting data within DSP, carried out largely by the DaSCH team, with my assistance when required. At the same time, I undertook the task of extending and aligning the data model with the Knora Base Ontology²⁸⁸, a built-in OWL model [see Schweizer and Geer 2021].

The Knora Base Ontology serves as the foundational meta-ontology for all projects using the DSP-API²⁸⁹. It is designed to support the complex data structures common in humanities research, where values often have their own metadata, including creation dates and permissions. Each project must define its data model through extensions to this base ontology, ensuring compatibility and integration within the DSP framework.

²⁸⁷ DSP-INGEST: <https://docs.dasch.swiss/latest/Dsp-Ingest/>

²⁸⁸ Knora Base Ontology: <https://docs.dasch.swiss/latest/DSP-API/02-dsp-ontologies/kb/>

²⁸⁹ The Knora Base Ontology is identified by the IRI <http://www.knora.org/ontology/kb> and typically abbreviated as `kb` in documentation.

Projects need to create their ontologies as extensions of Knora, defining specific resource classes that are subclasses of `kb:Resource`. These classes can represent different types of content, such as documents, images or audio files, and must be attached to the project using the `kb:attachedToProject` property. This structured approach facilitates a detailed and nuanced representation of data, enabling advanced queries and updates. In addition, the ontology specifies structured value types for storing detailed metadata for each value, increasing the granularity of data management. Resources in a project are associated not only by their type, but also by properties defined in the ontology. These properties, which can be linked to Knora values or other resources, obey the constraints set by the ontology to ensure data consistency and integrity.

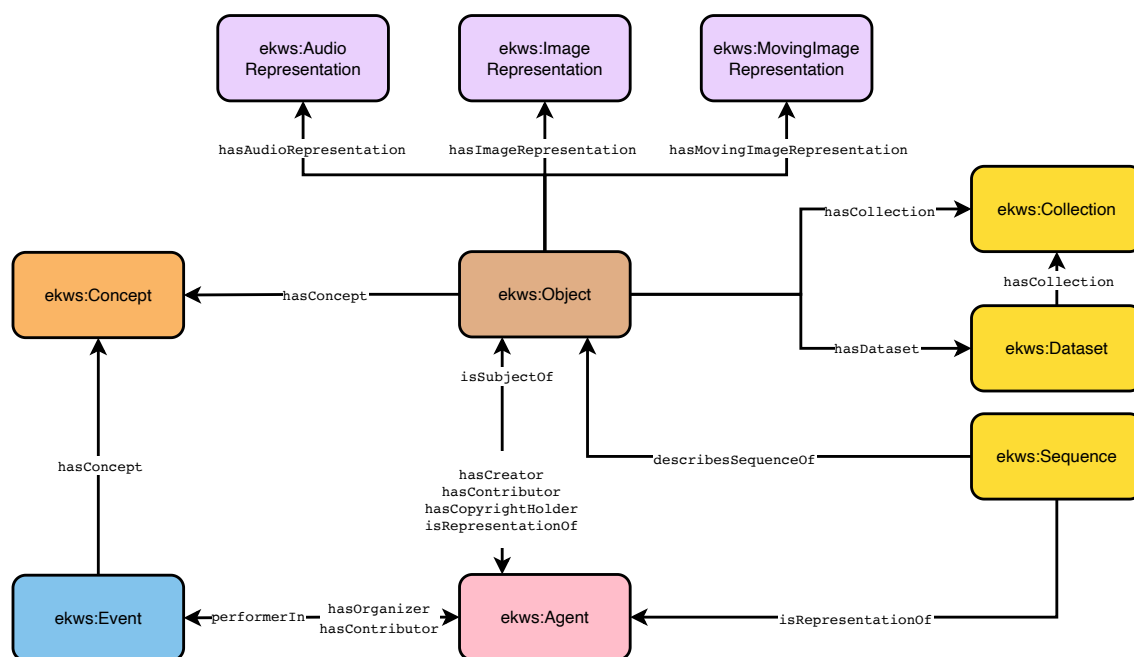


Figure 7.3: Overview of the Updated CAS Data Model

An overview of the main classes and the properties that link them within the updated CAS Data Model is presented in Figure 7.3²⁹⁰. This update was decisive not only to facilitate the migration of the database to DSP, but also to meet the specific requirements of the PIA project²⁹¹. This involved integrating new properties, deprecating or minting new classes, making naming consistent, as well as adding or pruning categories to the existing controlled vocabularies.

The renaming of CAS provided an opportunity to change the prefix from `sgv` to `ekws`, reflecting the updated organisational identity in German in the model's entity names. Entities such as `sgv:Bild`, `sgv:Film`, `sgv:Tonbildschau` and `sgv:Album` have all been superseded by `ekws:Object`, which is iterative to allow the creation of series of objects (such as photo albums or books).

To support the representation of different digital assets on DSP (e.g. when multiple media types portray the same object), especially by viewers or players, the following three classes related to `ekws:Object` have been created:

²⁹⁰ The colour schemes used in this representation follow best practices similar to those used in CIDOC-CRM and Linked Art data visualisations, and help to identify entity types. It is important to note, however, that the colour coding does not reflect mappings to the Knora Base Ontology, as this ontology focuses primarily on media-based entities.

This figure is also available in an enlarged format in Figure B.13.

²⁹¹ CAS Data Model (Ontology) on DSP-API: <https://api.dasch.swiss/ontology/0812/ekws/v2>

- `ekws:ImageRepresentation`
- `ekws:AudioRepresentation`
- `ekws:MovingImageRepresentation`

Two new classes have also been added: `ekws:Dataset`, a placeholder for sub-collections, used for example in the `SGV_17` collection to split it in two, as well as for any user-generated lists worth preserving, and `ekws:Event` for modelling activity-centric representations (e.g. an exhibition, an object biography, a crowdsourcing campaign, etc.). Also, `sgv:Subject` has been renamed to `ekws:Concept` and `sgv:Persons` changed to `ekws:Agent`.

A concerted effort was made to map classes and properties, primarily with Schema.org and to some extent with Dublin Core, SKOS and EDM, to ensure a common denominator with the data model being created for the PIA project within its Omeka S instance, a lightweight content management system to publish CH online collections [Sacramento et al. 2022, p. 122]²⁹². Furthermore, it is also anticipated that the CAS data model will be revised by the end of PIA to integrate new, enriched information. Throughout the project, the metadata of the `SGV_05`, `SGV_10`, and `SGV_12` collections have been further refined. These improved (meta)data should be imported into DSP at some point near the completion of PIA or shortly after. Table 7.2 provides a detailed overview of the classes from both the original and updated data models. In the revised model, each CAS class is mapped to Knora using `rdfs:subClassOf`, which ensures alignment, essential for integration within DSP.

Table 7.2: Overview of Classes: Legacy and Updated Data Models and its Mapping to the Knora Base Ontology

Legacy SSFS Data Model	Updated CAS Data Model	Knora Base Ontology Resource Type
<code>sgv:Image</code>	This class has been deprecated in favour of <code>ekws:Object</code> . Digital assets will be represented through <code>ekws:ImageRepresentation</code> .	-
<code>sgv:Film</code>	This class has been deprecated in favour of <code>ekws:Object</code> . Digital assets will be represented through <code>ekws:MovingImageRepresentation</code> .	-
<code>sgv:Tonbildschau</code>	This class has been deprecated in favour of <code>sgv:Object</code> but still exist as an object type category. Digital assets will be represented through <code>ekws:ImageRepresentation</code> .	-
<code>sgv:Sequence</code>	<code>ekws:Sequence</code>	<code>kb:VideoSegment</code>
<code>sgv:Album</code>	This class has been deprecated in favour of <code>ekws:Object</code> . Digital assets will be represented through <code>ekws:ImageRepresentation</code> .	-
<code>sgv:Subject</code>	<code>ekws:Concept</code>	<code>kb:Resource</code>
<code>sgv:Collection</code>	<code>ekws:Collection</code>	<code>kb:Resource</code>
<code>sgv:Persons</code>	<code>ekws:Agent</code>	<code>kb:Resource</code>
-	<code>ekws:Object</code>	<code>kb:Resource</code>
-	<code>ekws:Dataset</code>	<code>kb:Resource</code>
-	<code>eksw:Event</code>	<code>kb:Resource</code>
-	<code>ekws:ImageRepresentation</code>	<code>kb:StillImageRepresentation</code>
-	<code>ekws:MovingImageRepresentation</code>	<code>kb:MovingImageRepresentation</code>
-	<code>ekws:AudioRepresentation</code>	<code>kb:AudioRepresentation</code>

7.2.3 Implementation of the IIIF APIs and Web Annotation

To make the images compatible with the IIIF Image API, we decided to use our own SIPI instance, a IIIF image server developed by colleagues at the University of Basel. This server works, at the time of writing, with the JPEG2000 format and uses the Kakadu software to encode and decode the images on-the-fly [Rosenthaler et al. 2017]²⁹³.

²⁹² Omeka S: <https://omeka.org/s/>

²⁹³ Kakadu is the fastest library for encoding JPEG2000, yet it is not open source [see Robson et al. 2023].

In addition, SIPI includes a built-in web server capable of handling common file types and running Lua scripts, including Lua embedded in HTML pages. This server facilitates the creation and management of dynamic web content directly within the SIPI environment. For command-line use, SIPI provides utilities for converting images to and from formats such as TIFF, JPEG2000, JPEG and PNG, attempting to preserve all embedded metadata and ICC colour profiles, i.e. any set of data that characterises a colour input or colour space²⁹⁴. Deploying images using SIPI was fairly straightforward as it is provided as a Docker image, although our implementation relied on the free but temporary service provided by the University of Basel. This dependency required careful management of the directories to be served and their paths specified in the Docker settings. As this service often undergoes updates and maintenance, we were occasionally faced with disruptions where our web services were temporarily unavailable. To integrate these images into the PIA front-end, we implemented Leaflet-IIIF²⁹⁵, a plugin for the Leaflet mapping library specifically designed to display IIIF-compliant images. This setup allowed us to display images using structured URLs within our infrastructure, as shown below:

- <https://sipi.participatory-archives.ch/<collection>/<filename>>
- https://sipi.participatory-archives.ch/SGV_12/SGV_12N_08589.jp2

As the migration of images from Salsah to DSP has finally been completed²⁹⁶, the transition from a virtual research environment without IIIF capability to one that includes a IIIF Image API service via the DSP's SIPI, meaning that we now have two hosting spaces that are both compliant with the IIIF Image API 3.0. With these images now available on DSP, the need to maintain duplicate images on the virtual machine servers is reduced, at least for all images digitised prior to 2021. As a suggestion for streamlining our storage of digitised content, it would be efficient to consolidate all image hosting back to DSP. This would not only simplify the infrastructure, but also improve the accessibility and manageability of our digital assets. The URL structure for accessing these images served by their SIPI instance, uses a combination of a short code for the dataset and a unique identifier, on DSP is as follows²⁹⁷:

- <https://iiif.dasch.swiss/<shortcode>/<uuid>>
- <https://iiif.dasch.swiss/0812/276uIbjSulF-k5RrtYZ3LUA.jpjx>

The example URLs point to the same image. Locating these images requires some care due to the opaque nature of the identifiers used on DSP, which differs from our strategy of using CAS identifiers. Interestingly, the first extraction process from Salsah revealed discrepancies, such as the discovery that 661 images from the Ernst Brunner collection were present on Salsah but not retrievable, and 102 images, although digitised, were missing from both Salsah and the PIA database. The 661 missing images have since been located on DSP following the database migration. Ongoing research aims to resolve the remaining discrepancies.

Code Snippet 7.1 illustrates the SPARQL query employed to retrieve and construct all IIIF Image API URLs on the Apache Jena Fuseki²⁹⁸ server provided by DaSCH. The results of this query, along with a dedicated Python script to populate a spreadsheet, have been comprehensively documented on GitHub²⁹⁹. Currently, this approach has provided us with all necessary links to potentially streamline

²⁹⁴ ICC: <https://www.color.org>

²⁹⁵ Leaflet-IIIF: <https://github.com/mejackreed/Leaflet-IIIF>

²⁹⁶ Approximately 21TB from the CAS dataset were uploaded onto DSP.

²⁹⁷ The short code for CAS on DSP is 0812. See <https://ark.dasch.swiss/ark:/72163/1/0812>

²⁹⁸ Apache Jena Fuseki: <https://jena.apache.org/documentation/fuseki2/>

²⁹⁹ Finding IIIF Image API URLs from DSP: <https://github.com/Participatory-Image-Archives/pia-data-model/tree/main/dsp-migration/find-iiif-url>

our digital asset management by integrating these URLs into the PIA environment, thus bypassing our SIPI server and potentially improving the efficiency of our infrastructure.

Code Snippet 7.1: SPARQL Query Executed on the DaSCH Fuseki Instance to Retrieve and Construct IIIF Image API URLs

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX ekws: <http://www.knora.org/ontology/0812/ekws#>
4 PREFIX kb: <http://www.knora.org/ontology/kb#>
5
6 SELECT ?img ?internalLink ?label ?stillImageUUID ?stillImageInternalFilename ?ImageAPI ?ImageAPIinfo
   ?ImageAPIfull WHERE {
7   ?img a ekws:ImageRepresentation ;
8       rdfs:label ?label ;
9       kb:hasStillImageFileValue ?hasStillImageFileValue .
10
11   ?hasStillImageFileValue kb:valueHasUUID ?stillImageUUID ;
12                           kb:internalFilename ?stillImageInternalFilename .
13
14   # Construct IIIF Image API URLs based on ?stillImageInternalFilename
15   BIND(URI(CONCAT("https://iiif.dasch.swiss/0812/", STR(?stillImageInternalFilename))) AS ?ImageAPI)
16   BIND(URI(CONCAT("https://iiif.dasch.swiss/0812/", STR(?stillImageInternalFilename), "/info.json"))
   AS ?ImageAPIinfo)
17   BIND(URI(CONCAT("https://iiif.dasch.swiss/0812/", STR(?stillImageInternalFilename), "/full/max/0/
   default.jpg")) AS ?ImageAPIfull)
18
19   # Extract the internal ID part from ?img
20   BIND(STRAFTER(STR(?img), "http://rdfh.ch/0812/") AS ?internalLink)
21
22 }
23 ORDER BY ASC(?label)

```

After establishing a service aligned with the IIIF Image API, we created boilerplates³⁰⁰ for resources compatible with the Presentation API, largely inspired by cookbook recipes from the IIIF community. As illustrated in Table 7.3, these recipes have been instrumental in optimising the interaction capabilities. For instance, the ‘Single Image File Manifest’ recipe is particularly relevant as it aligns with the predominant composition of the CAS dataset, which consists mainly of individual images. Similarly, the ‘Book Manifest’ recipe has been applied effectively to structure our collections of photo albums. The deployment of the SIPI instance allows for deep viewing experiences, leveraging image service capabilities for interactive engagement with high-resolution images. Additionally, incorporating multi-language metadata fields supports accessibility and inclusiveness, reflecting Switzerland’s linguistic diversity³⁰¹. We also would like to ensure proper acknowledgement of content contributors, enhancing transparency and crediting all parties involved. The use of structured metadata links (`seeAlso`) facilitates better aggregation and discovery processes, while the annotations management system, should be able to handle both user inputs and ML outputs.

A script was then set up to generate JSON-LD files automatically by fetching descriptive and legal metadata from a database and extracting image dimension data from our IIIF Image API service. In our efforts to comply with the IIIF Presentation API 3.0, we not only aimed to improve the accessibility of resources, but also explored the implementation of annotations according to the WADM standard. As a result, the following two endpoints have been created to facilitate access to these resources:

- For accessing the IIIF manifests of individual resources:
<https://iiif.participatory-archives.ch/<id>/<manifest.json>>

³⁰⁰ Host Application for PIA IIIF Resources – Boilerplates:

<https://codeberg.org/PIA/pia-iiif-manifest-host/src/branch/main/boilerplates>

³⁰¹ Not all metadata fields have been translated in all official Swiss languages and English. As far as the metadata itself is concerned, the cataloguing has been done almost exclusively in German.

Table 7.3: Relevance of IIIF Cookbook recipes to our resources. Adapted from [Raemy and Demleitner \[2023\]](#)

IIIF Cookbook Recipe	Relevance to PIA
Simplest Manifest - Single Image File	Predominantly used for individual images in the CAS dataset
Simple Manifest - Book	Applied to photo albums
Support Deep Viewing	Leveraging SIPI for enhanced image interaction
Internationalization and Multi-language Values	Supports metadata (fields) in different languages
Acknowledge Content Contributors (providers)	Credits to CAS and other stakeholders
Linking to Structured Metadata (seeAlso)	Enhancing discovery and aggregation
Annotations Management	Pointing to one AnnotationPage per user (including ML outputs)

- For retrieving annotations associated with the resources:

```
https://iiif.participatory-archives.ch/annotations/<id.json>
```

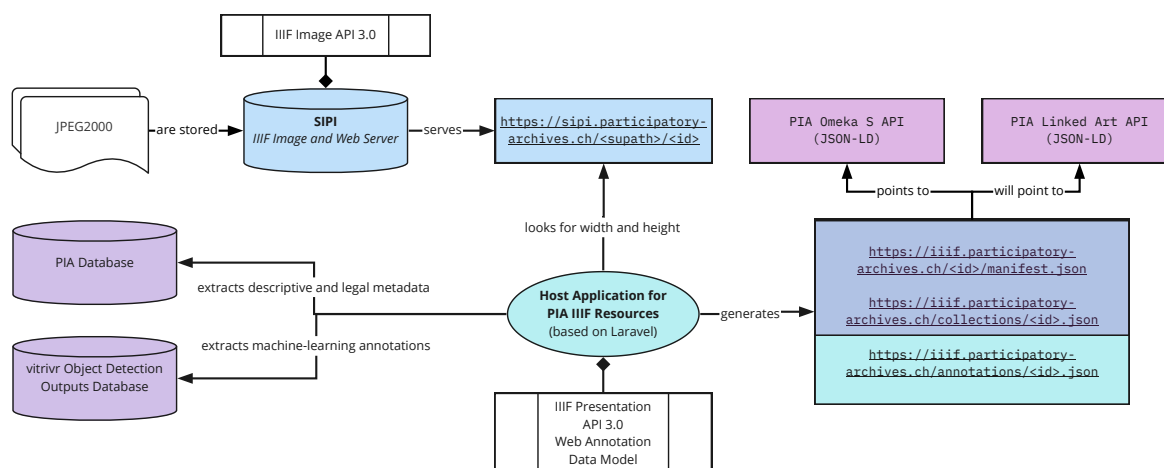
Figure 7.4: Overview of the Legacy Automatic Annotation Workflow [[Cornut et al. 2023](#), p. 12]

Figure 7.4³⁰² displays the annotation workflow, initially managed by a Laravel-based application³⁰³. This setup was responsible for generating our IIIF manifests and was fully integrated into our Omeka S instance. Although the management system has evolved, the basic principles of the workflow remain unchanged. The URL structure for the endpoints remains similar, accommodating both the machine-generated annotations from vitrivr³⁰⁴, a ML multimedia retrieval system developed at the University of Basel [see [Rossetto et al. 2016](#), [Spiess and Schuldt 2022](#)], and the IIIF Presentation API resources.

Code Snippet 7.2 shows an **AnnotationPage** with a machine-generated annotation using vitrivr applied to detect objects. This JSON-LD snippets is associated with a specific Manifest featuring a photograph from the Ernst Brunner collection³⁰⁵. Notably, this photograph, depicting people dancing as displayed on Mirador in Figure 7.5, was one of the works selected for inclusion in the iconic ‘The Family of Man’ exhibition, curated by Edward Steichen and first presented in 1955 at the Museum of

³⁰² This figure is also available in an enlarged format in Figure B.14.

³⁰³ Laravel: <https://laravel.com/>

³⁰⁴ vitrivr: <https://vitrivr.org/>

³⁰⁵ Brunner, Ernst. [Ringtanz während der Masüras auf der Alp Sura]. Guarda, 1939. Black and White Negative, 6x6cm. SGV_12 Ernst Brunner. SGV_12N.08589. CC BY-NC 4.0
CAS Photo Archive Website: <https://archiv.sgv-sstp.ch/resource/430824>.
DSP-APP: <https://ark.dasch.swiss/ark:/72163/1/0812/T6ITgGbkVomteyknedSDsA7>

Modern Art in New York³⁰⁶.

Code Snippet 7.2: Example of Machine-generated Annotations in a IIIF setting from PIA

```

1 {
2   "@context": "http://iiif.io/api/presentation/3/context.json",
3   "id": "https://iiif.participatory-archives.ch/annotations/SGV_12N_08589-p1-list.json",
4   "type": "AnnotationPage",
5   "items": [
6     {
7       "@context": "http://www.w3.org/ns/anno.jsonld",
8       "id": "https://iiif.participatory-archives.ch/annotations/SGV_12N_08589-p1-list/annotation-436121.json",
9       "motivation": "commenting",
10      "type": "Annotation",
11      "body": [
12        {
13          "type": "TextualBody",
14          "value": "person",
15          "purpose": "commenting"
16        },
17        {
18          "type": "TextualBody",
19          "value": "Object Detection (vitriivr)",
20          "purpose": "tagging"
21        },
22        {
23          "type": "TextualBody",
24          "value": "<br><small>Detection score: 0.9574</small>",
25          "purpose": "commenting"
26        }
27      ],
28      "target": {
29        "source": "https://iiif.participatory-archives.ch/SGV_12N_08589/canvas/p1",
30        "selector": {
31          "type": "FragmentSelector",
32          "conformsTo": "http://www.w3.org/TR/media-frags/",
33          "value": "xywh=319,2942,463,523"
34        },
35        "dcterms:isPartOf": {
36          "type": "Manifest",
37          "id": "https://iiif.participatory-archives.ch/SGV_12N_08589/manifest.json"
38        }
39      }
40    }
41  ]
42 }

```

In the beginning, our workflow was designed for single images and did not take into account objects consisting of a sequence of images. To address this, we worked with my colleague Murielle Cornut to identify a number of use cases to be built as IIIF Manifest. This development was largely facilitated by the IIIF Presentation API 3 Python Library, a community developed tool³⁰⁷. Specifically, we focused on two photo albums from the Kreis Family collection, which were then modelled as templates for subsequent implementation on Omeka S. These templates not only facilitated the standardisation of our digital representations, but also provided a framework for future extensions.

The digital artefacts from this process have been meticulously organised into three different types of IIIF Manifests³⁰⁸, each that I self-hosted for different viewing preferences: a comprehensive Manifest containing all pages with and without protective film, a paged Manifest for a cleaner view of the pages, and a layered Manifest that shows the overlay options.

As a practical application of this methodology, Code Snippet 7.3 illustrates how two digitised images – one with and one without protective film – of a particular page from a photo album are

³⁰⁶ This global tour showcased photographs that emphasised universal human experiences and cultural moments. The collection has been permanently displayed at Clervaux Castle in Luxembourg since 1994.

³⁰⁷ iiif-prezi3: <https://iiif-prezi.github.io/iiif-prezi3/>

³⁰⁸ IIIF Manifest for SGV_10 Photo Albums: <https://github.com/Participatory-Image-Archives/iiif-photoalbum>

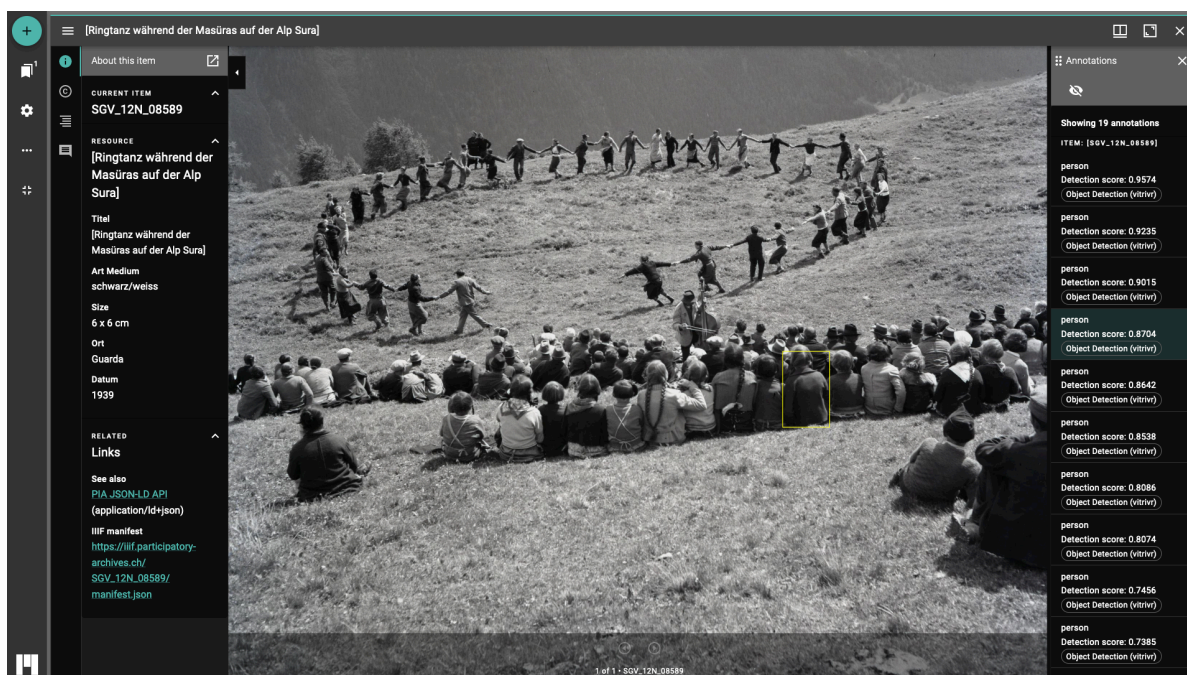


Figure 7.5: SGV_12N.08589 Displayed on Mirador with Machine-generated Annotations

layered onto the same Canvas. I have resized the width to the same dimension to ensure that the images retain their proportions and do not appear too much distorted in the viewer when the layers are switched on and off.

Code Snippet 7.3: IIIF Manifest with Two Images Painted onto the Same Canvas

```

1  {
2  "id": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/canvas/p3",
3  "type": "Canvas",
4  "label": {
5    "en": [
6      "SGV_10A_00050_003.jp2"
7    ]
8  },
9  "height": 3519,
10 "width": 5000,
11 "items": [
12   {
13     "id": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/page/p3",
14     "type": "AnnotationPage",
15     "items": [
16       {
17         "id": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/annotation/p3",
18         "type": "Annotation",
19         "motivation": "painting",
20         "body": {
21           "id": "https://sipi.participatory-archives.ch/SGV_10/album/SGV_10A_00050_003.jp2/full/max/0/default.jpg",
22           "type": "Image",
23           "height": 3519,
24           "width": 5000,
25           "service": [
26             {
27               "id": "https://sipi.participatory-archives.ch/SGV_10/album/SGV_10A_00050_003.jp2",
28               "type": "ImageService3",
29               "profile": "level2"
30             }
31           ],
32           "format": "image/jpeg"
33         },
34         "target": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/canvas/p3"

```

```

35     },
36     {
37         "id": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/annotation/p3/2",
38         "type": "Annotation",
39         "motivation": "painting",
40         "body": {
41             "id": "https://sipi.participatory-archives.ch/SGV_10/album/
42                 SGV_10A_00050_003_mit_Schutzfolie.jp2/full/max/0/default.jpg",
43             "type": "Image",
44             "height": 3524,
45             "width": 5000,
46             "service": [
47                 {
48                     "id": "https://sipi.participatory-archives.ch/SGV_10/album/
49                         SGV_10A_00050_003_mit_Schutzfolie.jp2",
50                     "type": "ImageService3",
51                     "profile": "level2"
52                 }
53             ],
54             "format": "image/jpeg"
55         },
56         "target": "https://julsraemy.ch/hostiing/manifests/SGV_10A_00050_layers/canvas/p3"
57     }
58 ]]]],

```

Through the initial workflow, the IIF Change Discovery API was successfully deployed within PIA. As of the end of April 2024, we have hosted 60,230 Manifests. This count represents a modest increase from the initial ingest in Spring 2021, with a few hundred additional images. However, the current script's approach of generating one Manifest per image may introduce a bias in representing the digital materiality of the collections, as it does not account for the structured nature of certain objects, such as photo albums. The following three URLs illustrate the hierarchical structure of our Change Discovery API, facilitating the tracking and dissemination of updates across the digital collection:

- **OrderedCollection** - Root Endpoint:
<https://iif.participatory-archives.ch/activity/all-changes.json>
- **OrderedCollectionPage** – Page 0:
<https://iif.participatory-archives.ch/activity/page-0.json>
- **Create** – SGV_12N.08589:
https://iif.participatory-archives.ch/activity/create/SGV_12N_08589.json

Code Snippet 7.4 illustrates the root endpoint of our Change Discovery API, following a similar implementation to that of the University of Oxford's Digital Bodleian³⁰⁹.

Code Snippet 7.4: PIA IIF Change Discovery API

```

1  {
2    "@context": "http://iiif.io/api/discovery/1/context.json",
3    "id": "https://iif.participatory-archives.ch/activity/all-changes.json",
4    "type": "OrderedCollection",
5    "totalItems": 60230,
6    "first": {
7      "id": "https://iif.participatory-archives.ch/activity/page-0.json",
8      "type": "OrderedCollectionPage"
9    },
10   "last": {
11     "id": "https://iif.participatory-archives.ch/activity/page-2409.json",
12     "type": "OrderedCollectionPage"
13   }
14 }

```

³⁰⁹ Digital Bodleian's IIF Change Discovery API: <https://iif.bodleian.ox.ac.uk/iif/activity/all-changes>

7.2.4 Deployment of APIs and Vocabulary to Support Semantic Interoperability

In the extended PIA infrastructure, several APIs fulfil key functions in providing semantic metadata, each supporting different facets of the project's digital ecosystem. Initially, the PIA research project used a bespoke JSON API which, although no longer maintained, laid the foundations for current and more sustainable solutions³¹⁰. We then moved on to Omeka S, which has its own REST API³¹¹. In this constellation there is the already mentioned DSP-API from DaSCH as well as our initiative to create a Linked Art API³¹², which is intended both as a benchmark for interoperability and as a practical implementation exercise.

In parallel to these API deployments, we created a controlled vocabulary in SKOS, specifically for the categorisation of Ernst Brunner's photographic material³¹³. Although all the technological building blocks mentioned are closely interrelated, I will focus here on the latter two of these efforts: the Linked Art API and the development of the thesaurus.

Figure 7.6 illustrates the various Linked Art entities expected to be deployed as API endpoints, a move designed to provide another community-driven norm, this time in terms of semantic affordances, to the PIA project's digital ecosystem³¹⁴. The primary focus will be on the full use of the **object**, **digital**, **concept** and **set** endpoints, with the additional use and to some degree the **person** and **group** endpoints to supplement the relational and contextual metadata. The use of these endpoints is not just a technical improvement, but a fundamental step towards a more networked and semantically accessible digital archive.

An example of a **HumanMadeObject** boilerplate for integration with the Linked Art **object** endpoint is shown in Code Snippet 7.5, which is a fundamental step in structuring the archives for improved semantic interoperability.

Code Snippet 7.5: Modelling in Linked Art of a **HumanMadeObject** from the Kreis Family Collection

```

1 {
2   "@context": "https://linked.art/ns/v1/linked-art.json",
3   "id": "https://data.participatory-archives.ch/object/220133.json",
4   "type": "HumanMadeObject",
5   "_label": "PIA ID 220133 - [Porträt von Emilie Kreis-Martz]",
6   "classified_as": [
7     {
8       "id": "http://vocab.getty.edu/aat/300127121",
9       "type": "Type",
10      "_label": "Albumen Print",
11      "classified_as": [
12        {
13          "id": "http://vocab.getty.edu/aat/300435443",
14          "type": "Type",
15          "_label": "Type of Work"
16        }
17      ]
18    },
19    {
20      "id": "http://vocab.getty.edu/aat/300127131",
21      "type": "Type",
22      "_label": "Cabinet Photograph",
23      "classified_as": [
24        {
25          "id": "http://vocab.getty.edu/aat/300435443",

```

³¹⁰ PIA Data API: <https://json.participatory-archives.ch/>

³¹¹ PIA Omeka S REST API: <https://participatory-archives.ch/api>

³¹² PIA Linked Art API: <https://data.participatory-archives.ch/> – Note that the production instance for this API hasn't been deployed.

³¹³ Ernst Brunner Thesaurus: <https://vocab.participatory-archives.ch/>

³¹⁴ PIA Linked Art (GitHub): <https://github.com/Participatory-Image-Archives/linkedart/>

```

26     "type": "Type",
27     "_label": "Type of Work"
28   }
29 ]
30 }
31 ],
    
```

To generate data for the Linked Art API, I initiated a collaboration with the University of Oxford in late 2022³¹⁵, drawing on their experience with CH data and the Linked Art model [Raemy et al. 2023b]. Together we developed a workflow to transform the already digitised SGV_10 and SGV_12 collections into Linked Art. This process involved reusing the existing boilerplates that encapsulate the data characteristics and cataloguing practices, thus standardising the representation of different object types within these collections.

The transformation workflow³¹⁶ used a three-stage process where the collection data, initially queried via the existing PIA JSON API, was first mapped to an intermediate JSON format. This intermediate format acts as a flexible stage to facilitate the subsequent transformation into Linked Art JSON-LD, allowing greater adaptability to different data sources and object types. We used the Python library Cromulent to create basic Linked Art representations, which were then enriched with detailed attributes of photographic objects, such as names, web pages and digital services. The adaptability of this workflow not only meets the immediate needs of the PIA project, but also sets a precedent for

³¹⁵ Linked Art II: Developing Community, Practice, and Scholarship:

<https://linked.art/community/projects/linkedartii/>

³¹⁶ Linked Art Collection Data Workflow: <https://github.com/tgra/Linked-Art-Collection-Data-Workflow>

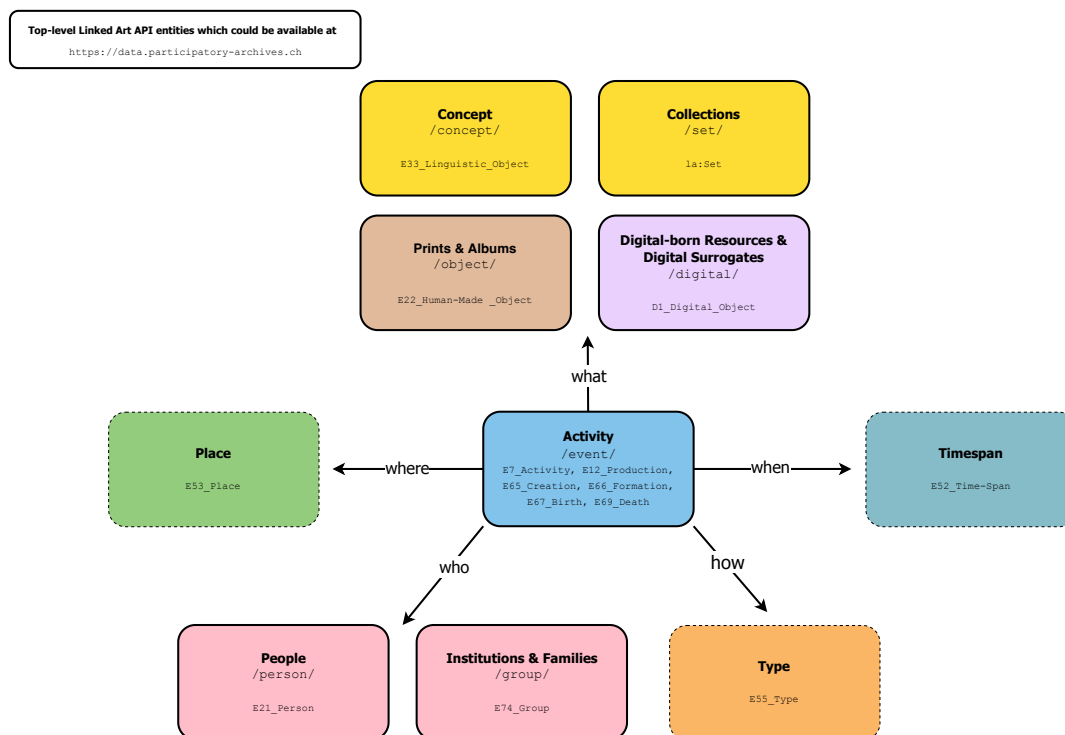


Figure 7.6: Suggested Top-level Entities of the PIA Linked Art API

future adaptations. It provides a framework that can be reconfigured for use with other data sources, or extended to include new object types and Linked Art patterns.

Despite the initial success and adaptability of the workflow, it has become increasingly clear that the JSON API, which had not been populated for some time, was long out of date and no longer met the evolving needs of the PIA project. In light of this and the delayed implementation of the PIA Linked Art API, I have decided to develop an alternative approach that is better suited to our specific needs. This new process is inspired by the basic Extract, Transform, Load (ETL) principles of the original workflow, but has been adapted to produce Linked Art data that better suits the needs of the PIA project and my thesis. This revised method allows for more precise control over data transformations and is designed to handle the latest data inputs from Omeka S and DSP, ensuring that our Linked Art representations are both current and relevant. These changes were essential, especially as the Linked Art API 1.0 had not been officially released at the time of these developments, and remained a moving target. Moreover, the plan to deploy the Linked Art API through this ETL pipeline for the research project has unfortunately never seen the light of day and remains an outstanding task that should be addressed.



Figure 7.7: Boxes Holding Printed Photographs by Ernst Brunner According to his System of Classification

The Ernst Brunner Thesaurus was developed in collaboration with my colleague Fabienne Lüthi in order to systematise the vast number of terms used by the photographer Ernst Brunner to categorise his photographs, or rather a selection of his black and white negatives that were printed and mounted on index cards. The printed photographs were systematically stored in a series of boxes by the photographer and his assistant, as depicted in Figure 7.7.

This meticulous classification covers a wide range of topics, from occupational activities to regional differences within Switzerland, from animal species, especially those encountered in the countryside, to infrastructure elements. For example, under the broad category ‘work’ there are subcategories for specific tasks such as ‘watering’, ‘haymaking’ and ‘baking’. Each term not only reflects the subject of the photograph, but also provides insights into the cultural and social contexts of the time. In total, Ernst Brunner made use of 255 terms. Figure 7.8 shows on the left hand side with a few terms from

Ernst Brunner's classification, the place (Eggiwil, Bern) as well as an identifier (HY 59), on the right hand side there is the actual photograph on an other card³¹⁷.

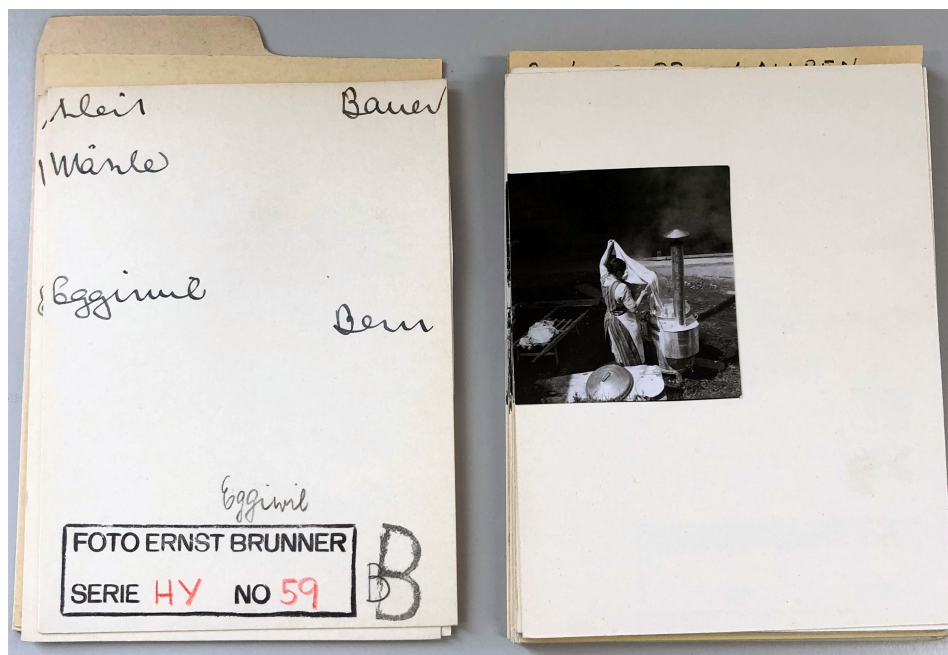


Figure 7.8: Ernst Brunner Thesaurus: Index Card of HY 59

Code Snippet 7.6 presents the Turtle serialisation for a subset of the Ernst Brunner Thesaurus, specifically illustrating the term ‘Brauch’ and its related narrower concepts. This serialisation is part of a broader effort to encode the Thesaurus using SKOS, in which SkoHub³¹⁸ plays a central role. SkoHub is a tool for managing and publishing SKOS vocabularies, effectively facilitating the linking and dissemination of structured semantic data [Pohl and Ostrowski 2019].

Code Snippet 7.6: Turtle Snippet of the Ernst Brunner Thesaurus: Brauch

```

1 brunner:04 a skos:Concept ;
2   skos:prefLabel "Brauch"@de ;
3   skos:topConceptOf brunner: ;
4   skos:narrower brunner:04a, brunner:04b, brunner:04c, brunner:04d, brunner:04e,
5     brunner:04f, brunner:04g, brunner:04h, brunner:04i, brunner:04j,
6     brunner:04k, brunner:04l, brunner:04m, brunner:04n .
7
8 brunner:04a a skos:Concept ;
9   skos:prefLabel "Religion"@de ;
10  skos:inScheme brunner: ;
11  skos:broader brunner:04 .
12
13 brunner:04b a skos:Concept ;
14   skos:prefLabel "Politik"@de ;
15   skos:inScheme brunner: ;
16   skos:broader brunner:04 .
17
18 brunner:04c a skos:Concept ;
19   skos:prefLabel "Alp"@de ;
20   skos:inScheme brunner: ;
21   skos:broader brunner:04 ;
22   skos:narrower brunner:04c01, brunner:04c02, brunner:04c03 .

```

³¹⁷ The source of this image is: Brunner, Ernst. [Wäsche im Freien]. Eggiwil, March 1944. Black and White Negative, 6x6cm. SGV_12 Ernst Brunner. SGV_12N.19859. CC BY-NC 4.0

CAS Photo Archive Website: <https://archiv.sgv-sstp.ch/resource/442094>.

DSP-APP: https://ark.dasch.swiss/ark:/72163/1/0812/_Xn9IpyYWzSwkrObzG=q0Qp

³¹⁸ SkoHub: <https://skohub.io/>

```
23
24 brunner:04c01 a skos:Concept ;
25   skos:prefLabel "Alp-Abfahrt"@de ;
26   skos:inScheme brunner: ;
27   skos:broader brunner:04c ;
28   skos:scopeNote "Vermutlich nicht von Ernst Brunner erstellt (kommt wahrscheinlich von Loosli)."@de
.
```

In the early stages of developing our digital infrastructure, we used Archival Resource Keys (ARKs), an open PID system recognised for its robust community support [Koster 2020, Kunze and Bermès 2024] and managed by the ARK Alliance³¹⁹. The decision to use ARKs was influenced by their flexibility and established practice in Swiss digital infrastructure projects [Cevey and Raemy 2020, Raemy and Schneider 2019a]. Each term in our Ernst Brunner Thesaurus has been assigned an ARK identifier through a collaboration with Arketype, an ARK allocation service. Although it is no longer maintained and that we decided to not assign other ARKs³²⁰, the resolution for these identifiers are still operational as the binding took place directly on the Name-To-Thing (N2T) global resolver³²¹. This assignment process for the thesaurus consisted of leveraging the suffix passthrough functionality³²², which is a feature of the N2T resolver. This functionality allows a single identifier to be assigned to a complex object or collection, in this case the root vocabulary of the vocabulary, while still allowing clear references to each individual component or term within the collection. Each term within the thesaurus was then given a distinct identifier based on an internal numbering system. This method allows efficient referencing and management of hierarchical or grouped data, with the primary ARK acting as a gateway to a structured set of resources.

The use of ARKs has streamlined the process of integrating and mapping the thesaurus concepts across various platforms. Now, each concept can be easily linked to our Omeka S instance and represented in Linked Art by using the `equivalent` property. This approach ensures that our metadata remains interconnected and semantically enriched across different systems.

This section has provided a condensed yet detailed exploration of how LOUD standards and their associated components have been implemented within PIA in an attempt to provide common semantic interoperability affordances. The successful integration of these standards across our infrastructure is the result of a collaborative effort that extended beyond individual endeavours. It involved not only the concerted efforts of fellow PhD candidates and software developers within our project but also drew significantly on the expertise of the IIF and Linked Art communities. This community-driven approach has been instrumental in enriching our infrastructure with interoperable solutions that will continue to support research and broader engagements.

7.3 Synthesis and Insights

To encapsulate the implementation of the LOUD standards within this exploratory framework: we adhered as much as possible to the best practices and guidelines derived from these two communities. My ability to serialise and represent data was contingent on its digital availability and to some extent to the database migration. Consequently, while I was able to effectively interact with and explore the Kreis Family and Ernst Brunner collections, the ASV collection presented challenges due to its late digitisation in the project timeline. Despite this timing issue, the impact of the ASV collection will ultimately be realised through the application of the aforementioned APIs. This delay has partly been

³¹⁹ ARK Alliance: <https://arks.org/>

³²⁰ PIA ARK (public archive): <https://github.com/Participatory-Image-Archives/ark>

³²¹ N2T: <https://n2t.net/>

³²² Cf. N2T feature: suffix passthrough: <https://arks.org/about/ark-naans-and-systems/>

offset by a productive collaboration between PhD candidates Birgit Huber and Max Frischknecht, who have jointly contributed to the digital integration of this collection [see [Huber and Frischknecht 2024b](#)].

Overall, different APIs have been progressively deployed to address different requirements while allowing parallel exploration of data modelling. Each API offers unique advantages, but their collective integration promotes a semantic interoperability layer. This not only represents high-resolution images with their structural and essential metadata, but also incorporates both machine-generated and human annotations, which can be retrieved via the IIIF Change Discovery API. The Manifests are also linked to semantic metadata captured in Linked Art, which in turn can semantically model the various IIIF services. As a result, this setup permits round-tripping across the APIs.

As a concrete example, the IIIF Image API has been instrumental in rationalising the distribution of images across our various prototypes, providing efficient access to high quality digital surrogates and the ability to easily resize them for different use cases. The API's capabilities ensure that images can not only be viewed in detail, but can also be interactively manipulated, increasing user engagement and accessibility³²³.

Similarly, the IIIF Presentation API has enabled rich storytelling through external annotation tools such as Exhibit, which we have used on several occasions. This API allows for the dynamic presentation of images and metadata, creating engaging narratives that highlight the cultural and historical significance of the collections.

However, this implementation of complex digital objects has primarily been on a case-by-case basis, focusing on accurately representing the digital materiality of photographic albums. This selective approach echoes practices even at major institutions such as the Getty Institute, which uses a similar method for edge cases such as the Bayard Album. This album, a digital replica of a long disassembled 19th century photo album, has been reconstructed using the IIIF Presentation API. Over 300 images are composed on a single canvas, closely mimicking the original layout, and enhanced by using the `Choice` class³²⁴ to provide views of the recto and verso of the inserted photos, demonstrating nuanced materiality³²⁵. Such detailed representations, while enriching for scholarly discourse, pose challenges for scalability, even across larger collections. PIA explores such detailed visualisations, albeit on a very different scale, and provides users with alternative views that significantly improve the understanding of photographic albums.

Reflecting on the digital presentation of the Kreis Family Collection, [Edwards and Hart \[2004, p. 12\]](#) remind us that photographs should be seen as '*objects in a historically marked time*'. [Edwards \[2009, pp.130-131\]](#) argues that photographs should be seen as '*material performances*' that involve more than the forensic and semiotic analysis of content; they also circulate through complex networks. This perspective is relevant to PIA because it reveals the need to capture comprehensively the materiality of digital objects. [Edwards and Hart \[2004, p. 15\]](#) go on to argue that '*even in the digital world, these material decisions are integral to the social saliency of the photograph*', underscoring the importance of thoughtful digital curation. Adapting such detailed visual representations for scalable application presents significant challenges, particularly because the structural information that seems obvious in analogue form does not always translate seamlessly into digital formats. Furthermore, some aspects of materiality may never be fully captured digitally, limiting the scope of what can be effectively displayed.

³²³ For example, we deployed a meme generator that could easily leverage the IIIF Image API URL syntax.

³²⁴ Cf. Multiple Choice of Images in a Single View (Canvas): <https://iiif.io/api/cookbook/recipe/0033-choice/>

³²⁵ Bayard Album: <https://www.getty.edu/art/collection/object/1040J2>
Custom IIIF Manifest: <https://data.getty.edu/media/manifest/bayard-custom>

This recognition compels us to continually rethink ways to record and convey the rich material and historical context of photographic archives, in order to broaden their interpretive richness and maintain their relevance in the wider scholarly discourse.

The deployment of the Change Discovery API served as a practical demonstration that establishing such a protocol on top of the IIF Presentation API is not only viable but quite straightforward. This particular implementation highlights the potential for wider adoption by other IIF implementers. However, the true utility of this setup will be realised through the development of aggregators that can leverage this capability, particularly for harvesting by leveraging structured metadata available through the `seeAlso` property in Manifests. This requirement demonstrates an opportunity for further innovation in how digital collections are accessed and integrated across platforms.

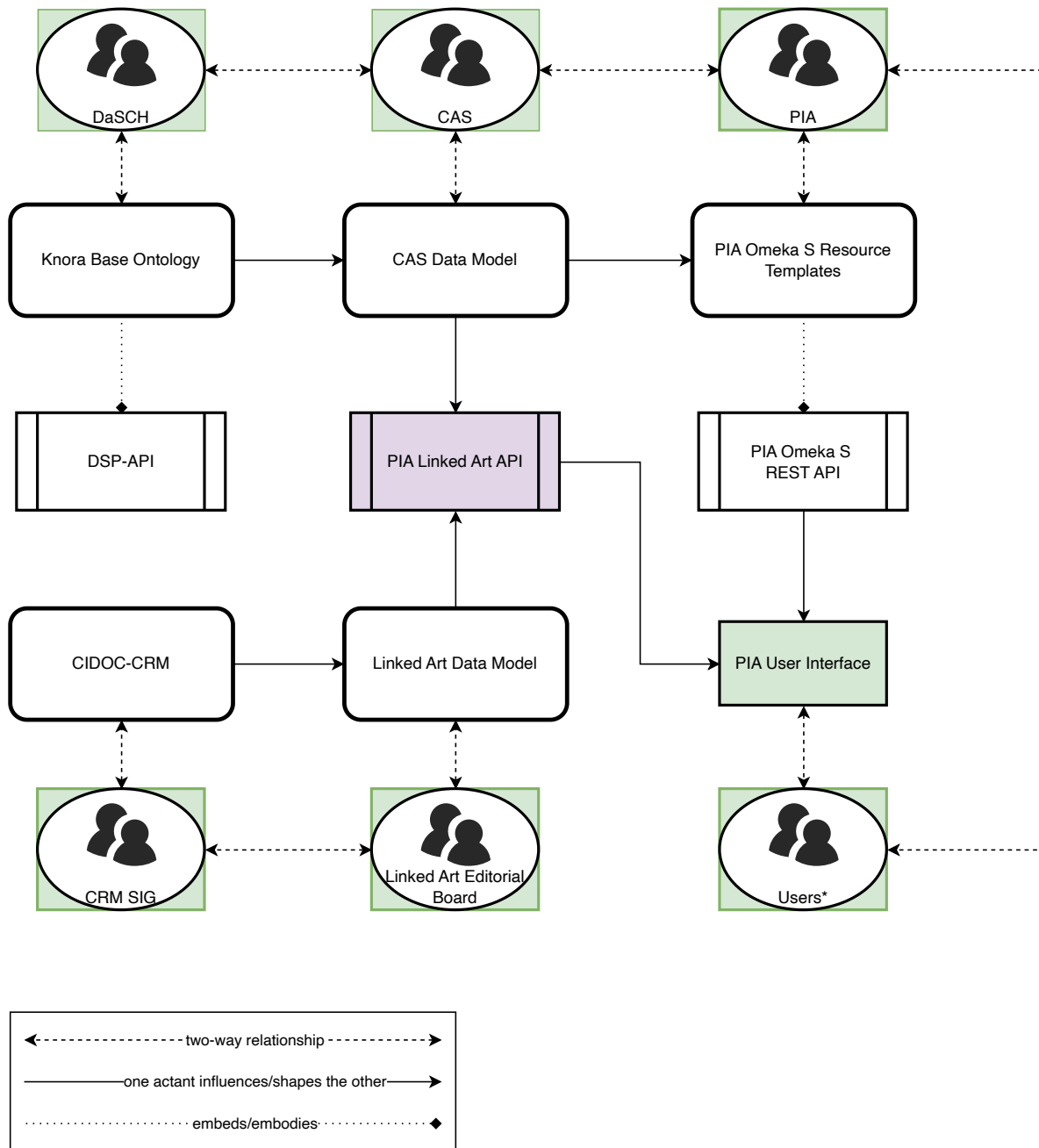


Figure 7.9: The Main Actors and their Associations around the PIA Linked Art API

While the Manifest serves as the fundamental unit within IIF, Linked Art protocol can assume a similarly central role as semantic gateways in broader contexts. Figure 7.9 illustrates an ANT perspective to highlight the potential of Linked Art as a central hub for connecting metadata, analogous to the role of IIF in promoting common modelling practices. The diagram illustrates how the Linked Art data model, together with the CAS Data Model, influences the structuring and granularity of the PIA Linked Art API and guides its integration. This network acts as a critical conduit for weaving together disparate data threads.

The exploration within PIA went beyond simply establishing common practices and adhering to open standards; it involved thinking about the future of such an ephemeral infrastructure. Echoing [Star \[1999\]](#), this process highlighted the inherent precariousness of computing, which in turn necessitated regular real-world testing to assess its resilience and effectiveness.

The materiality of digital objects is not limited to their immediate electronic form, but is extended by the infrastructure that supports them, encompassing micro-services, software and hardware, as [Müller \[2018, p. 56\]](#) considers. These components are integral, not peripheral, and can either enhance or limit the utility and prominence of digital artefacts. In addition, the scalability of such a system should be further explored, and it will be necessary to see to what extent these open standards, maintained by communities or built around community practice, can serve as an example and thus become what is referred to in [Hyvönen \[2020\]](#) as a third generation of portals that can automatically solve research questions according to the constraints set by researchers and experienced users.

Moving on to the final section of this chapter, I will consider the role of LOUD standards and their associated components in relation to participation, setting the stage for a discussion of perspectives ahead and the wider implications of these findings.

7.4 Perspectives

The use of LOUD standards within this exploratory context has demonstrated the feasibility of rapid implementation of these APIs and underlines the need for a sustainable long-term strategy for their maintenance. While images can still be compliant with the IIF Image API using the DSP's SIPI image server, thereby alleviating back-end maintenance concerns, the IIF Presentation, and IIF Change Discovery would benefit from a consistent, albeit read-only, post-project operational approach³²⁶. This approach would primarily involve managing the costs associated with domain hosting, which could be sustained over an extended period. Alternatively, these costs might be further reduced or even eliminated if DaSCH decides to extend its hosting services to include these APIs³²⁷. A workflow proof of concept for the deployment of the IIF Presentation API at DaSCH is illustrated in Figure 7.10³²⁸. This strategy not only ensures the long-term accessibility of these digital resources but also aligns with future organisational hosting policies.

This strategy facilitates access without the need for ongoing back-end updates, ensuring that the digital assets remain accessible and the metadata continues to enrich the user experience. However, managing end-user annotations introduces another layer of complexity. Decisions about whether to

³²⁶ As mentioned earlier, the Linked Art API and the ETL pipeline have not been deployed in a production environment. Ensuring that these components are fully implemented and operational remains an ongoing task that may serve as a benchmark for collaboration with other CHIs.

³²⁷ It is anticipated that the IIF Presentation API 3.0 will also be accessible through DSP, a development I have been coordinating (cf. <https://github.com/dasch-swiss/daschiify>). However, the hosting of machine-generated annotations would still require additional infrastructure, and hotlinking could be employed so that the annotations remain visible on a given client.

³²⁸ This figure is also available in an enlarged format in Figure B.15.

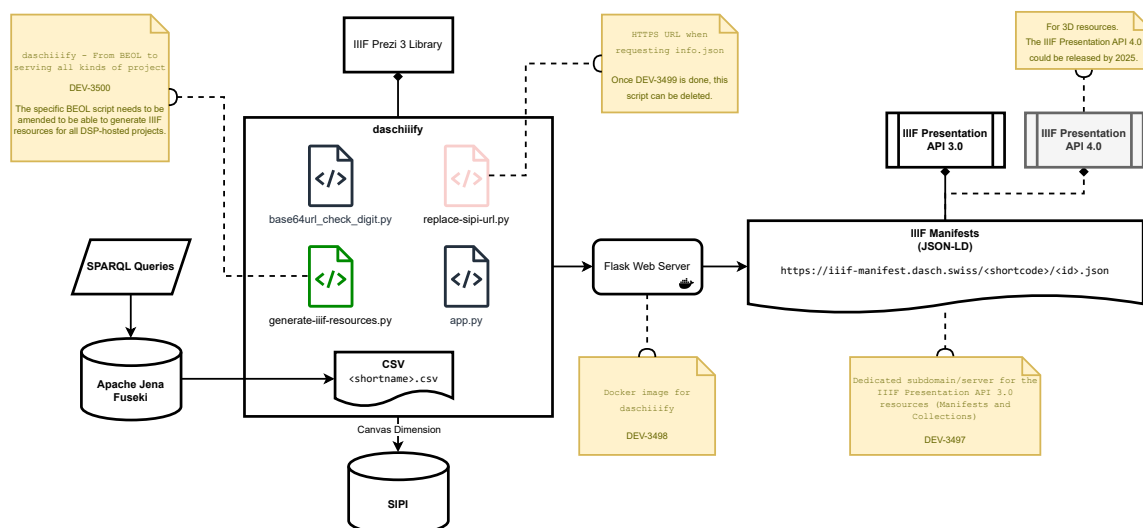


Figure 7.10: Overview of the IIIF Presentation API Workflow Proof of Concept at DaSCH

allow users to add annotations, either anonymously or with authentication, are not only technical, but also involve significant socio-political considerations. These decisions reflect wider implications for user engagement and the democratisation of data interaction, which are essential for the long-term viability and relevance of digital archives.

With regard to the IIIF Change Discovery API, if future maintainers, including CAS, choose to continue its development, several optimisations could be considered³²⁹ Parallelising fetch requests can significantly reduce data retrieval times by allowing simultaneous access to multiple data points, rather than waiting for each request to complete sequentially. This is particularly effective if the URL structure of the API is predictable, allowing URLs to be pre-constructed for concurrent fetching. Using HTTP/2 and asynchronous requests can also optimise the handling of multiple concurrent connections and reduce response times. It allows multiple requests to be interleaved on a single connection, reducing the latency associated with HTTP protocols [see Thomson and Benfield 2022].

In addition, improving data transfer involves adapting the **Accept-Encoding** to support compression formats such as brotli³³⁰, which can significantly reduce the bandwidth required by compressing the data before transmission [see Alakuijala and Szabadka 2016]. This is beneficial if servers support such encoding techniques, as it minimises the data payload and speeds up the transfer. Finally, rethinking API pagination to accommodate larger datasets per page or fetching from both ends of the dataset reduce the total number of requests required.

In exploring further possibilities for the Ernst Brunner Thesaurus, the integration of HTR and LLM technologies offers a promising avenue. This approach could automate the extraction of terms and their original identifiers from the index cards on which they are catalogued. This would not only facilitate the process of digitising and preserving this valuable metadata, but would also greatly enhance the richness of the semantic information available in our databases.

With regard to the use of ARK identifiers, if there is a need to host the controlled vocabulary elsewhere or to restart ARK assignment for all published digital artefacts, the authority number assigned to PIA could still be used effectively. It is conceivable either to work with another third party that

³²⁹ Some of these considerations are elements that have been discussed on the IIIF Slack Workspace technical channel.

³³⁰ brotli: <https://github.com/google/brotli>

supports ARK resolution, or to develop an in-house mechanism similar to that used by DaSCH with its ARK resolver³³¹. Such flexibility in the management of ARK identifiers ensures continuity and scalability of access to digital resources, in line with practices at other institutions [see [Huynh et al. 2022](#), [Peyrard et al. 2014](#)].

As I reflect on the PIA project's use of LOUD standards to enhance digital practices, it is also appropriate to consider broader participatory trends in the CH sector. While PIA has incorporated various innovative technologies and methodologies to facilitate access and interoperability, broadening its scope to include participatory and crowdsourcing efforts offers further avenues for enrichment. Such strategies have not yet been fully explored within PIA on a practical and large scale, but observing and learning from other projects could provide valuable insights and models for future development.

The innovative approach taken by the Swedish National Archives provides a compelling case study in participatory archival practices, as detailed by [Kasperowski et al. \[2024\]](#) in their examination of the project 'The Detective Section'³³². This initiative used AI-assisted transcription of historical documents to effectively involve the public. The project involved volunteers in transcribing 25,000 pages of 19th century police records, using citizen humanities to improve the accuracy and cultural relevance of HTR. The authors found that the emotional and personal ties of the volunteers to the archival content significantly improved the AI's learning process, resulting in a lower error rate in text recognition. Technical discussions in the paper reveal that the project used the Transkribus platform³³³, which allowed non-expert users to contribute to AI training by correcting machine-generated transcriptions. In doing so, the archives not only democratised access to historical information, but also enriched Sweden's CH with diverse interpretations that reflect broader community involvement. The success of this project could provide an important model for the PIA research project, suggesting that similar participatory methods – which [Kasperowski et al. \[2024\]](#) praise for its ability to merge technical efficiency with rich, community-driven content curation – can be used to effectively engage the public. While HTR itself is not central to the PIA research project³³⁴, the methods used for public engagement and data enrichment provide valuable insights.

With the implementation of IIF APIs it will be easier to use proxy mechanisms to send and receive metadata, further enriching the archival resources. The PIA project can thus explore the synergy between advanced digital tools and humanistic insights to enrich its heritage and expand public engagement in historical research.

To conclude this chapter, I have, with the help of my colleagues, as well as members and critters from the wider IIF and Linked Art communities, navigated the entanglements of implementing LOUD standards and following their design principles within the PIA project, demonstrating both the successes and the complexities involved. Looking forward, the next chapter moves to a different scale, examining how these standards have been adopted and adapted within Yale's discovery platform, LUX. This comparison will highlight differences in scale, scope and methodology, providing a broader perspective on the adaptability and benefit of LOUD in different environments.

³³¹ DSP ARK resolver: <https://github.com/dasch-swiss/ark-resolver>

³³² Detektiva avdelningen: <https://eu-citizen.science/project/326>

³³³ Transkribus: <https://www.transkribus.org/>

³³⁴ Apart from perhaps handwritten records made by people involved in collecting evidence for the ASV collection, and the identification of terms used by Ernst Brunner to label some of his photographs in the eponymous collection.

8

Yale's LUX and LOUD Consistency

'The LUX initiative has created a model of collegial and productive collaboration that colleagues at Yale can look to. This collaborative approach extends beyond Yale, and I hope it has a trickle-down effect on the thinking of those who develop systems for libraries and museums. It remains to be seen how this initiative will influence the broader community and shape the future of museum system development.'

Heather Gendron, Director, Robert B. Haas Family Arts Library, YUL

This chapter initiates the third and final empirical investigation of the thesis, exploring the application of LOUD specifications on the LUX platform at Yale University, as detailed in Section 8.1. Additionally, it examines the consistency of Linked Art and IIIF Presentation API resources within LUX, as discussed in Section 8.2.

The impetus for this investigation arose from preliminary insights obtained from the PIA research project, which suggested that relying almost exclusively on those empirical findings may not fully capture the complexities of developing and sustaining LOUD-compatible infrastructures. Following an important PhD midway meeting in February 2023 [Raemy 2023d], which underscored the need for more practical engagement, I arranged a ten-day visit to Yale in May 2023. This visit, a direct follow-up to an in-person Linked Art meeting, aimed to deepen my understanding of the specific challenges and strategies involved in implementing LOUD.

The overarching goal of this investigation is to illustrate the diverse practices Yale has undertaken in its extensive adoption of LOUD standards, and to identify the ongoing challenges posed by specialised skill requirements and resource constraints within CHIs. The study aims to demonstrate how overcoming these challenges and ensuring the overall alignment of Linked Art and IIIF across the LUX platform and further afield can significantly advance shared methodologies and accessibility. This in turn could facilitate wider and more effective use and sharing of data, underlining the critical role of LOUD norms in the development of data stewardship in the CH sector.

As with the previous chapters, the findings of this research are compiled in Section 8.3. This section provides a comprehensive summary and delves into the findings regarding the use of LOUD standards at Yale, as well as examining how well their resources align with the Linked Art and IIIF Presentation APIs. The chapter culminates in Section 8.4, which offers forward-looking perspectives for Yale and other CHIs. This final section reflects on the lessons learned during the course of the study and discusses potential strategies. It also considers the wider implications for the CH sector, suggesting ways in which institutions might address similar challenges and seek to improve semantic interoperability through community and local socio-technical practices.

8.1 Large-Scale Deployment of LOUD at Yale

This section examines the use of LOUD at Yale University. It is divided into three parts: the background to the LUX initiative is detailed in Section 8.1.1, the technology stack used is outlined in Section 8.1.2, and a series of interviews conducted in 2023 with Yale staff members involved in developing LUX is covered in Section 8.1.3. This structure provides an account of the deployment, from its conceptualisation to its technological realisation, as well as practical insights from its implementers.

8.1.1 Background

The trajectory of the LUX platform³³⁵ at Yale began over a decade ago, marked by a series of strategic developments aimed at improving digital access and infrastructure across CH units. The journey began in 2008 with the establishment of the Office for Digital Access and Infrastructure and the creation of the first shared DAM system. This system initially connected the collections of Yale University Art Gallery (YUAG), Yale Center for British Art (YCBA) and Yale Peabody Museum of Natural History (YPM), providing a basic framework for subsequent digital integration efforts. In fact, the development of LUX was built on a series of incremental steps.

In 2011, Yale University adopted an OA policy and launched its first-generation cross-collection discovery platform, which included digital assets from its major CH collections. This platform was a precursor to the LUX system and demonstrated Yale's commitment to making its vast resources more accessible and connected.

The appointment of Susan Gibbons in 2016 as first Vice Provost for Yale's Collections and Scholarly Communications (CSC) significantly advanced these initiatives. Her leadership coincided with the creation of the Yale's Cultural Heritage Information Technology (CHIT) Pillar in 2017, which brought together CH directors and Yale's Information Technology Services (ITS) representatives on a regular basis, fostering a collaborative environment critical to the subsequent phases of the project.

2018 saw the launch of a five-year project to improve access to Yale's collections, supported by initial project funding and an expanded scope to include Yale University Library (YUL) bibliographic records and archival finding aids. The following year, 2019, was pivotal, with the adoption of IIIF as the image standard for the collections, the beginning of project management, and a significant grant from the Andrew W. Mellon Foundation to support metadata reconciliation and harmonisation efforts.

The proof of concept, based on Blacklight³³⁶, for LUX in 2020 united data from over 15 million object records into a shared index, marking a major milestone in the project. Despite budget cuts due to the COVID-19 pandemic, the project continued with the refinement of the metadata schema and the addition of Robert Sanderson to lead these efforts³³⁷.

Although this proof of concept was an achievement in that it demonstrated the ability to integrate data from all units. However, it also showed that the interface did not meet the functional expectations. As a result, a technology evaluation process called a 'bake off' was initiated. Teams of community members from across CHIT - the museums, YUL and ITS - looked at a number of technology stacks and made an informed recommendation within a few months. LUX has been governed by several committees and groups³³⁸. The CHIT Steering Committee oversees all strategic decisions, supported

³³⁵ LUX: <https://lux.collections.yale.edu/>

³³⁶ Blacklight: <https://projectblacklight.org/>

³³⁷ As a matter of fact, the development and refinement of LUX actually benefited significantly from the effects of the pandemic, as meetings were moved to Zoom, increasing productivity and inclusiveness [see [Bishop 2023](#)].

³³⁸ LUX - Who We Are: : <https://lux.collections.yale.edu/content/who-we-are>

by technical and interdisciplinary collaboration teams such as the Cross-Collections Discovery Working Group, Cultural Heritage Information Technology Architecture, the Metadata Working Group, and the Bias Awareness and Responsibility Committee.

In 2021, after a thorough six-month evaluation of technologies, the LUX project entered the build phase of LUX 1.0. By 2022, all metadata from participating units had been ingested into LUX, and internal reviews of the beta version began, leading to its public release in May 2023.

At the time of writing, LUX integrates over 17 million objects, over 5 million agents (people and groups), over 600,000 places, some 4.8 million concepts, over 38,000 events, and over 13 million works from Yale's diverse collections (as seen in Figure 8.1), facilitating access and connectivity through a platform built on Linked Art. This format is not only appropriate for art museums, but has also proven adaptable to mapping library and archival data, thereby addressing cross-disciplinary challenges and ensuring that the LUX platform remains a vital resource for scholars, researchers and the general public, enabling detailed exploration of Yale's vast CH resources.

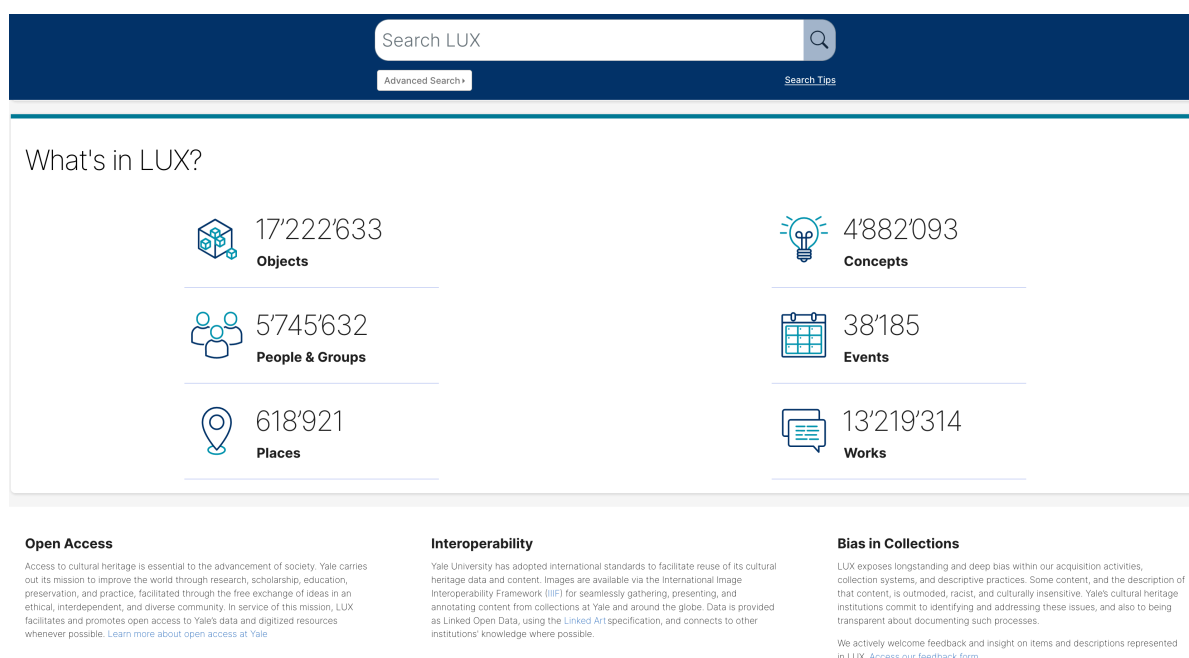


Figure 8.1: Snapshot of the LUX GUI (23 May 2024)

Through collaboration and a shared understanding of responsibilities, Yale has established a data architecture in which each unit is responsible for providing data in a format and structure that supports a consistent approach to data harvesting. This approach allows for easy updates and future enhancements. This collaborative effort has not only facilitated numerous improvements to the dataset, but has also led to advances in the technologies used.

The following subsection on the technology stack, in particular the LUX Data Transformation Pipeline, looks at the specific technologies and strategies used to achieve these optimisations and functional integrations.

8.1.2 Technology Stack

This subsection explores the comprehensive technology stack that underpins the LUX platform at Yale, detailing its architectural components, collaborative development processes and integration with

external data sources. It covers the initial design principles, system components, data management procedures and continuous data quality improvement efforts. Key elements of the architecture and data pipeline are illustrated in Figure 8.2 and Figure 8.3, providing a visual representation of the workflows and technologies involved.

LUX was initially designed with a strong emphasis on standards compliance and interoperability [Sanderson 2023a]. This design approach includes the use of AS and JSON-LD for efficient data harvesting, Linked Art to facilitate data sharing across domains, and IIIF for seamless image delivery. To enable efficient data interpretation and simplify the work for software engineers, each collection exports their content using JSON-LD, which can be interpreted as triples. The choice was made to employ a low-level ontology, specifically Linked Art. LUX also uses HAL to generate the accordions in the GUI through the `_links` properties.

As noted by Sanderson and White [2023], LUX functions primarily as an aggregator, with individual institutions retaining ownership of their data. This arrangement fosters a highly collaborative environment, characterised by engagement and experimentation that drives consensus building. Described as ‘contagious’, this cooperative spirit fosters community-wide collaboration. Guided by agreed models for governance and cost-sharing, LUX integrates LOUD standards and adheres to the self-defined Sustainable, Harvestable, Available, Reconciled, Enriched, Diverse (SHARED) technical principle. The platform’s front-end facilitates user interaction by seamlessly handling all Linked Data under the hood, eliminating the need for stakeholders to learn SPARQL. In addition, LUX’s advanced search capabilities feature graph queries that balance search depth and performance, limiting computation times to one minute to optimise responsiveness and usability.

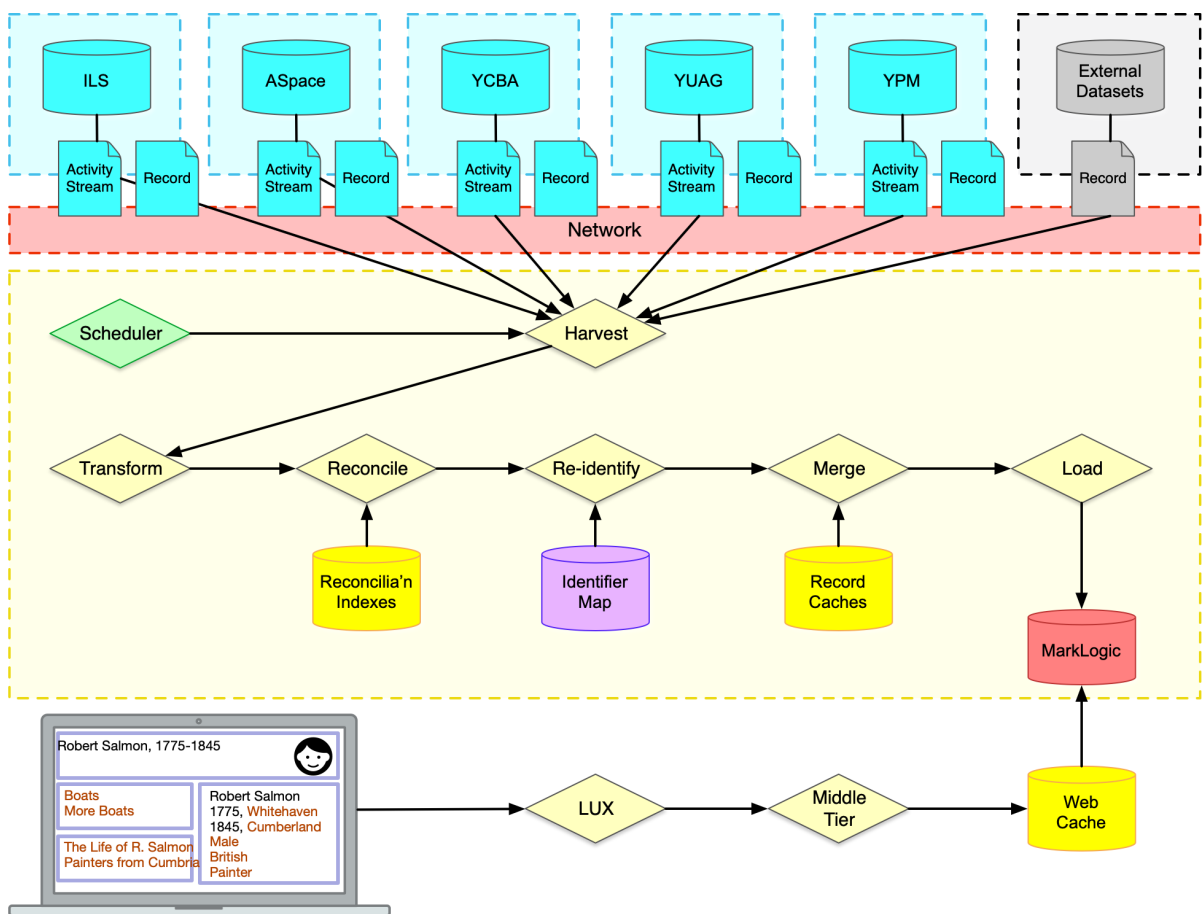


Figure 8.2: LUX Data Pipeline and Architecture

The architecture of LUX consists of several interconnected components: data harvesting, data pipeline, back-end database, middle tier and front-end. These components have been integrated according to some of the established standards mentioned earlier, allowing any individual component to be replaced without requiring a complete rewrite of the system. Hosted by Amazon, LUX benefits from the widespread adoption and robust support within the software development community for technologies such as Python, JavaScript, Node.js and React.

Moreover, LUX harnesses the power of an enterprise-class, multi-modal back-end database, increasing its efficiency. To encourage collaboration and knowledge sharing, the code for all components of LUX, including the custom code for MarkLogic³³⁹, has been made open source and available to the wider community³⁴⁰ [see Cahill 2024].

The LUX platform is characterised by its extensive connections, not only within Yale units, but also beyond, through the inclusion of external data sources during data processing. These sources cover different subject areas and perspectives and are available as Linked Data, enriching the information available to users by matching records in LUX. For example, one of the key processes used to match works and objects is the integration of Wikidata records, providing the ability to make meaningful associations between objects from units, such as those within the YCBA collection, and related works. They also include additional names and terms from authority records and subject headings of libraries like GND³⁴¹, LoC³⁴², or the National Diet Library (NDL)³⁴³, as well as the Virtual International Authority File (VIAF)³⁴⁴. This strategic approach requires careful anticipation of the requirements for Linked Data integration, accompanied by a concerted effort to foster extensive collaboration within the YCBA unit. In particular, the scope of authority control procedures at YCBA encompasses all collections, further emphasising the need for extensive collaboration. In addition, collaborative efforts extend beyond the YCBA unit to include partnerships with external services and inter-unit collaborations between the YCBA and the YUAG. This collaborative approach is particularly evident in the joint exhibitions mounted by the two units.

Figure 8.2 depicts the overall data transformation, reconciliation, enrichment and publication workflow for LUX. The base records come from both internal and external sources, with internal records being harvested through the IIF Change Discovery API [see Raemy and Sanderson 2024]. The processes (represented by diamonds) are described hereafter. The source code of the pipeline, which notably includes the identifier map of equivalencies and their associated indexes, was also released on GitHub in March 2024³⁴⁵.

Harvest It runs nightly and is triggered by an operating system level scheduler to poll each stream to find and retrieve records that have changed since the previous harvest.

Transform The records are passed through source specific transformation routines in order to either map from arbitrary data formats, or to validate and clean up records already provided in Linked Art.

Reconcile This process is conducted to discover further identities from the various datasets to be able to collect all information about a particular entity eventually into a single record.

³³⁹ MarkLogic: <https://www.marklogic.com>

³⁴⁰ LUX code base: <https://github.com/project-lux/>

³⁴¹ GND: <https://gnd.network>

³⁴² LoC Linked Data Service: <https://id.loc.gov>

³⁴³ Web NDL Authorities: <https://id.ndl.go.jp/auth/ndla> and <http://id.ndl.go.jp/auth/ndlsh>

³⁴⁴ VIAF: <https://viaf.org>

³⁴⁵ Data Transformation Pipeline Code: <https://github.com/project-lux/data-pipeline>

Re-Identify It maps the original URIs of the records to the internal identifiers.

Merge The records from multiple sources that have been mapped to the same identifier are merged together to form the single record.

Load The resulting dataset is then annotated with some additional features for indexing and exported to MarkLogic.

Figure 8.3 shows an annotated version of the data pipeline, detailing specific Python scripts required for each step of the process or for handling different datasets. For example, when mapping and establishing dependencies, it's necessary to consider the different systems used by different units, such as Integrated Library System (ILS) and Gallery Systems' The Museum System (TMS). A specific example is the YCBA, where data harvesting moves from LIDO format to Linked Art. In addition, the middle tier of the architecture manages four separate instances of LUX – primary, backup, development and sandbox – each customised for specific operational needs. The platform uses code libraries such as LinkedArt.js (JavaScript) and Cromulent (a Python library), which include built-in validation and convenience functions [Sanderson 2024b]. These tools exemplify the level of engineering sophistication of the LUX platform, guaranteeing its adaptability and efficiency in a variety of data integration scenarios.

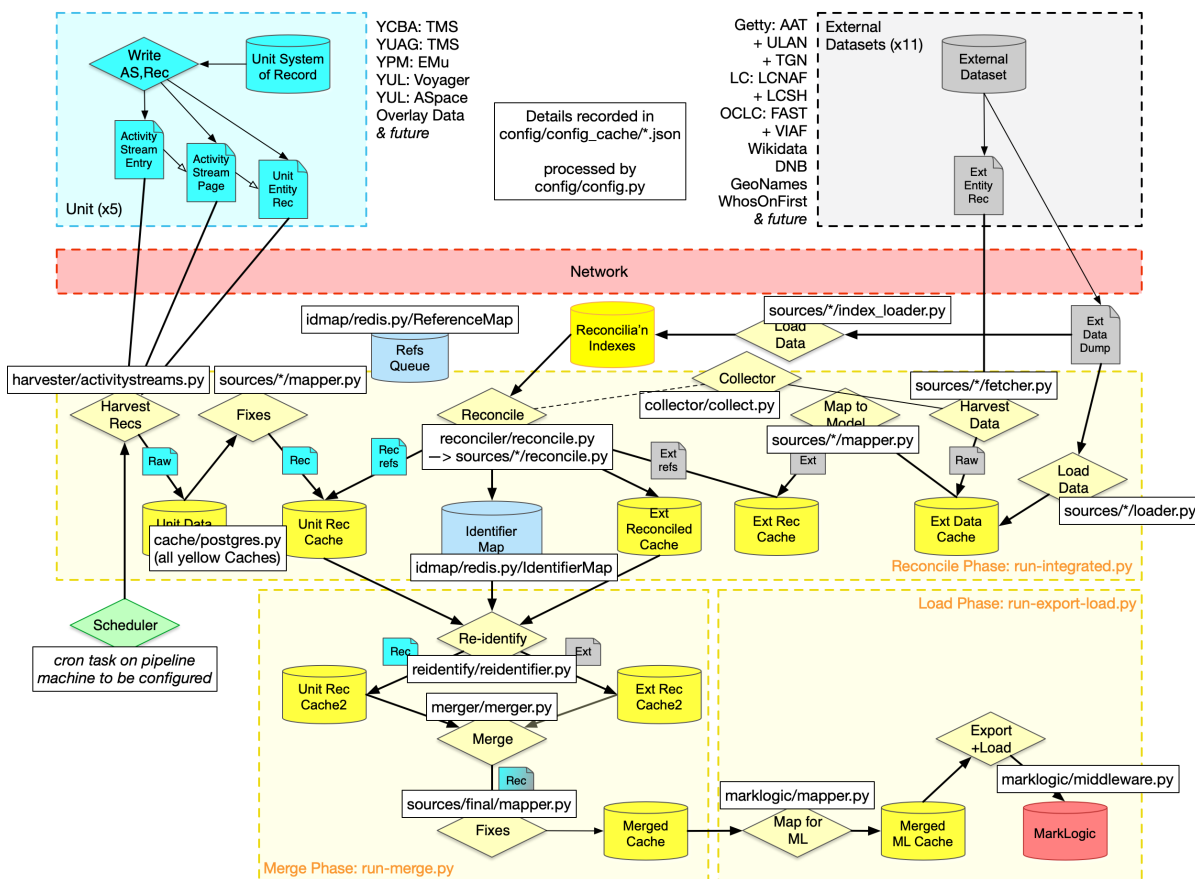


Figure 8.3: Annotated LUX Data Pipeline and Architecture

This integration significantly enhances the record by combining knowledge from different Yale units, the Getty's ULAN, European national libraries and other external sources. This is an example of the positive impact and improvement that can be achieved by linking disparate data sources. The linking process is automated within the data processing code, using equivalent URIs and intelligent matching of names associated with people, places and things. However, matching and merging data

into a single LUX record can be a complex task. Data quality is affected by human imperfections, as all data is derived from human input. As a result, efforts are constantly being made to refine and improve the matching process as part of a continuous improvement process.

Having discussed the technology behind the LUX platform, I now turn to the perspectives of a few of the Yale staff who worked on it. The next section is an account of interviews conducted in 2023 with team members involved in the development of LUX, offering a reflection on the project shortly after its official launch, their experiences, the challenges they faced, and the results of their efforts. It provides a personal account of the collaboration and technological development that went into LUX.

8.1.3 Interviews

In May 2023, a number of interviews were conducted with 20 individuals associated with the LUX project at Yale University³⁴⁶. These interviews, which included both one-on-one and group sessions with up to four participants, took place over nine separate sessions. Two of these sessions were conducted via Zoom to accommodate remote participants, while the remaining seven were held face-to-face on the Yale campus in New Haven, CT, USA. Each interview was designed to last approximately 45 minutes and prior to these sessions, each participant was provided with a consent form and a prepared questionnaire to streamline the discussion and ensure that all relevant topics were adequately covered.

The consent form explained the aims of the study, the voluntary nature of their participation, their right to withdraw at any time, and detailed the process of audio recording and verbatim transcription for analysis as part of my PhD. It emphasised that the data collected would be stored securely, accessible only to myself, and could be used for future research with all identifying information removed or pseudonymised to ensure confidentiality. One copy of the transcript was also given to all participants to review and confirm their comfort with the recorded content.

To further protect the integrity of participants, those who agreed to have their names disclosed and their statements published underwent a secondary verification process. This process allowed them to review and amend their contributions in order to ensure accuracy and satisfaction with the representation of their views. Of the 20 participants, thirteen agreed that their names and any quotes from them could be published in this thesis or any related report.

8.1.3.1 Themes and Questions

The interview questions were systematically organised around four main themes: Mission & Values, Collaboration, Technology and Principles. This organisation helped to guide the interviews, although not all questions were asked of each participant. Instead, the selection of questions was specifically tailored to the expertise and role of each interviewee, allowing for more focused and meaningful discussions. For example, technical questions about the infrastructure or data interoperability of the platform were directed at software developers and data engineers, while broader questions about the impact and overarching values of LUX were asked of a wider audience. This approach ensured that each session was both focused and adaptable, providing comprehensive feedback on both the operational and strategic dimensions of the LUX platform, as detailed in Table 8.1.

A dual method of capturing data was adopted during the interviews: I transcribed key points directly into a spreadsheet for real-time analysis and tracking, while simultaneously recording the interviews for comprehensive documentation. This approach allowed for immediate adjustments to the

³⁴⁶ The 20 individuals interviewed represented various Yale units: 6 from ITS, 2 from CSC, 8 from YUL, 1 from YUAG, 2 from YCBA, and 1 from YPM.

Table 8.1: LUX Interview Questions

Topic	Questions
Mission & Values	1) How would you define the LUX platform in a few words? 2) What impact has LUX had internally and/or outside Yale? 3) What does LUX do that previous systems haven't?
Collaboration	4) Has LUX succeeded in bringing together people from different units at Yale? Were you equipped with all the necessary tools or methods to work together? 5) How did you manage to build consensus? Can you give me an example of a successful and unsuccessful scenario of consensus building? 6) Has the LUX platform brought you closer to national or international communities and working groups (such as LD4, IIF or Linked Art)?
Technology	7) Infrastructure a) For everyone: What do you find innovative about the technological foundation you have deployed? 7) Infrastructure b) For software developers and data engineers: Can you explain the steps involved in designing the LUX data pipeline? 7) Infrastructure c) For software developers: How do you plan to measure usage of the system? 8) Data Usability a) For everyone: To what extent is the data on the LUX platform easily reusable (on the user interface or/and with the APIs)? 8) Data Usability b) For software developers and data engineers: How usable was the data for building LUX? 8) Data Usability c) For software developers and data engineers: How can you ensure data concurrency and consistency? 9) Interoperability a) For everyone: How do you assess the mapping process that has been done to expose the metadata in Linked Art? Would you have used other standards and/or serialisation formats? 9) Interoperability b) For software developers: How much computational effort is required to generate IIF and Linked Art resources? 9) Interoperability c) For software developers, metadata experts, data engineers, and semantic architects: How to assess the semantic interoperability of the data produced for LUX? How to ensure that the data can be easily reconciled with other datasets?
Principles	10) The LOUD standards, which include the IIF Presentation API and Linked Art, have their own design principles and LUX is based on the FAIR data Principles and SHARED technical principles. a) Have these principles provided you with any real guidance? b) Which of these principles do you feel have been achieved? c) Are there any other principles (or norms, best practices) that you feel are relevant to LUX?
Misc	Further comments

questioning process - questions could be rephrased or redirected on-the-fly based on the flow of the conversation and preliminary findings from the real-time notes.

To convert the recorded audio into text, Whisper³⁴⁷, an Automatic speech recognition (ASR) system, was used. Whisper is designed to provide robust and accurate speech-to-text capabilities and supports multiple languages. It has been trained on a wide range of data collected from the internet and has shown considerable ability to understand different types of accents and dialects, making it an ideal choice for transcribing the varied responses of the interview participants [Radford et al. 2022]. The transcription process was facilitated using the CLI. The specific commands executed to transcribe the interviews from audio files into text are detailed in Code Snippet 8.1.

³⁴⁷ Whisper: <https://github.com/openai/whisper/>

Code Snippet 8.1: Command to Transcribe the Interviews into Audio Files Using Whisper

```
1 ~ % for f in *.wav ; do whisper $f --fp16 False ; done
```

This sequence is designed to systematically process each `.wav` file located in a specified directory to transcribe the audio content into plain text. It is a Bash loop that iterates over all `.wav` audio files. For each file, it invokes Whisper with the parameter `--fp16 False`, which specifies that the floating point precision should be maintained at 32 bits for better accuracy during the transcription process. This option is particularly useful when the quality of audio files varies or when the highest possible accuracy is required, as it ensures the neural network computations retain more information during processing.

The command does not specify a particular model, which means it defaults to the `base` model unless otherwise configured. This model strikes a balance between speed and accuracy, making it suitable for general purposes where the audio quality is reasonable and the content does not require intensely detailed linguistic analysis. Each transcription is then saved in a separate text file corresponding to its `.wav` file, ensuring that the data remains organised and each session's content is easily accessible for further review and analysis. Whisper automatically detected that the recordings were in English and adjusted its processing to optimise for English language transcription.

Once transcribed, the raw text files were converted into a more manageable format for content analysis. The conversion process involved formatting the transcribed text into `.md` files, which are easier to handle and annotate than plain text files. This was achieved by using the `sed` command to insert newline characters between successive blank lines in the text files, effectively creating a new paragraph for each spoken segment. The `awk` utility was then used to ensure proper line spacing in the final Markdown documents. The command used for this conversion is shown in Code Snippet 8.2.

Code Snippet 8.2: Command to Convert Text files into Markdown with New Lines

```
1 ~ % sed -e 's/^\$/\n/g' transcript.txt \  
2 | awk '{ printf "%s\n\n", $0 }' > transcript.md
```

These Markdown files then served as the primary source for detailed analysis, allowing easy navigation through the interview content and facilitating data management and retrieval.

8.1.3.2 Findings

Upon reviewing the transcriptions, the interviews revealed a rich description and perception of LUX. The following passages provide an overview of these findings, supported by selected quotes from participants. This is followed by a distant reading process using topic modelling to uncover further threads of interest.

LUX is perceived as a ground-breaking search and discovery tool that brings together disparate CH data, enabling users to explore and discover connections between collection items. The platform is seen as having transformative potential for teaching, learning and research. Internally, it has fostered collaboration, increased the visibility of participating units, and demonstrated the feasibility of integrating current technology. Externally, LUX has attracted significant interest from other institutions, highlighting its potential impact on a wider scale.

Emmanuelle Delmas-Glass, Head of Collections Information Access at YCBA, expressed some nuanced concerns:

'I do have concerns about how LUX is perceived at times and its positioning as a highly sophisticated technical tool. While it certainly possesses advanced capabilities, it is important to remember that it is not solely defined by its technical complexity.'

Emmanuelle Delmas-Glass, Head of Collections Information Access, YCBA

Collaboration across units at Yale has been a notable success of the LUX project, fostering interactions between libraries and museums. However, challenges in decision-making have been identified, particularly around the issue of ownership. Despite Susan Gibbons' leadership, the question of who owns LUX remains unresolved. Efforts to promote LUX through training and campus discussions have faced hurdles in terms of resource allocation and division of responsibilities. The impact of COVID-19 has been unexpectedly positive, increasing participation and trust through virtual meetings. Heather White, User Experience Designer at ITS, highlighted the project's dedication to sustainable development:

'Unlike previous projects at Yale, LUX stands out as the first one with a post-launch plan, demonstrating a commitment to sustainability and ongoing development. This is in stark contrast to the common practice of addressing issues and implementing requested features on an ad hoc basis without a clear roadmap. LUX, on the other hand, has upcoming releases and planned features that were not initially included, indicating a long-term staffing and development strategy. However, it is worth noting that, similar to most projects at Yale, there is still a lack of dedicated product management.'

Heather White, User Experience Designer, ITS

The consensus-building process for mapping metadata in Linked Art, while generally successful, faced some disagreements and challenges, particularly between data owners and mapping teams. The iterative, collaborative process involving Linked Art and JSON-LD proved beneficial. Future projects could benefit from consolidating collaborative norms and achieving coherence in both technical and social aspects. Long-term support and prioritisation of LUX amidst staff turnover and changing institutional priorities requires careful planning and commitment. Peter Clarke, a Full Stack Engineer at ITS, shared his experience with Linked Data:

'Working with Linked Data was an interesting experience for me, as it revealed the fascinating connections that can be discovered across different products. It provided me with valuable exposure to triples and expanded my understanding of document databases and JSON-based indexing. This shift in technology, particularly in how we handle data, has been the most significant change for me personally.'

Peter Clarke, Full Stack Engineer, ITS

Key individuals played a pivotal role in the success of LUX. Susan Gibbons' high-level support as Vice Provost facilitated trust and established LUX as a layer of discovery without replacing existing units. Sarah Prown's role as project manager was central to fostering collaboration and audience engagement, with norms and boundaries that built trust. Robert Sanderson's facilitation and collaboration with developers was also significant, despite some challenges in communicating the potential of Linked Art. His strategic role, particularly with a documented API, is expected to encourage future cross-collection projects such as hackathons.

Alison Clemens, Access Strategist for Special Collections at the Beinecke Rare Book & Manuscript Library, emphasised the impact of Susan Gibbons' leadership.

'The success of this iteration of cross-collection discovery can be attributed, in large part, to Susan Gibbon's role as Vice Provost, who initiated the project and revitalised cross-campus collaboration efforts. This new approach to cross-collection discovery has resulted in a significant investment of staff time and resources.'

Alison Clemens, Access Strategist for Special Collections, Beinecke Rare Book & Manuscript Library, YUL

LUX's technological foundation is considered innovative, providing a cross-collection discovery platform that integrates and links digital data more effectively than previous systems. LUX is known as the largest known CH Linked Data platform, based on open standards such as Linked Art, IIF and AS. The platform provides new entry points for accessing data, improving data integration and organisation. However, the usability of the data varied between units, with significant challenges around data quality and consistency, particularly from YUL. The lack of dedicated resources for data cleansing and reconciliation at unit level hinders transparency and effectiveness, highlighting the need for specialised roles within each unit. Kelly Davis, Data Engineer at CSC, underscored the importance of data ownership and quality:

'I firmly believe that the owners of the data should be responsible for fixing any issues that arise, as the pipeline or external sources may not have the necessary context. However, addressing such issues requires significant time and resources. Improving the mapping process and leveraging high-quality external sources can enhance the usability of the data, although errors are inevitable.'

Kelly Davis, Data Engineer, CSC

There is also a consensus that the advanced search function needs further improvement to make it more intuitive. Anna Arays, Librarian for Slavic & East European Studies at YUL, shared her insights into the challenges faced by libraries and museums:

'The library's role is to provide quick, reasonable, and easy access to a large volume of materials. Our users typically have specific search goals, and while they do browse, their approach to discovery differs from object-based environments. Collaborating with museums revealed challenges that may seem straightforward to them but become real problems for our users. For instance, the object/works split in the platform, which stems from the distinction between intellectual content and objects. Balancing these elements to ensure user-friendly integration poses significant friction. Additionally, discussions on data privacy highlighted librarians' strong concerns about recording user search data. Achieving consensus required labour-intensive efforts to address the distinct needs of both user communities. Despite the complexity, this project fostered mutual learning and serves as a foundation for further improvements based on user feedback.'

Anna Arays, Librarian for Slavic & East European Studies, YUL

In addition, the separation of objects and works on the GUI has been confusing for many users. This challenge is deeply rooted in the distinction made by CIDOC-CRM and Linked Art between the intellectual content of a work and the physical Human-Made Object. While this separation facilitates more nuanced metadata and relational connections in databases, it often complicates the UX, particularly for those accustomed to viewing an object and its intellectual content as a single, unified entity.

Yer Vang-Cohen, Data and Database Administrator at YUAG, explained this difficulty:

'It was challenging to convey the distinction between work and the Human-Made Object, particularly to our curators who perceive the object as a unified entity. Although they have a general understanding of the visual entity record, it proved difficult for them to grasp the concept of splitting the record and having metadata about the visual content reside separately from the Human-Made Object.'

Yer Vang-Cohen, Data and Database Administrator, YUAG

LUX adheres to principles such as FAIR and LOUD, which have provided essential guidance, ensured interoperability and influenced its design and development. However, perceptions of these principles varied considerably. Project managers placed a strong emphasis on these principles, while developers, librarians and archivists considered them to be of secondary importance in the creation of LUX, with some individuals simply unaware of them. This variance underlines the need for wider awareness and integration of these principles across all roles involved in the project.

In terms of future potential, there is significant interest in the potential re-use of the LUX code base. Many believe that with a small, dedicated team possessing foundational knowledge in core technologies like JavaScript and React, it is feasible to adapt and extend the LUX system for new applications. Leveraging the work already done with MarkLogic could further streamline this process, making it accessible for research-oriented systems or other institutional needs. Jeff Campbell, Associate Director of Digital Asset Technologies at ITS, elaborated on this potential:

'I believe that a small team of two or three individuals, who are open to exploring new avenues and possess a sense of curiosity, could accomplish it. Ideally, they should have a foundational understanding of core technologies like JavaScript and React, so they can focus on specific requirements without starting from scratch. While MarkLogic is a robust system, the work we've already done with it can potentially be leveraged for their use cases, making the process smoother. If someone is interested in a research-oriented system similar to LUX, they could start with a single-node environment or opt for a powerful machine.'

Jeff Campbell, Associate Director, Digital Asset Technologies, ITS

After collecting these qualitative findings, I aimed to obtain a distant reading perspective of the interviews using simple topic modelling methods. To do this, I used an online topic extraction tool³⁴⁸. The tool performs the following steps:

1. Clean up the text: flatten to ASCII, remove URLs and punctuation marks.
2. Lemmatisation.
3. Decompose the text into n-grams up to four grams and remove less relevant n-grams.
4. Counting co-occurrences: identifying which pairs of n-grams occur frequently together.
5. Create a network from the list of co-occurring n-grams: connect frequent n-grams.
6. Apply a community detection algorithm to the network using the Louvain algorithm.
7. Apply the parameter chosen by the user: a large value detects a few large communities, while a small value detects many small clusters.
8. Each cluster in the network is identified as a topic. The key terms in each topic are the n-grams contained in the cluster.

³⁴⁸ Topic extraction tool for documents and social media: https://nocodefunctions.com/topics/topic_extraction_tool.html

I fed all the raw .txt files, which contained no names or identification, into the tool. I set the parameter to detect large topics, limiting the number to ten, and used a special stopwords file (e.g. LUX, data) to avoid overcrowding the results. The identified clusters can be seen in Figure 8.4³⁴⁹:

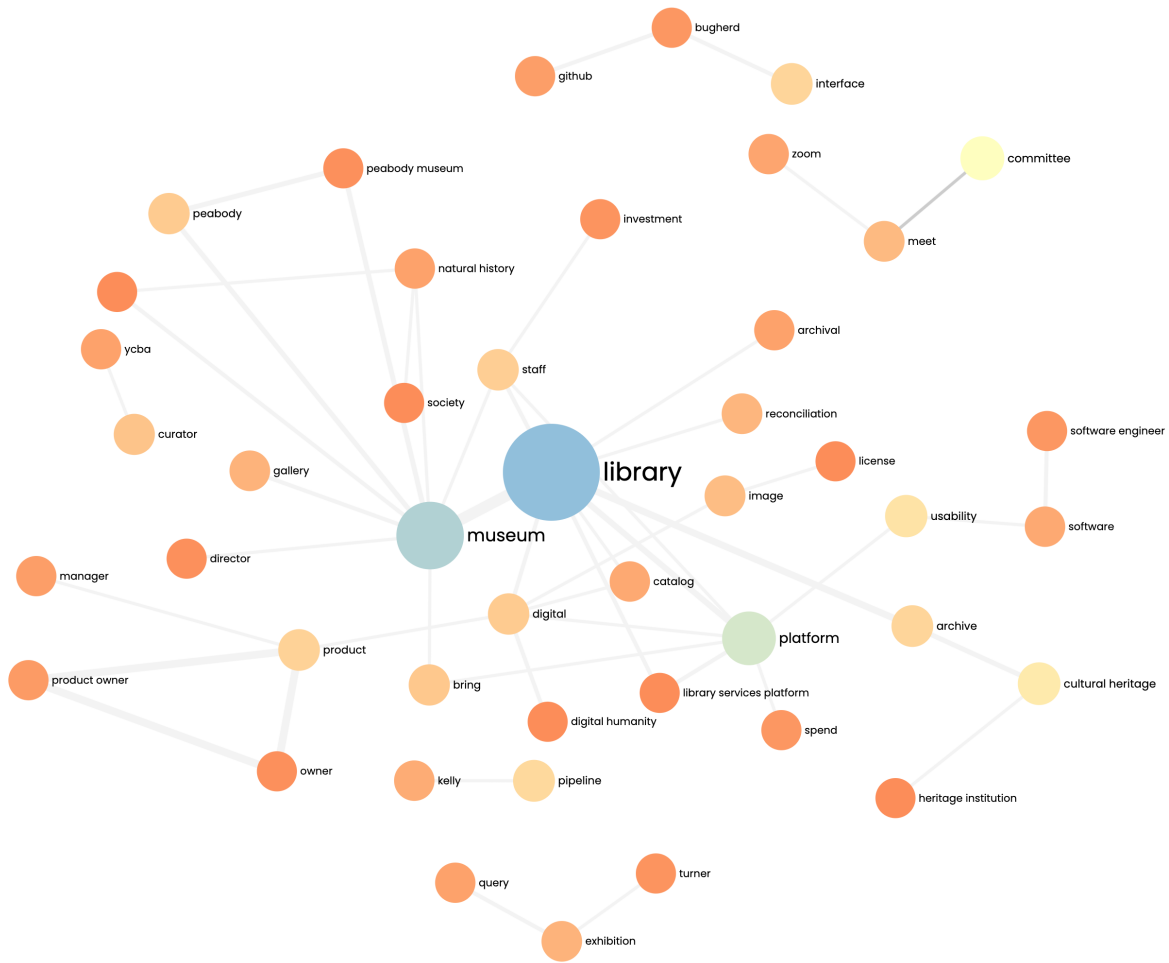


Figure 8.4: LUX Interviews: Key Topics and Their Relationships

The topic modelling results reveal a broad range of themes and focal points across the interviews. Below is a detailed analysis of each identified cluster:

Topic 1 This cluster centres around the pipeline, prominently featuring Kelly Davis, the Data Engineer. This indicates a significant emphasis on the technical aspects of data processing and infrastructure within LUX. The focus on a single individual highlights the pivotal role played by certain key personnel in the project's technical operations.

Topic 2 This topic focuses on YCBA and curators, suggesting strong involvement and interest from this particular unit and its staff. It could indicate that the YCBA has been particularly proactive or influential in the LUX project's development.

Topic 3 This cluster pertains to usability and software, reflecting ongoing discussions about the user experience and the practical implementation of software tools. It underscores the importance of ensuring that the GUI and functionality are accessible and effective for end users.

³⁴⁹ The figure can be recreated using the data provided in this Gist, which was fed into Gephi to create this SNA-based visualisation: <https://gist.github.com/julsraemy/027a194a1a1090b249ea08c2d03fbbee>

Topic 4 This topic highlights semantic and interoperability issues, which are critical for achieving seamless integration and use of data across different systems and platforms.

Topic 5 This cluster revolves around meetings and committee activities, with 'Zoom' being a prominent term. It reflects the impact of virtual meetings on collaboration.

Topic 6 This topic discusses product ownership and management, highlighting the need for clear leadership and accountability.

Topic 7 This cluster includes terms related to technical aspects such as 'Github' and 'Bugherd', indicating a focus on software development and issue tracking tools.

Topic 8 This topic is focused on museums, particularly the YPM, emphasising the role of specific institutions within the LUX initiative.

Topic 9 This cluster covers exhibitions and CH, with mentions of 'Turner' and 'Gallery', suggesting an emphasis on public engagement and the presentation of collections. It highlights the importance of exhibitions in making collections accessible and engaging to the public.

Topic 10 This topic focuses on library services and DH, with key terms like 'platform' and 'catalogue', indicating discussions about the integration of digital tools and resources in library contexts. It reflects the evolving nature of CH services in the digital space.

While these clusters provide valuable insights into the main themes discussed during the interviews, some limitations should be noted. For instance, the prominence of individual names, such as Kelly Davis, may skew the perceived importance of certain topics. Additionally, the clustering process might overlook nuanced discussions or smaller, yet significant, themes that are not captured by the dominant n-grams. Nonetheless, this topic modelling exercise offers a useful overview of the key areas of focus and concern within LUX.

Having explored the qualitative findings and performed a distant reading of the interviews data using topic modelling, I now turn my attention to the second objective of this investigation: assessing the consistency of Linked Art and IIF resources within LUX. The following section examines the tools and methods used to ensure compliance with LOUD standards and highlighting key findings from this analysis.

8.2 Linked Art and IIF Consistency

This section addresses the second, more technical, objective of the investigation around LUX. It is divided into two parts. Subsection 8.2.1 presents the available software tools used to validate Linked Art and IIF resources. The second part, subsection 8.2.2, examines how some of these tools were applied to verify the LOUD consistency of a portion of the resources available on the platform.

8.2.1 Ensuring Compliance: Tools and Validators

Verifying the compliance of JSON-LD documents can be done at several levels: first, from a purely syntactic perspective; then by checking if the `@context` is correct, meaning whether the correct properties and patterns (lists, arrays, and path construction) are used; and finally, by ensuring that the values of the elements within are correct. The latter is more complex when external controlled vocabularies are used, as is the case with Linked Art. Although certain terms are recurrent and some are mandatory or recommended, such as AAT terms, there is always some leeway for institutions implementing the API.

In this dissertation, I am primarily concerned with the second level, which unfortunately only partially addresses full validation at the level of semantic interoperability. However, the tools for this level are available and maintained. This validation process involves ensuring that the JSON-LD structure is correct, which includes validating that the `@context`, properties, and patterns adhere to the expected formats. This ensures that data can be hopefully interpreted and integrated across different systems.

Within Linked Art, two validators have been created: the first is a JSON validator³⁵⁰ testing resource compliance according to various entities, and the second is a SHACL validator³⁵¹. The former is still maintained, while the latter, though a commendable individual effort, hasn't been updated to accommodate all community needs.

Regarding validation and schema consistency, there are some affordances that have been allowed by the Linked Art community which can be found in the Linked Art JSON Validator and not directly in the schema documentation, such as having internal additional properties starting with underscores. For instance, LUX does this by appending `._links`, borrowed from HAL.

For IIIF, the available validation tools cover both the Image API and the Presentation API. The Image API validator³⁵² supports all major versions of the API and thoroughly tests the server's behaviour in delivering images, not just the quality of the images themselves. This validator can detect most potential issues by requiring the upload of a specific test image. It is available as a Python package, allowing integration into pipelines for automated testing.

The IIIF Presentation API validator³⁵³ works similarly by parsing resources (such as Manifest and Collection) and detecting common issues, particularly related to JSON-LD. While it ensures that the manifests are correctly structured, it does not verify the existence of referenced images. This validator can be used from its homepage, by making JSON requests, or by running it locally³⁵⁴. It is important to note that a valid Manifest or Collection does not guarantee that all its functionalities – patterns – will be properly displayed by IIIF viewers or players. Complex manifests might be valid but may not be fully supported by all clients, which might not yet have the capacity to render all the features³⁵⁵. Conversely, non-compliant manifests could still be partially or fully understood by some clients, which might not need to interpret every aspect of a resource.

Tripoli, an unofficial validator for the IIIF Presentation API 2.1, written in Python, provided an API for integrating IIIF validation into code. However, it has not been updated since 2017. Hyperion, a collection of NPM packages including a validator for version 3.0 of the IIIF Presentation API, has been integrated into the IIIF Commons initiative. Its dedicated validator functionality is now part of parsers and other helpers.

These tools ensure that IIIF services and resources are compliant with the standards, although some validators, such as Tripoli, are outdated. Despite these challenges, the tools provided by IIIF-C and the integrated functionalities within the IIIF Commons initiative support validation and help maintain the integrity and interoperability of IIIF resources. For validating them, I chose to leverage the official IIIF-C validators as described in the following subsection.

³⁵⁰ JSON Schema Validation of Linked Art Resources: <https://github.com/linked-art/json-validator>

³⁵¹ SHACL Linked Art Validator: <https://github.com/linked-art/shacl>

³⁵² IIIF Image API Validator: <https://iiif.io/api/image/validator/>

³⁵³ IIIF Presentation API Validator: <https://presentation-validator.iiif.io/>

³⁵⁴ Validator for the Presentation API: <https://github.com/IIIF/presentation-validator>

³⁵⁵ This is to some extent documented in the Viewer Matrix: <https://iiif.io/api/cookbook/recipe/matrix/>

8.2.2 LUX Data Consistency

This subsection discusses the data consistency of a large-scale LOUD application, focusing primarily on Linked Art in subsection 8.2.2.1 and peripherally on IIIF in subsection 8.2.2.2.

The consistency of the Linked Art data was assessed in three different ways: against the JSON validator based on the API endpoint constraints, unit-to-unit consistency between YUAG and YCBA, and by comparing concept terms from LUX with Getty's AAT. Finally, the consistency of IIIF Presentation API resources (V2.1 and/or V3.0, depending on their availability across Yale units).

8.2.2.1 Linked Art Consistency

In May 2023, I received one 24th of the full LUX dataset, i.e. a slice of records which correspond to the highest number of parallel processes possible. It is a JSON Lines (JSONL) file which is composed of lines of LUX JSON-LD that include Marklogic's rather verbose triples format. The full dataset is about 435GB uncompressed, or about 45GB compressed with `gzip`. At the time of verifying the consistency of the data, a downloadable LUX dataset or a custom dataset builder were not available.

To ensure data consistency across LUX, I conducted in October 2023 an extensive validation of Linked Art data. This process involved testing the Linked Art JSON validator and schemas with data from both the Linked Art API Endpoints and the LUX platform. The goal was to verify if the Linked Art examples and schemas were up-to-date and to validate exemplary data from LUX before a larger scale validation.

Originally, my intention was to check all the resources I received from LUX. However, after testing an initial batch, I found that they all yielded similar results for each entity endpoint. Consequently, I did not continue to check every resource. I communicated the results to Robert Sanderson, who maintains the Linked Art JSON validator, manages the Linked Art website, and also works on LUX, to ensure that the findings were accurate.

The validation process was conducted using a Python script that was forked and adapted with specific arguments from the Linked Art JSON validator³⁵⁶. This script processed data from the examples found on the Linked Art website and LUX, checking for compliance with the respective JSON schemas. Code Snippet 8.3 and Code Snippet 8.4 are validation results for a `Concept`³⁵⁷:

Code Snippet 8.3: Validating a Linked Art Resource: Example Concept

```

1 -----
2 Processing: https://linked.art/example/concept/0
3 Validation failed. Validation errors:
4 Error 1: /member_of/0 --> 'id' is a required property
5 Error 2: /created_by/influenced_by/0 --> 'id' is a required property
6 Error 3: /created_by/influenced_by/1 --> 'id' is a required property
7 Error 4: /broader/0 --> 'id' is a required property

```

Code Snippet 8.4: Validating a Linked Art Resource: Concept on LUX

```

1 -----
2 Processing: https://lux.collections.yale.edu/data/concept/000052b6-a350-4bcb-afca-15dceb72920a
3 Validated!

```

³⁵⁶ Forked Linked Art JSON Validator: <https://github.com/julsraemy/json-validator>

³⁵⁷ A comprehensive analysis is documented on GitHub at <https://github.com/julsraemy/loud-consistency/tree/main/linked-art>.

Table 8.2 summarises the validation results for different Linked Art endpoints and selected LUX instances. The table includes the status of the JSON schema validation, specific issues identified, and the required actions to achieve consistency. The main validation errors were due to missing identifiers in the Linked Art Endpoint examples and issues with the `identified_by` → `Identifier` pattern and the `referred_to_by` property in `Set` instances for the LUX data. All instances that bear the TBD* in the table had the same message warning regarding the use of `Identifier` within `identified_by`. It is a pattern that seems correct according to the documentation, informed by the Linked Art core schema³⁵⁸ which led to validation errors.

Table 8.2: Validation Summary of Linked Art Example Endpoints and LUX Instances

Endpoint	JSON Schema	LA Example	LUX Instance
Concepts	✓	Add missing IDs	✓
Digital Objects	✓	Add missing ID	TBD*
Events	Add <code>Period</code> as accepted Class when using the equivalent property	✓	✓
Groups	✓	Add missing IDs	✓
People	✓	Add missing IDs	✓
Physical Objects	✓	Add missing IDs	TBD*
Places	✓	Add missing IDs	TBD*
Provenance Activities	✓	✓	N/A
Sets	✓	✓	<code>referred_to_by</code> - LUX specific hack to deal with HTML
Textual Works	✓	Add missing IDs	TBD*
Visual Works	Add <code>shown_by</code> . Schema name (endpoint path) to be changed to <code>visual</code> ?	✓	TBD*

As part of this Linked Art consistency effort, I received two different spreadsheets from Robert Sanderson. The first contained 113,151 Getty terms from either AAT, TGN, or ULAN that are also present in LUX. The spreadsheet indicated how many times these terms appeared at both YCBA and YUAG. This data was used to check the consistency of concepts between LUX and the Getty Vocabularies, as well as the unit-to-unit consistency of both art museums. The Linked Art endpoints in question are categorised as either `concept` (AAT), `person` (ULAN and occasionally AAT), `group` (ULAN and occasionally AAT), or `place` (TGN and occasionally AAT).

For example, when comparing the term ‘goddesses’ (`aat:300343852`) between the AAT³⁵⁹ and LUX endpoints³⁶⁰, which is also present in both YCBA and YUAG digital collections, we can observe some differences. The Linked Art representation of the term on the Getty endpoint is loosely consistent with the data model but not with the Linked Art API, whereas LUX strictly follows the `concept` schema. Additionally, some SKOS properties present in the Getty endpoint, such as `skos:inScheme`, are not allowed by either the model or the API. Others, like `skos:narrower`, are permitted in the model but not in the Linked Art API. While it is understandable that Getty wants to maintain the hierarchical structure of a controlled vocabulary in both directions, Linked Art typically constrains this type of relationship in one direction. Therefore, external institutions building upon any of the Getty Vocabularies’ Linked Art representations need to do so carefully.

³⁵⁸ Linked Art Core Schema Definition: <https://linked.art/api/1.0/schema/core.json>

³⁵⁹ AAT – goddesses: <http://vocab.getty.edu/aat/300343852> / <https://vocab.getty.edu/aat/300343852.json> (Linked Art)

³⁶⁰ LUX – goddesses: <https://lux.collections.yale.edu/data/concept/6891f2d0-455f-47f9-96d6-b718c0a05fe2>

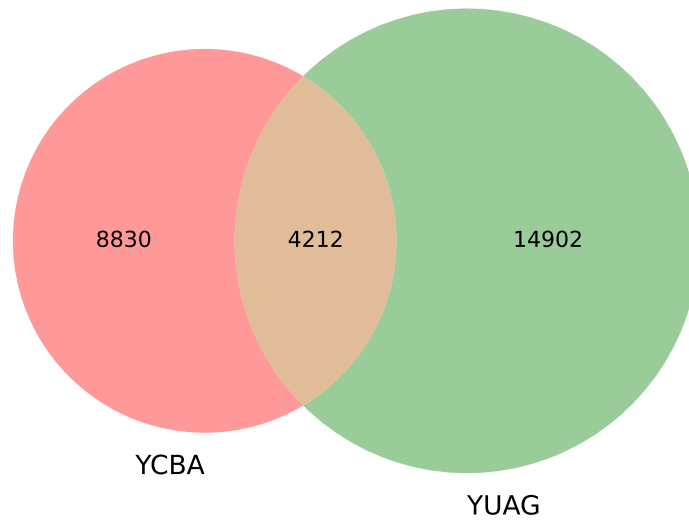


Figure 8.5: Getty Vocabularies Term Presence in the YCBA and YUAG Digital Collections

The second spreadsheet focused on the 27,944 Getty terms that are also present in both of the art museums, either as direct references or after reconciliation through the LUX pipeline. YCBA uses 13,042 concepts and YUAG uses 19,114, with 4,212 terms present in both collections³⁶¹, as illustrated in Figure 8.5.

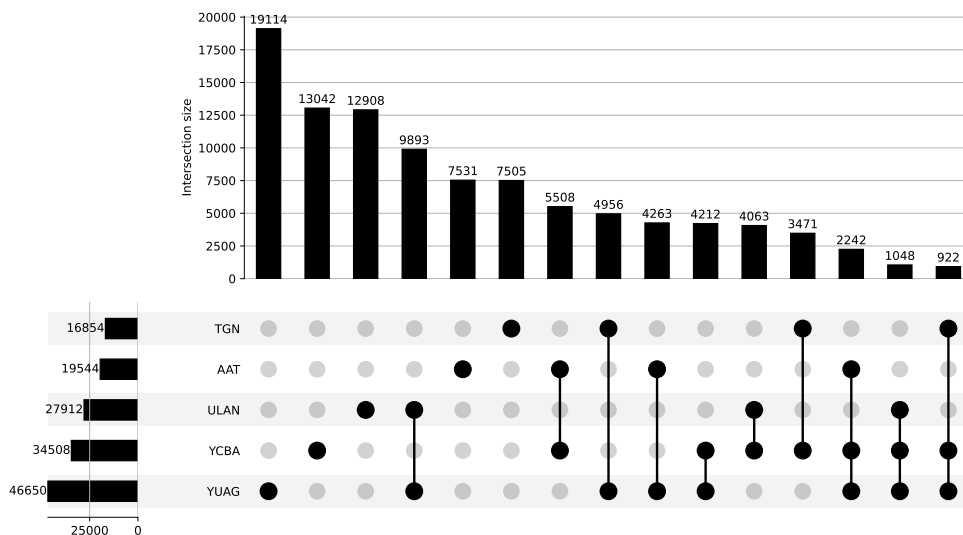


Figure 8.6: Shared Terminology in YCBA and YUAG via Getty Vocabularies

³⁶¹ A dedicated CSV with the Getty terms found in YCBA and YUAG can be found at <https://github.com/julsraemy/loud-consistency/blob/main/linked-art/ycba-yuag-vocab.csv>.

Figure 8.6 is an UpSet plot which was used to analyse the intersections between these terms. It reveals that ULAN is the most used controlled vocabulary, especially to reference people, with 12,908 terms, including 4,063 shared between YCBA and ULAN and 9,893 between YUAG and ULAN. However, AAT is the most shared controlled vocabulary, with significant overlaps: 5,508 terms shared between YCBA and AAT, and 4,263 terms between YUAG and AAT. Additionally, there are 2,242 terms shared among YCBA, YUAG, and AAT; 1,048 terms shared among YCBA, YUAG, and ULAN; and 922 terms shared among YCBA, YUAG, and TGN. In contrast, TGN is the least leveraged vocabulary, contributing 7,505 terms, with 3,471 shared between YCBA and TGN and 4,956 between YUAG and TGN. This analysis shows the predominant use of ULAN for referencing people and the extensive sharing of AAT within LUX for these particular units.

Table 8.3 and Table 8.4 illustrate how YCBA and YUAG use the Getty controlled vocabularies. The columns represent the terms directly referenced, terms added after reconciliation through the LUX platform, and the total terms that are both directly referenced and added after reconciliation. Direct reference indicates terms that these units have directly put into their systems, while reconciliation is thanks to the LUX pipeline mechanism.

Vocabulary	Direct Reference	After Reconciliation	Both
AAT	4949	417	142
ULAN	1	4062	0
TGN	4	3454	13

Table 8.3: YCBA - Usage of Getty Controlled Vocabularies

Vocabulary	Direct Reference	After Reconciliation	Both
AAT	4189	72	4
ULAN	0	9893	0
TGN	4939	12	5

Table 8.4: YUAG - Usage of Getty Controlled Vocabularies

This breakdown shows the significant role of the LUX platform in enriching the datasets of both museums. It is notable that AAT is heavily used directly by both units and TGN by YUAG, whereas ULAN is heavily incorporated through reconciliation.

Overall, the consistency of Linked Art across LUX is quite satisfactory. The main challenges were navigating the Linked Art JSON schemas, which were occasionally out of date. In addition, the Linked Art examples have been heavily modified since the initial validation and no longer have the `Identifier` issue. Despite these improvements, the findings indicate that while terms referenced by AAT are the most shared entities between YUAG and YCBA, there is probably still room for further reconciliation of other entities to enhance the cross-interoperability within LUX.

8.2.2.2 IIIF Consistency

In this subsection, I will evaluate the IIIF Presentation API resources from the LUX platform in terms of compliance with the specifications. Manifests from YUAG, YPM, and YCBA are available both as versions 2.1 and 3.0 of the Presentation API. The following URL patterns are used for either serving Version 2.1 (default for YCBA) or 3.0 (default for the two other museums):

`https://manifests.collections.yale.edu/v2/<unit>/<obj|nat>/<id>`

`https://manifests.collections.yale.edu/<unit>/<obj|nat>/<id>`

YPM's IIIF resources have `nat` in their URLs and the art museums have `obj` as a slug before the identifier. As for YUL, Manifests are available in version 3.0 at the following URL:

`https://collections.library.yale.edu/manifests/<id>`

Code Snippet 8.5 is an example command to validate a IIIF Presentation API resource against the validator installed locally. The command sends a request to the local instance of the validator, specifying the version of the IIIF Presentation API (either 2.1 or 3.0) and the URL of the IIIF resource to be validated.

Code Snippet 8.5: CURL Command Against a Local Instance of the IIIF Presentation API Validator

```
1 curl "http://localhost:8080/validate?version=<2.1|3.0>&url=<iiif-resource-url>"
```

The results are returned as JSON, as shown in Code Snippet 8.6 for a Version 2.1 Manifest and in Code Snippet 8.7 for a Version 3.0 Manifest. These results indicate whether the IIIF resource is valid (1) or invalid (0) by the value of the `okay` field. Additionally, the response includes any potential **warnings** and errors found in the resource. The way errors are handled differs slightly between the two versions.

Code Snippet 8.6: JSON Responses against the Presentation API Validator: V2.1 Resource

```
1 {
2   "okay": 1,
3   "warnings": [
4     "The remote server did not use the requested gzip transfer compression, which will slow access.
5       (Content-Encoding: )",
6     "WARNING: Setting non-standard field 'rendering' on resource of type 'sc:Canvas'\n",
7     "WARNING: Setting non-standard field 'rendering' on resource of type 'sc:Canvas'\n",
8     "WARNING: Resource type 'sc:Manifest' should have 'description' set\n",
9     "WARNING: Resource type 'dctypes:Image' should have 'format' set\n",
10    "WARNING: Resource type 'dctypes:Image' should have 'format' set\n"
11  ],
12  "error": "None",
13  "url": "https://manifests.collections.yale.edu/v2/ycba/obj/53822"
14 }
```

Code Snippet 8.7: JSON Responses against the Presentation API Validator: V3.0 Resource

```
1 {
2   "okay": 0,
3   "warnings": [],
4   "error": "",
5   "errorList": [
6     {
7       "title": "Error 1 of 2.\n Message: Additional properties are not allowed ('logo' was
8         unexpected)",
9       "detail": "",
10      "description": "",
11      "path": "/[Additional properties are not allowed ('logo' was unexpected)]",
12      "context": {
13        "@context": "http://iiif.io/api/presentation/3/context.json",
14        "id": "https://manifests.collections.yale.edu/ycba/obj/53822",
15        "type": "Manifest",
16        "label": "{ ... }",
17        "metadata": "[ ... ]",
18        "rights": "https://creativecommons.org/publicdomain/zero/1.0/",
19        "requiredStatement": "{ ... }",
20        "logo": "[ ... ]",
21        "homepage": "[ ... ]",
22        "seeAlso": "[ ... ]",
23        "items": "[ ... ]"
24      }
25    }
26  ],
27  "error": ""
28 }
```

```

26     "title": "Error 2 of 2.\n Message: 'https://creativecommons.org/publicdomain/zero/1.0/' does
        not match 'http://creativecommons.org/licenses/.*',
27     "detail": "",
28     "description": "",
29     "path": "/rights/['https://creativecommons.org/publicdomain/zero/1.0/' does not match 'http://
        creativecommons.org/licenses/.*']",
30     "context": "https://creativecommons.org/publicdomain/zero/1.0/"
31   }
32 ],
33 "url": "https://manifests.collections.yale.edu/ycba/obj/53822"
34 }

```

With the help of a forked version of the Presentation API Validator³⁶², I created a shell script to automate the validation of multiple IIIF Presentation API resources using the CLI. The script prompts the user for the unit, version, base URL, starting identifier, and the desired number of valid manifests. It then incrementally generates URLs by adding sequential identifiers to the base URL and sends validation requests to the local instance of the validator for each generated URL. This method was particularly useful as the exact range of identifiers was unknown and, besides, there were no IIIF Collections available from Yale to be parsed.

The script was designed with embedded logic to handle situations where certain identifiers might result in 404 errors, intending to skip these invalid URLs and continue crawling until the specified number of valid Manifests was attained. However, this logic did not work as intended. Instead of continuing beyond the specified range to achieve the desired count of valid manifests, the script stopped once it reached the end of the specified range, even if it had not yet processed the desired number of valid resources. As a result, some intended validations were skipped and not replaced by additional identifiers.

Table 8.5: Base Yale Manifest URLs and Respective Versions Used for Validation

Unit	V.	Base URL	Start ID
ycba	2.1	https://manifests.collections.yale.edu/v2/ycba/obj/	53800
ycba	3.0	https://manifests.collections.yale.edu/ycba/obj/	53800
yuag	2.1	https://manifests.collections.yale.edu/v2/yuag/obj/	53800
yuag	3.0	https://manifests.collections.yale.edu/yuag/obj/	53800
ypm	2.1	https://manifests.collections.yale.edu/v2/ypm/nat/	53800
ypm	3.0	https://manifests.collections.yale.edu/ypm/nat/	1629301
yul	3.0	https://collections.library.yale.edu/manifests/	16198300

The outputs are then saved into a consolidated JSONL file, maintaining a structured and easily readable format. This facilitates further analysis and reporting by providing a comprehensive and structured dataset of the validation results. Table 8.5 lists the base URL and respective version used for validating the IIIF Presentation API resources.

Using the shell script, 100 Manifests per version and per unit were crawled. Since YUL only had IIIF Presentation API 3.0, a total of 700 URLs potentially discoverable on LUX were parsed with the validator. This resulted in the following outcomes:

V2.1 Manifests from YCBA, YUAG, and YPM These Manifests all validated successfully but had the following warnings: the remote server did not use the requested gzip transfer compression, which can significantly slow down access to the resources. Gzip³⁶³ compression reduces the size

³⁶² Validator for the Presentation API (Fork): <https://github.com/julsraemy/presentation-validator/>

³⁶³ GNU Gzip: <https://www.gnu.org/software/gzip/>

of the data being transferred, thereby speeding up the download process and reducing the load on network bandwidth. Additionally, the warnings included non-standard field `rendering` on resource of type `sc:Canvas`, and resource type `dctypes:Image` should have `format set`.

V3.0 Manifests from YCBA, YUAG, and YPM These Manifests were all invalid and had minor error issues: the `logo` property was not properly structured as it should be embedded within the provider; the URL to Creative Commons should be HTTP rather than HTTPS³⁶⁴; and in some YUAG Manifests, the language map construction was not leveraged correctly, as `en` cannot be an array.

V3.0 Manifests from YUL These Manifests all validated successfully.

Figure 8.7 shows the distribution of valid Manifests, valid Manifests with warnings, invalid Manifests, and HTTP 404 responses across the different units and versions.

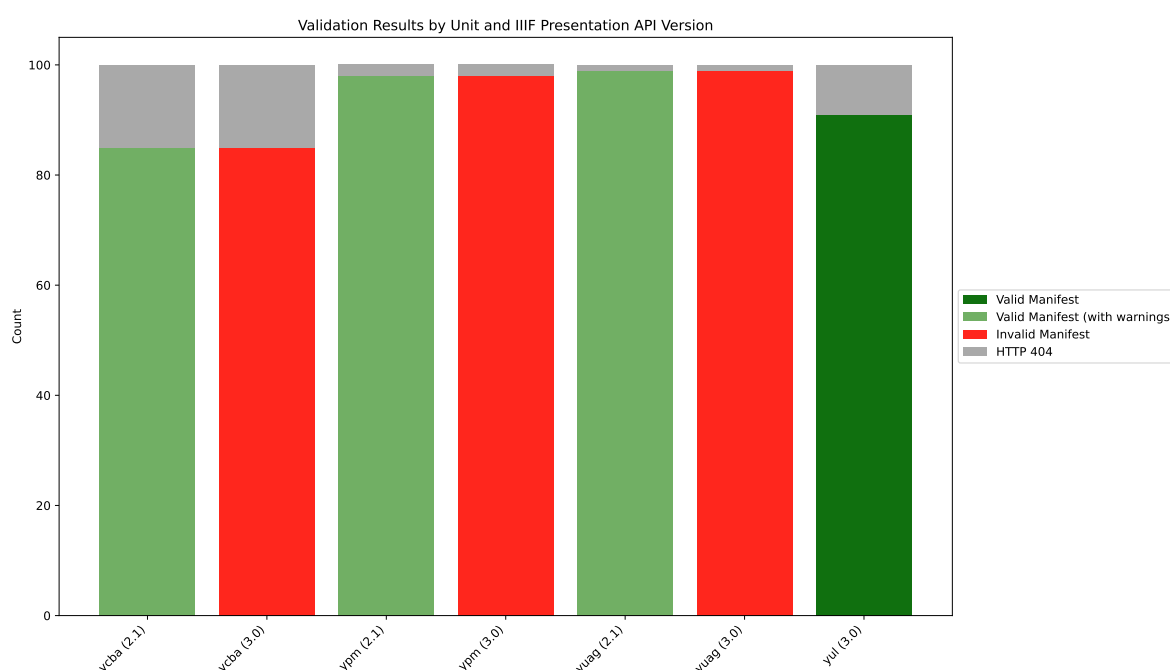


Figure 8.7: Validation Results of IIIF Presentation API Resources from Yale (Sample)

All of the errors in the Version 3.0 Manifests likely stem from conversions from Version 2.1 to 3.0. For example, the `logo` property could be at the top level in older versions but must now be embedded within the `provider` object, which is an array that includes the provider's details. The `logo` itself should be a list of image objects associated with the provider. Similarly, the language map construction logic has been updated in the more recent specification to improve internationalisation efforts. Additionally, the Presentation API specifies that the HTTP form of Creative Commons license URIs should be used for machine readability, even though HTTPS is recommended for human-readable contexts.

The following section synthesises these findings and provide further insights into the deployment of LOUD standards within LUX.

³⁶⁴ Although HTTPS URIs are listed as the 'license deed' and are recommended for general usage, the canonical form for machine-readable layers, as used in the IIIF Presentation API should be HTTP. This is because the Presentation API's enumeration of values is primarily software-driven, and thus, the HTTP form is considered the canonical form for software infrastructures. See <https://github.com/IIIF/trc/issues/32>

8.3 Synthesis and Insights

This section, which summarises the empirical research into the large-scale use of LOUD within the LUX platform, is divided into two subsections, one for each research question or objective, emphasising the collaborative aspect that emerged from the interviews and then pointing out the relative character of sheer consistency with standards.

8.3.1 Cross-Unit Collaboration

The success of the collaboration at Yale can be attributed to a shared vision that recognises the value of highlighting the connections between our diverse collections across different domains. This shared vision has enabled the resources of all participating units to contribute significantly over a number of years through active participation in committees, working groups and unit-level development efforts.

The LUX project has successfully fostered collaboration among various units at Yale, bringing together libraries and museums to work on a unified platform. The technological foundation of LUX, based on open standards, offers innovative features for data integration and organisation, facilitating cross-collections discovery. This collaborative environment has the potential to transform teaching, learning, and research, generating interest and demonstrating potential for other institutions. High-level support from leadership, particularly Susan Gibbons as Vice Provost, has been crucial in building trust and ensuring the project's success as a valuable discovery layer.

Despite the positive outcomes, several challenges and areas for improvement have been identified. One major issue is the unresolved question of ownership and decision-making for LUX, which complicates resource allocation and responsibility assignments. Data quality and consistency present another challenge, particularly with library data, and the lack of dedicated resources for data cleanup and reconciliation hinders transparency and effectiveness. Additionally, the advanced search feature of LUX requires further enhancement to improve its intuitiveness, and the separation of **Objects** and **Works** on the GUI has been confusing for some users. There is also a variation in the perception of principles such as FAIR and LOUD, with some stakeholders placing high importance on them, while others view them as secondary or are unaware of them.

The topic modelling exercise revealed actor-networks composed of organisations, individuals, and non-human actors that are intricately intertwined. This analysis provides insights into the complex web of relationships that sustain the LUX initiative. These findings resonate with the work of Ashwin Jacob Mathew, who in an ethnographic study of the internet, highlighted that it is maintained by a variety of social and technical relationships built on trust and technique [Philip 2021]. This concept, echoing the notion of 'infrastructuring' [see Pawlicka-Deger 2022], underscores the importance of ongoing processes and relationships in maintaining and evolving infrastructure, similar to the collaborative efforts seen in LUX.

Ultimately, the lessons learned from cross-unit collaboration at Yale underscore the importance of a shared vision and strong leadership support in driving successful projects. While challenges remain, LUX's collaborative model offers a promising framework for future initiatives within and beyond the confines of this university.

8.3.2 As Consistent as Required

Ensuring full consistency in data across different institutions and platforms is challenging. The balance lies between achieving semantic consistency that can be reconciled or matched across diverse

datasets and maintaining sound UX that remains intuitive and functional across various GUIs, mostly embodied by compliant viewers or players, from one institution to another.

Adherence to LOUD standards and schemas within LUX has generally been positive. However, specific issues arise, particularly when transitioning between versions of a specification. For instance, the `logo` property, which could be appended at the top level in Version 2.1 of the IIIF Presentation API, must now be embedded within the `provider` object in Version 3.0. This transition can be problematic for validators, as they might flag such properties as errors. Nevertheless, clients like Mirador can still display the logo correctly upon parsing the Manifest. This makes it currently acceptable but client-dependent. Maintaining strict consistency with the API is crucial for large institutions that could serve as examples for others, particularly those that are IIIF-C members.

For Linked Art representations in LUX, achieving consistency within the API entity endpoints is fundamental to ensuring interoperability beyond the portal, positioning it as a flagship for Linked Art implementation. For IIIF, as long as key clients such as UV and Mirador can accurately interpret the data patterns, the focus on strict consistency can be loosened to some extent. However, maintaining a high level of consistency remains a prerequisite for success, ensuring that data can be seamlessly integrated and used across different platforms and institutions.

In conclusion, while LUX's adherence to LOUD standards and schemas is commendable, the initiative must continue to address specific consistency issues to enhance (semantic) interoperability. The following section will explore the broader perspectives of the LUX initiative and its implications for the CH field, highlighting the lessons learned and potential future developments.

8.4 Perspectives

This chapter has explored the complexities and achievements of the LUX initiative at Yale, highlighting the importance of cross-unit collaboration, technical innovations, and the adherence to standards. Looking forward, the LUX platform continues to evolve, offering valuable lessons for other institutions in the CH field.

The LUX platform is dedicated to a multi-faceted approach for its development and expansion. Actively collecting user feedback through a form available on each page, the platform leverages this input to continually refine its services and features. Additionally, LUX will broaden its scope by incorporating more collections, including those from renowned institutions like the Paul Mellon Centre in London³⁶⁵. Furthermore, their commitment to updating descriptive metadata aims to enhance the accuracy of collection descriptions and reduce any inherent biases. These efforts are guided by the Bias Awareness and Responsibility Committee, ensuring a more equitable representation of cultural heritage [Metcalf Hurst 2023, p. 4].

The integration of external data sources, such as Wikimedia statements for public domain images, introduces new dimensions and challenges to the LUX platform. While enriching the dataset, it also poses issues where queries may return unrelated results due to these external statements. For instance, searching for 'lightning talk' might return results related to 'Bird of paradise'. This highlights the need for ongoing refinement in query handling and data integration processes to maintain relevance and accuracy in search results.

³⁶⁵ Paul Mellon Centre: <https://www.paul-mellon-centre.ac.uk/>

One of the key insights from the LUX initiative is that a Linked Data-driven infrastructure does not necessarily necessitate SPARQL for its implementation. The use of community-driven JSON-LD-based specifications has enabled the platform to maintain interoperability and data consistency without the complexity of SPARQL queries. This approach simplifies the adoption and integration process for institutions with varying levels of technical expertise, promoting wider usage and collaboration. Additionally, by leveraging MarkLogic, a multi-modal database, LUX benefits from robust data management capabilities [Sanderson 2024a]. Although MarkLogic is a paid service, all the processes related to its use – such as API usage, build tools and tasks, data constants, database indexing, or performance testing procedure – along with the data pipeline, have been made open source as part of the LUX project. This transparency further encourages collaboration and adoption by other institutions.

The experiences and lessons learnt from LUX provide valuable insights for other institutions looking to implement similar projects. The success of the LUX platform underscores the importance of high-level support, interdisciplinary collaboration, and a commitment to standards compliance. Other institutions can draw from these experiences to enhance their own digital collections and discovery platforms. That said, the requirements of the API ecosystem, such as the implementation of HAL and the IIIF Change Discovery API for deploying Linked Art, as LUX has done, are quite significant. Consolidation work within the Linked Art community is needed to help support smaller institutions in adopting this ecosystem. For LUX, the ecosystem they have built will need to be updated once Linked Art comes to an agreement on how to have sound naming conventions using HAL for back-links, which is necessary to drive the GUI affordances as well.

In reflecting on the broader implications of the LUX initiative, it becomes clear that innovation within the digital space is a multifaceted concept that extends beyond technological advancements. It encompasses innovative approaches to knowledge building and production methods, often influenced by interdisciplinary fields. Nell Smith [2014] highlights the need to question what counts as innovation, particularly the contributions of critical and feminist studies to DH.

‘Innovation continues to be a hot term, especially in the digital world. What counts as innovation should persistently be questioned, especially since technological innovation has repeatedly overshadowed innovations in knowledge building, including methods of knowledge production. These suggestions all come from methods generated by feminist criticism and theory, critical race studies, sexuality studies and queer theory, and class studies and have advanced and otherwise improved my own work, and, as is surely obvious, can improve the work of digital humanities, scholarly editing, computer science, information studies, library science and humanities computing. The frozen social relations of old orders can and should be thawed in order to enable sociological innovations, which should be key and self-consciously incorporated into the production of any postcolonial digital archive.’ [Nell Smith 2014]

Large-scale Linked Data projects such as LUX need to be enabled by the adoption of sound socio-technical principles. These principles ensure that technological innovations are balanced with social and institutional considerations, fostering an environment where both data and human (f)actors are aligned towards common goals. Interdisciplinary committees, such as those focused on metadata or bias awareness, play a crucial role in achieving mutual agreement and, most importantly, in recognising the diverse perspectives and contributions of all current and past contributors.

As LUX moves forward, it will continue to focus on user feedback, integrating new collections, and updating metadata to improve accuracy and representation. These efforts will ensure that LUX remains a flagship project for the integration and discovery of CH data leveraging LOUD specifications and their design principles, setting a benchmark for other institutions in the field.

9

Discussion

‘[Il] faut renoncer à l’idée d’une interopérabilité syntaxique ou structurelle par l’utilisation d’un modèle unique, qu’il s’agisse de la production, de stockage ou de l’exploitation au sein même d’un [système d’information].’ [Poupeau 2018]³⁶⁶

This chapter presents a comprehensive discussion where I interpret, analyse and critically examine my findings in relation to the thesis and the wider application of LOUD. Through an in-depth analysis of the design principles of LOUD and their implications for CH, this discussion aims to demonstrate the many challenges and opportunities inherent in this framework. The focus is on achieving community-driven consensus, rather than simply pursuing technological breakthrough.

The following sections are organised to provide a comprehensive review of the empirical findings, an evaluation abstracting LOUD, and a retrospective analysis of the research journey. Firstly, in Section 9.1, I will present a summary of the empirical findings from my research. This will include key themes and insights, structured to reflect the different areas of study and practice within LOUD.

Secondly, in Section 9.2 I will provide an evaluation of LOUD by means of using the LoA approach. This evaluation will focus on the impact of LOUD on the perception of Linked Data within the CH domain and the wider DH field. This will include the key themes and insights that have emerged, structured in a way that reflects four levels of abstraction. I will also explore the dual nature of LOUD implementation, involving both simplicity and complexity, and discuss the various factors that influence such dynamics.

Finally, in Section 9.3, I will offer a retrospective analysis of the research journey. This section will interpret the findings to situate LOUD as fully-fledged actors. It will reflect on the challenges, achievements, and lessons learned throughout the research process, providing a holistic view of the project’s trajectory and its implications for the future of LOUD.

9.1 Empirical Findings

This section summarises the empirical findings of my research and already offers some suggestions. The structure does not follow the exact order of the three empirical chapters but is organised around overarching topics that emerged throughout the study. The seven topics include Community Practices and Standards, Inclusion and Marginalised Groups, Maintenance and Community Engagement, Interoperability and Usability, Future Directions and Sustainability, Digital Materiality and Representation, as well as Challenges of Scaling and Implementation.

³⁶⁶ Author’s translation: ‘We need to give up on the idea of syntactic or structural interoperability through the use of a single model, whether for producing, storing or managing data within an information system.’

Community Practices and Standards GitHub serves as a vital hub for community involvement, with a core group of active contributors often attending meetings regularly. This platform simplifies decision-making within the community, although it also reflects biases similar to those in FLOSS communities. Behind visible activities like meetings, there is substantial preparatory work managed by co-chairs, editorial boards, or driven by community-generated use cases. This foundational work often determines the direction and outcomes of formal gatherings. The LUX project at Yale, as seen in Chapter 8, has successfully fostered collaboration across various units, bringing together libraries and museums on a unified platform. The technological foundation of LUX, based on open standards, facilitates data integration and cross-collections discovery.

Not only does the deployment of FLOSS tools contribute to these achievements, but it also emphasises the social advantages of working collaboratively. The concept of the Tragedy of the Commons, as described by [Hardin \[1968\]](#), highlights the potential for individual self-interest to deplete shared resources. However, [Ostrom \[1990\]](#) offers a counterpoint by demonstrating how communities can successfully manage common resources through collective action and shared norms. In this context, initiatives like the Community Health Analytics in Open Source Software (CHAOSS) initiative³⁶⁷ play a significant role by providing metrics that help evaluate the health and sustainability of open source communities. These metrics include contributions, issue resolution times, and community growth, offering valuable insights into how collaborative efforts can be maintained and improved.

Reaching consensus is another critical aspect of community practices and standards. While the minutes of meetings are valuable artefacts, they often reflect an Anglo-Saxon approach to decision-making characterised by few substantive points and critical turning points. The formal aspects of conversations captured in minutes do not fully encompass the decision-making process, which frequently involves informal conversations, consensus-building through open dialogue, and subtle cues that influence outcomes. These elements are integral to the English and American approach and hold valuable lessons for an international community. IIF and Linked Art are international communities, but decisions are made in English and the majority of participants are based in North America and the UK, significantly imprinting this approach. Understanding these nuances can help us improve our collaborative efforts within the IIF and Linked Art communities. By recognising and appreciating these different facets of decision-making, we can learn from each other and enhance our collective ability to make effective and inclusive decisions.

Some of the challenges associated with these practices include the major demand on resources for community building, the slowness inherent in distributed development, and the difficulty in achieving consensus. Additionally, the concept of social sustainability can be seen as an imaginary construct that papers over differences, as discussed by [Fitzpatrick \[2019\]](#). Addressing these challenges is crucial for the long-term success and effectiveness of the IIF and Linked Art communities.

Inclusion and Marginalised Groups The demographic homogeneity in these communities can perpetuate biases and neglect issues relevant to underrepresented or marginalised groups, as seen in Chapter 6. Participation in these standardisation processes is itself a privilege. The assumption that internet access and digital devices are universally available is critically examined, revealing key actors in the digital landscape. This mirrors issues within the IIF community, where generating IIF resources presupposes means that may not be accessible to all.

We need clear terms of inclusion, as highlighted by [Hoffmann \[2021\]](#). She argues that effective inclusion requires a critical examination of the frameworks and conditions under which inclusion

³⁶⁷ CHAOSS: <https://chaoss.community/>

is offered. The framework should ensure that inclusion initiatives do not merely add diversity to existing power structures but work to transform these structures fundamentally. This involves questioning who defines the terms of inclusion, who benefits from them, and who may be inadvertently excluded. Hoffmann [2021] suggests a participatory approach, where marginalised communities are actively involved in shaping inclusion policies and practices, thus making inclusion an ongoing, reflective process rather than a static goal.

The inclusion of marginalised groups is a necessary step, but it is not sufficient. To truly make a difference, there must be a strategic and concentrated effort to appropriate technologies, as emphasised by Morales [2009, 2017, 2018] and further articulated by Martinez Demarco [2019, 2023]. This strategic approach highlights the political significance of challenging dominant neoliberal and consumerist perspectives on technology and individual engagement.

Martinez Demarco [2023] underscores the critical importance of focusing on practices that go beyond mere inclusion. Instead, it requires a deep understanding and critical assessment of how technology is intertwined with social, economic, and ideological contexts. It implies a reflective and deliberate process of technology adoption in which individuals creatively tailor technology to their specific needs, beliefs, and interests. Moreover, a key aspect highlighted by Martinez Demarco [2023] is the implicit and explicit critique of a universalist approach to inclusion, which often lends itself to all too easy instrumentalisation. Understanding and studying resistance to inclusion in an oppressive digital transformation context is paramount, particularly given the highly unequal conditions that prevail.

In this light, a comprehensive study of socio-material and symbolic processes, practices, and ‘imaginaries’ involved in embedding technologies into individuals’ lives is needed. This approach also recognises technology as a catalyst for change. It envisions the use of technology to drive meaningful change at multiple dimensions and realities—national, societal, or personal. By focusing on these practices, empowering individuals to navigate and use technology thoughtfully and purposefully becomes a reality, bridging the gap between technological advances and societal progress [Martinez Demarco 2019].

Maintenance and Community Engagement The tension between creating advanced specifications and their practical implementation by platforms is evident in the IIF Cookbook recipes and Linked Art patterns, as discussed in Chapter 6. This ongoing development shows that the community is still finding the best ways to achieve broad adoption and interoperability. The deployment of the Change Discovery API, as illustrated in Chapter 7, demonstrates that establishing such a protocol on top of the IIF Presentation API is feasible and straightforward. High-level support from leadership, particularly Susan Gibbons as Vice Provost, has been crucial in building trust and ensuring the project’s success as a valuable discovery layer at Yale. This integration of diverse collections through a unified platform, based on open standards, highlights the potential for transforming teaching, learning, and research by leveraging collaborative efforts. The topic modelling exercise in LUX reveals the intricate actor-networks composed of organisations, individuals, and non-human actors. This analysis underscores the importance of ongoing processes and relationships in maintaining and evolving infrastructure, akin to the concept of ‘infrastructuring’. As detailed in Chapter 8, following best practices and guidelines such as the SHARED Principles is essential for better involvement, but it is also crucial to uphold these commitments consistently over the long term to ensure meaningful participation. Between the PIA team members, there were sometimes ‘*disconnects between different communities who undertake collaborative research*’ [Vienni-Baptista et al. 2023]. This was something we had to navigate and learn from, which was manageable within the context of a laboratory setting. However, for any follow-up projects or whatever forms the digital infrastructure we built may take, it is imperative that these disconnects are addressed and solidified to ensure cohesive and sustained community engagement.

Interoperability and Usability Within PIA, different APIs have been progressively deployed to meet various requirements while allowing parallel exploration of data modelling. Each API offers unique advantages, but their collective integration promotes semantic interoperability. For example, the IIF Image API has been instrumental in rationalising image distribution across prototypes, providing efficient access to high-quality digital surrogates and the ability to resize them for different uses. Adherence to LOUD standards and schemas within LUX has generally been positive, although transitioning between versions of a specification can present challenges, highlighting the need to improve the consistency of compliant resources.

Linked Art, for instance, has the capacity to generate various insights and sources of truth around different entities. However, additional or entirely new vocabularies from sources like the Getty may need to be used – such as Homosaurus. Complementary to Linked Art, using WADM allows for assertions that go beyond purely descriptive narratives, though it may sacrifice some semantic richness. This complexity in managing vocabularies and maintaining semantic richness directly ties into broader usability considerations within the community.

Addressing these usability concerns, Robert Sanderson has suggested focusing on the use of full URIs in Linked Art to ensure computational usability, in contrast to IIF’s approach of minimising URIs to enhance readability. This difference highlights a fundamental question in usability: balancing readability and computational usability. Understanding developers’ perspectives on these approaches is critical.

I would suggest as a way forward for the IIF and Linked Art communities to focus on further improving usability of the specifications. This includes conducting comprehensive usability assessments of APIs to evaluate the experiences of new developers versus existing ones, understanding the steepness of the learning curve associated with each API, and guiding improvements in documentation, on-boarding processes, and overall developer support. Efforts should be made to lower the barriers to entry for new developers by developing more intuitive and user-friendly tutorials, providing example projects, and creating a robust support community. Ensuring that developers can quickly and effectively leverage APIs will foster greater adoption. Addressing the challenges of transitioning between different versions of specifications is critical, and developing tools and guidelines that help maintain consistency across versions will reduce friction and ensure smoother updates.

Future Directions and Sustainability Survey findings, as discussed in Chapter 6, underscore the need for ongoing efforts to develop LOUD standards that foster an inclusive, dynamic digital ecosystem. Future strategies should include creating educational resources and frameworks that support interdisciplinary collaboration and reduce barriers to participation. While the Manifest serves as the fundamental unit within IIF, the Linked Art protocol can play a similar central role as semantic gateways in broader contexts, allowing round-tripping across the APIs. The topic modelling exercise in LUX, detailed in Chapter 8, reveals complex actor-networks of organisations, individuals, and non-human actors, providing insights into the relationships sustaining the LUX initiative.

The next steps for Linked Art might involve forming a new consortium independent of a CIDOC Working Group, which could provide the necessary support to sustain the initiative. Alternatively, integrating Linked Art into IIF as a new TSG and specification could address the discovery challenges within IIF, as discussed during the ‘Connecting IIF and Semantic Cultural Heritage Metadata for Discovery’ birds of a feather session led by Robert Sanderson [see Raemy 2024b] at the 2024 IIF Conference in Los Angeles³⁶⁸. Design principles that act as bridges across different

³⁶⁸ IIF Annual Conference and Showcase - Los Angeles, CA, USA - June 4-7, 2024: <https://iif.io/event/2024/los-angeles/>

disciplines, as proposed by [Roke and Tillman \[2022\]](#), are crucial. IIF has demonstrated that this collaborative approach is feasible, and Linked Art could follow in its footsteps. However, achieving this requires increased dedication from passive members and broader adoption of the model and the API ecosystem in the near future.

Digital Materiality and Representation As explored in Chapter 7, the detailed digital representation of photographic albums, such as the Kreis Family Collection, demonstrates the need to comprehensively capture the materiality of digital objects. This includes the structure and context of images, which are crucial for maintaining their historical and social significance. The implementation of the IIF Presentation API in creating a detailed digital replica of the Getty's Bayard Album shows how digital materiality can be enhanced through thoughtful use of technology, but also highlights the scalability challenges for such detailed representations.

Creating these detailed digital representations can be seen as a 'boutique' approach, which, while labour-intensive and resource-demanding, is necessary for preserving the integrity and contextual significance of cultural heritage objects. The challenge lies in developing the appropriate means and methodologies to achieve this level of detail consistently. Future endeavours, whether through research projects or collaborative efforts between GLAM institutions and DH practitioners, should aim to address these challenges and create sustainable practices for digital materiality and representation. As Edwards aptly notes:

'Presentational forms equally reflect specific intent in the use and value of the photographs they embed, to the extent that the objects that embed photographs are in many cases meaningless without their photographs; for instance, empty frames or albums. These objects are only invigorated when they are again in conjunction with the images with which they have a symbiotic relationship, for display functions not only make the thing itself visible but make it more visible in certain ways.' [[Edwards and Hart 2004](#), p. 11]

Challenges of Scaling and Implementation As seen in Chapter 6, the IIF Cookbook recipes and Linked Art patterns reflect the tension between creating advanced specifications and their practical implementation. This gap between ideation and real-world application underscores the challenges faced by the community in achieving broad adoption and interoperability. In Chapter 7, the exploration of APIs like the IIF Change Discovery API illustrates the practical challenges and potential of scaling these technologies for wider adoption. The successful implementation in PIA demonstrates viability, but also points to the need for continued development and community engagement to fully realise the benefits.

Furthermore, assessing the scalability of IIF image servers, as discussed by [[Duin 2022](#)] and exemplified by the firm Q42 with their Edge-based service Micrio³⁶⁹, highlights the importance of optimising data performance. Erwin Verbruggen aptly noted that 'optimizing data performance in my opinion means sending as little data over as needed'³⁷⁰, emphasising the need for efficient data handling to enhance scalability. This insight reinforces the necessity of continual refinement in scaling digital infrastructure to support broader use and integration.

Reflecting on these findings, I would like to assert that continuous participation, particularly for institutions that can afford to be part of initiatives like IIF-C, is essential. Active members should not only focus on their own use cases but also consider the needs and perspectives of other, perhaps marginalised, groups. Achieving the dual goals of making progress within one community, whether it be

³⁶⁹ Micrio: <https://micr.io/>

³⁷⁰ Message written on the IIF Slack Workspace on 28 October 2022.

IIIF or Linked Art, while also engaging in effective outreach and creating a solid baseline, will benefit everyone in the CH sector and beyond. Addressing where LOUD fits in, how people perceive this new concept or paradigm, and understanding how LOUD differs from Linked Data in general are essential. These questions help to clarify the stages at which themes related to one of the LOUD design principles emerge, crystallise, and potentially disappear. My thesis does not fully resolve these queries but offers insights and hints for further exploration.

In conclusion, the empirical findings reveal the richness of the implementation and maintenance of LOUD standards in the CH domain. From the critical role of community practices and standards to the challenges of achieving interoperability and inclusivity, each theme underlines the complex interplay of social, technical and organisational factors. Section 9.2 will look at the evaluation of LOUD and explore its overall impact, delving into the delta of what to do with it, particularly in terms of Linked Data versus LOUD, where my thesis provides pointers but does not provide definitive answers.

9.2 Evaluation: Abstracting LOUD

In this section I will assess the impact of LOUD within the CH domain and the wider DH field, examining its implications for community practices and semantic interoperability, and secondarily whether LOUD has affected the perception of Linked Data.

Referring to Figure 4.2, the following is a descriptive attempt to provide a LoA of LOUD based on my empirical findings, focusing particularly on the deployment of IIIF within PIA and Linked Art within the LUX framework, aside from the data model abstraction level.

Representation and Display For PIA, the implementation of Leaflet provided an immediate and easy-to-integrate viewer to display high-resolution digitised images of CAS photographs. The context is accessible through accompanying metadata and related links on the GUI. Although not LOUD-driven *per se*, it functions as a mediator through the IIIF Image API. Balancing between immediacy and hypermediacy, the Mirador instance enabled the display of IIIF Presentation API resources with machine-generated annotations. We also incorporated the V3 plug-in to manipulate images³⁷¹. However, we failed to provide a robust authentication layer allowing users to add their own annotations easily, highlighting the limitations of a four-year research project not primarily aimed at tool development but at proposing a participatory system. IIIF-compliant software can aid in this, yet development needs to be community-driven rather than individualised. Exhibit was the only tool used for educational and teaching purposes that was well-received, though integration issues persisted.

LUX exemplifies Linked Art hypermediacy, where the structure of the JSON-LD representation drives their GUI, including URL syntax³⁷².

For both PIA and LUX, JSON-LD serves as an interface for certain users (software developers, data curators, data scientists). While resources can become BOs depending on the viewer, a few inconsistencies can still be overcome and will likely be understood by humans reading the files.

Data Model The data model of IIIF is primarily driven by its design principles and WADM. Also the main unit is the Manifest, often a digitisation or representation of a physical object, meaning the Presentation API is key to achieve an acceptable level of interoperability. The Shared Canvas Data Model is still somewhere, baked into the specifications, but one does not really need to know

³⁷¹ mirador-image-tools: <https://github.com/ProjectMirador/mirador-image-tools>

³⁷² For instance, this user interface view of Claude Monet (1840-1926): <https://lux.collections.yale.edu/view/person/642a0152-1567-4f8e-93f3-66f11c5cab9a> and its Linked Art counterpart: <https://lux.collections.yale.edu/data/person/642a0152-1567-4f8e-93f3-66f11c5cab9a>

about this model to understand how IIIF works from Version 2 of the main APIs onwards, it is a piece of history though. IIIF is Linked Data, but has no real semantics value and should really not be treated as RDF triples.

One could almost say the same about Linked Art, as it is not necessary to fully understand CIDOC-CRM to either start using its model or to deploy an API endpoint. However, some basics do need to be understood, such as the event-based model viewpoint, the important classes and their `rdfs:domain` and `rdfs:range`. However, Linked Art has bent some rules and created some properties and classes to meet the needs of the community. As far as implementation goes, I would suggest to directly implement and be consistent with the Linked Art API endpoints, rather than starting with the data model, as I see cross-institutional interoperability through interfaces as a more important milestone than data modelling as a pastime for the few specialists.

For both PIA and LUX, dedicated data models were done to be consistent with the specifications, with some internal structure and data in LUX for their own purposes. Semantic data in PIA, as already realised through the Omeka A JSON-LD API was not done beyond templating of a few Linked Art resources and the workflow done with the University of Oxford, but there is not a productive PIA Linked Art API at the moment.

Infrastructure Serialisation mock-ups and JSON-LD templates on GitHub were the starting point to model IIIF Manifests and Collections for the PIA research project. Laravel and then Omeka A were the two main elements, in two different iterations, that were leveraged to present the IIIF resources. If single image Manifests were quite easy to serialised, the integration of more detailed representations of photographic albums presented challenges. A more robust infrastructure is definitely needed for the long-run, but efficient enough in a laboratory setting.

Algorithmic and Computational Processes PIA relied on virtual machines and had the necessary Kakadu licence embedded in our SIPI instance to encode the images. If the former proved difficult as performance was sometimes an issue, the latter was a good option as serving JPEG2000 images cannot currently rely on FLOSS solutions which are too slow.

The LUX pipeline and the use of MarkLogic as a multi-modal database are examples of the data engineering expertise and outsourcing solutions required for such a platform. Some open source solutions, such as QLever³⁷³, a high-performance SPARQL engine, may also offer some hope to institutions that are not well-funded and need robust knowledge graph-oriented solutions.

The dual simplicity and complexity of implementing LOUD specifications and participating in community-led efforts can be attributed to the need for a reorientation of research projects. It is essential for these projects to actively engage in community processes rather than intermittently presenting their progress and subsequently withdrawing. This ongoing engagement fosters a more robust and collaborative environment, ultimately contributing to the advancement of shared goals and standards. Such a reorientation necessitates a fundamental change in how universities and GLAMs institutions operate, extending their involvement beyond the immediate project scope to ensure sustained participation and impact.

Despite the introduction of LOUD, the perception of Linked Data has not evolved significantly. Most software engineers continue to treat resources primarily as JSON, often overlooking the graph structure that underpins Linked Data. For IIIF, this approach is appropriate given its focus on content interoperability and presentation. However, for Linked Art, overlooking the graph structure could be

³⁷³ QLever: <https://github.com/ad-freiburg/qllever>

problematic to some extent, as it limits the full realisation of the semantic relationships and rich interconnections that Linked Data can provide. This highlights the need for more focused efforts to integrate semantic web principles, particularly in contexts where these principles can significantly improve the quality of data.

I have faced challenges in moving many of the models developed within PIA into (beta) production, and the usability requirements of APIs have scarcely been addressed. However, the findings from this thesis should be viewed as starting points rather than conclusive solutions. The unseen aspect of this dissertation is my active involvement in both communities and my attempts to reciprocate this engagement within PIA. Each investigation presented could have warranted a dedicated thesis, indicating the breadth and depth of the topics explored. Ultimately, this work merely scratches the surface of numerous subjects, laying the groundwork for future research and development.

The next section will offer a retrospective on the work accomplished during this PhD thesis. It will reflect on the various milestones achieved, the lessons learned, and the potential directions for future research.

9.3 Retrospective: Trudging like an Ant

In this retrospective³⁷⁴, I will offer an analysis of the research journey. This section will interpret the findings to situate LOUD as fully-fledged actors within the CH field. It will reflect on the challenges, achievements, and lessons learned throughout the research process, providing a holistic view of the project's trajectory and its implications for the future of LOUD.

The empirical findings of my research reveal the nuanced interplay between socio-technical practices and implementations, synthesising insights through both thematic and abstract lenses. This dual approach underscores the importance of fostering collaboration and effective decision-making, while addressing biases and promoting inclusivity. The need for ongoing maintenance, interoperability and usability remains paramount, as does the development of educational resources and consortia to sustain initiatives. In addition, capturing digital materiality and addressing scalability challenges are critical to the widespread integration of LOUD standards. These findings lay the groundwork for future research and development aimed at bridging operational applications with more extensive design approaches.

How can LOUD be situated as fully-fledged actors within the CH field? Reflecting on the notion of 'graceful degrading', frequently mentioned during the 2024 IIF Conference, LOUD specifications embody this concept perfectly. Even if not all embedded patterns of a given API-compliant resource are correctly interpreted or rendered by a client, some of its basic features should still be displayed. This flexibility is crucial for ensuring the broad usability and adaptability of LOUD standards, allowing them to transcend institutional boundaries and serve as robust mediums of knowledge transfer. To paraphrase [Poupeau \[2018\]](#)'s quote at the beginning of this chapter, there isn't a unique model for interoperability, but there are definitely best sociotechnical practices to be learned from IIF and Linked Art. The act of participation prevails over the relatively easy and one-off deployment of specifications for the short term.

By using LOUD, CH data can be effectively interlinked with different datasets, resulting in numerous potential benefits. An overriding benefit is the improved discoverability and accessibility of CH resources, facilitating enhanced search and retrieval capabilities. In addition, the adoption of LOUD

³⁷⁴ The title of the section is an homage to Bruno Latour and a passage found in his book 'We Have Never Been Modern'.

promotes seamless data sharing and reuse within academic and memory institutions, fostering a culture of collaboration and interdisciplinary knowledge exchange. This approach not only enhances the overall utility and comprehensiveness of CH repositories, but also promotes collective understanding and appreciation of diverse cultural assets and historical narratives.

However, it is essential to critically evaluate the application of LOUD in the context of CH data. While LOUD offers promising prospects for improved data interlinking and accessibility, challenges and concerns persist. The transition to LOUD principles necessitates significant investments in resources, including infrastructure, expertise, and time, which may pose barriers for smaller institutions or those with limited funding. Moreover, ensuring the accuracy, consistency, and quality of Linked Data is a complex task, demanding meticulous attention to detail and ongoing maintenance efforts. Furthermore, potential issues related to data ownership, rights management, and the potential misuse or misinterpretation of interconnected data should be carefully considered. Standardisation across different CH domains, each with unique data structures, formats, and contexts, may present formidable obstacles to seamless integration. These concerns underscore the need for a nuanced and cautious approach to the implementation of LOUD standards, taking into account the complexity and specificity of CH data and its diverse custodians.

This thesis has been a journey of discovering Linked Art and a confirmation that the ethos of IIF is yet to be fully manifested beyond product implementation. The sense of belonging to a community is an ongoing endeavour, much like the ants in Latour's metaphor. This dissertation underscores that active participation in community processes is essential to achieving the dual goals of advancing the technological framework for semantic interoperability and fostering an inclusive and collaborative CH ecosystem.

10

Conclusion

*'For a better understanding of the past,
Our images have to be enhanced,
A new dialogue in three dimensions,
Must have openness at its heart,
For somewhere within the archive
Of our aggregated minds
Are a multitude of questions
And a multitude of answers,
Simply awaiting to be found.'*
[\[Mr Gee 2023\]](#)

This chapter brings to a close the journey undertaken since February 2021, aiming to clearly articulate the answers to the research questions, discuss how the research aligns with the objectives, elucidate the significance of the work, outline its shortcomings, and suggest avenues for future research.

I had the privilege of hearing the above poem at EuropeanaTech in The Hague in October 2023. What struck me most, and what I have tried to convey in this thesis, was the powerful dialogue and collective spirit striving to harness the potential of our (digital) heritage. With a sense of conviction after this conference, I approached the next one in Geneva in February 2024 with confidence, believing that I had made a compelling case for the concept of LOUD. When a participant asked how LOUD differed from Linked Data, however, I found myself explaining the socio-technical ethos of IIF and Linked Art, the richness of the individuals who make them up, the ability to combine these different standards, and the common use cases that emerge from these collaborations. Whether my answer was convincing remains uncertain, but I knew it was too brief. Perhaps it is here, in this conclusion, that my thoughts can find their full expression.

I believe that LOUD should be at the forefront of efforts to improve the accessibility and usability of CH data, an endeavour that is increasingly relevant in a web-centric environment. This paradigm has gained considerable traction, particularly with the advent of Linked Art and the recognition that the IIF Presentation API has been an inspiration for the LOUD design principles. The development and maintenance of LOUD standards by dedicated communities are characterised by collaboration, consensus building, and transparency. In the interstices of the IIF and Linked Art communities, frameworks for interoperability are not only exposed, but revealed as profound testaments to the power of transparent collaboration across institutional boundaries. Both communities, it is true, are still very much Anglo-Saxon efforts, where the specifications have mainly been implemented in GLAM and/or DH research projects, or at least when we have been aware of them. It has clear guidelines on how to propose use cases, mostly using GitHub, and hides the sometimes unnecessary RDF complexity behind a set of JSON-LD `@context`. IIF is at the presentation layer and can really play its role as a mediator,

with the Manifest as its central unit connecting other specifications, including semantic metadata, and preferably with simpatico specifications such as Linked Art.

An important hypothesis arises from the observation that adherence to the LOUD design principles makes specifications more likely to be adopted. The primary benefit of adopting LOUD standards lies in their grassroots nature. This grassroots approach not only aligns with the core values of openness and collaboration within the DH community but also serves as a common denominator between DH practitioners and CHIs. This unique alignment fosters a sense of shared purpose and common ground. However, it's essential to acknowledge that while LOUD and its associated standards, including IIF, hold immense promise, their limited recognition in the wider socio-technical ecosystem may currently hinder their full potential impact beyond the CH domain.

Consideration of socio-technical requirements and the promotion of digital equity are essential to the development of specifications in line with the LOUD design principles. In the context of the IIF and Linked Art communities, this means both recognising current challenges and building on existing practices. This includes forming alliances that support diverse forms of inclusion at both project and individual levels. For example, organisations should be encouraged to send representatives from diverse professional and personal backgrounds, such as underrepresented groups or non-technical fields. This can be facilitated by initiatives that lower the barriers to participation, such as financial support for travel and participation, flexible participation formats, and targeted outreach efforts. Furthermore, as these standards often align with open government data initiatives, they present opportunities for broader public engagement and institutional transparency.

In the broader context of DH, understanding LOUD involves tracing the historical development of the field and its evolving relationship with technology. The interdisciplinary nature of DH has always integrated diverse scholarly and technical practices. In recent years, DH has seen a notable increase in interest in the use of Linked Data and semantic technologies to improve the discoverability and accessibility of CH collections. LOUD's emphasis on user-centred design and usability aligns well with these goals. Consequently, the principles of LOUD hold great promise for advancing the integration and use of community-driven APIs and/or Linked Data within DH. This can be seen within PIA, where the benefits of implementing IIF helped us to streamline machine-generated annotations, integrate different thumbnails into GUI prototypes, model photo albums with different layers from the Kreis Family collection, and enable project members and students to engage in digital storytelling, an important participatory facet that can be seamlessly explored by DH efforts and CHIs with the help of the IIF Image and Presentation APIs. Data reuse is definitely a key LOUD driver, which could have been done more extensively with a productive instance of Linked Art. As for widening participation, this is definitely a strategic and political decision, rather than a technical one. That said, LOUD specifications can definitely be embedded through strategic citizen science initiatives.

A recent example that highlights the comprehensive value of Linked Data was presented by [Newbury \[2024\]](#) at the Coalition for Networked Information (CNI) Spring 2024 Meeting. He delineated its significance as extending well beyond single entities, such as the Getty Research Institute, to enrich a vast ecosystem. Specifically, he identified three principal areas of value: Firstly, within the ecosystem itself, where the utility of information is amplified through its application in diverse contexts. Secondly, for the audience, by directly addressing user needs and facilitating various conceptual frameworks. And finally, within the community, by enabling wider use and adaptation of data and code. This approach to Linked Data, as articulated by Newbury, not only enhances its utility across these dimensions, but also aligns seamlessly with the LOUD proposition, underscoring a shared vision for a digital space where

the interconnectedness and accessibility of (meta)data serve as foundational principles for progress and community engagement.

LUX, as a catalyst for LOUD, exemplifies a practical approach to implementing Linked Data that has garnered significant local engagement and support at Yale. This initiative demonstrates how sound socio-technical practices can be effectively applied within a supportive institutional environment. The consistency of the data within LUX aligns well with IIF and Linked Art standards, with only a few minor adjustments required for full compliance. These quick fixes are manageable and do not detract from the overall robustness of the initiative. While it may be too early to fully assess the wider impact of using LOUD specifications on the LUX platform within the CH domain, the initiative has already attracted considerable interest in recent months. This growing attention suggests that the LUX approach is resonating with other organisations, suggesting the potential for wider adoption and impact. The enthusiastic local engagement at Yale provides a strong foundation for LUX and highlights its potential to serve as a model for similar projects aimed at enriching digital heritage through effective collaboration and agreed-upon standards.

In carrying out this thesis, I have adhered to the five main objectives set out at the beginning of the PhD. These objectives have been accomplished to a high degree, reflecting a substantial and well-executed project. Furthermore, most of the outputs – such as data models and scripts – from this work are available on GitHub, providing open access to the wider community. In addition, I have published several papers, both individually and collaboratively, further disseminating the findings and contributions of this research.

Additionally, this thesis is relevant because it sheds light on communities and implementations that can be celebrated not only for their standards but also for their operating ethos; IIF and Linked Art present models ripe for emulation beyond their immediate digital confines. Here, agency and authority are most typically granted to the collective over the isolated, with each actor – be it an individual, an institution or an interface – intricately interconnected. Yale's LUX initiative also embodies this ethos, demonstrating how collaborative efforts can lead to innovative solutions and wider impact. It is to be hoped, then, that these practices of openness and multiple partnerships will not be seen as limited to their origins in digital representation. At the very least, I hope that these socio-technical approaches can serve as exemplars or sources of inspiration in broader arenas, where the principles of mutual visibility and concerted action can point the way towards cohesive and adaptive collaborative architectures.

Despite its contribution, this thesis is far from perfect and certainly contains several shortcomings. I will name here three significant ones. First, the visualisations included and the use of FOL are primarily designed to support my own self-reflection and may be more beneficial to me than to the broader academic community. While they provide insights into my research process and findings, their applicability and usefulness to others might be limited. Second, the theoretical framework I employed, while instrumental to my research, may not serve as a universally applicable toolbox. Nevertheless, I urge readers to pay close attention to STS methodologies and practices. The works of Bruno Latour, Donna Haraway, and Susan Leigh Star have been invaluable companions throughout this dissertation. Additionally, for those involved in conceptualising semantic information, I recommend exploring Floridi's PI, which offers profound insights into the nature and dynamics of information. These readings have greatly influenced my approach and understanding, and I believe they can offer valuable perspectives to others as well. Third, while the thesis aims to address both community practices and semantic interoperability, it leans more heavily towards the former. This emphasis on community practices may overshadow the broader discussion of semantic interoperability, potentially limiting the appeal of the thesis to those primarily interested in the technical aspects.

Other shortcomings include the broad scope of the thesis, with three empirical chapters exploring different avenues. While this comprehensive approach provides a broad understanding of the research topic, it has also resulted in a rather lengthy thesis. This may be a challenge for the reader, as a topic of interest in one chapter may not be as compelling in another. The diversity of empirical focus, while enriching the research, may dilute the coherence for some readers, making it more difficult to maintain a consistent engagement throughout the dissertation. Despite these limitations, I hope that the different perspectives and findings contribute to a richer, more nuanced understanding of LOUD for CH.

Avenues for future research are numerous and promising. One interesting area to explore is the comparative benefits experienced by early adopters of IIF and Linked Art specifications versus those who implemented these standards later. Early adopters have the advantage of having their use cases discussed and resolved within the community, and it would be insightful to analyse the long-term impacts on their projects. Such a study is already feasible for early adopters of IIF and will become possible to compare further implementations of Linked Art within a few years. Furthermore, future exploration could focus on the full implementation of Linked Art within PIA or similar efforts, as well as more performance-oriented testing with the deployed LOUD APIs. These efforts should further validate the robustness and scalability of these specifications. Another important area for future investigation is the participation of institutions and individuals from the Global South in both the IIF and Linked Art communities. It is crucial to explore how we can better support their uptake of these specifications and encourage their active involvement in these initiatives to ensure a more inclusive and globally representative environment.

As I reflect on the journey of this thesis, I am reminded of the powerful dialogue and collective effort that has been at its heart. Mr Gee's poem resonates deeply with my own aspirations for this work: to enhance our understanding of the past through openness and collaboration, as can be seen in IIF and Linked Art. As I bring this dissertation to a close, I am filled with a sense of accomplishment and a renewed commitment to promoting sound socio-technical practices. It is my hope that the insights and methodologies presented here will inspire others to engage in this ongoing dialogue, continually asking and answering the many questions that arise as we collectively explore our cultural heritage landscapes.

Bibliography

- Alessandro Adamou. Shout LOUD on a road trip to FAIRness: experience with integrating open research data at the Bibliotheca Hertziana. *Journal of Art Historiography*, 2022(27s), December 2022. ISSN 2042-4752. URL <https://arthistoriography.wordpress.com/27s-dec22/>.
- Alessandro Adamou, Davide Picca, Yumeng Hou, and Paula Loreto Granados-García. The Facets of Intangible Heritage in Southern Chinese Martial Arts: Applying a Knowledge-driven Cultural Contact Detection Approach. *Journal on Computing and Cultural Heritage*, 16(3):63:1–63:27, August 2023. ISSN 1556-4673. doi:10.1145/3606702. URL <https://dl.acm.org/doi/10.1145/3606702>.
- Yahaya Ahmad. The Scope and Definitions of Heritage: From Tangible to Intangible. *International Journal of Heritage Studies*, 12(3):292–300, March 2006. ISSN 1352-7258. doi:10.1080/13527250600604639. URL <https://www.tandfonline.com/doi/full/10.1080/13527250600604639>.
- Nadeem Akhtar. Social Network Analysis Tools. In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pages 388–392, Bhopal, India, April 2014. IEEE. ISBN 978-1-4799-3070-8. doi:10.1109/CSNT.2014.83. URL <https://ieeexplore.ieee.org/abstract/document/6821424>.
- Madeleine Akrich, Michel Callon, and Bruno Latour, editors. *Sociologie de la traduction: Textes fondateurs*. Sciences sociales. Presses des Mines, Paris, France, 2006. ISBN 978-2-35671-023-9. URL <https://doi.org/10.4000/books.pressesmines.1181>.
- J. Alakuijala and Z. Szabadka. Brotli Compressed Data Format. Request for Comments RFC7932, Internet Engineering Task Force, July 2016. URL <https://doi.org/10.17487/RFC7932>.
- Vladimir Alexiev. Museum Linked Open Data: Ontologies, Datasets, Projects. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 8:19–50, September 2018. ISSN 2535-0366, 1314-4006. doi:10.55630/dipp.2018.8.1. URL <https://dipp.math.bas.bg/dipp/article/view/dipp.2018.8.1>.
- Henk Alkemade, Steven Claeysens, Giovanni Colavizza, Nuno Freire, Jörg Lehmann, Clemens Neudecker, Giulia Osti, and Daniel van Strien. Datasheets for Digital Cultural Heritage Datasets. *Journal of Open Humanities Data*, 9(1): 1–17, October 2023. ISSN 2059-481X. doi:10.5334/johd.124. URL <https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.124>.
- Laurie Allen. Why Experiment: Machine Learning at the Library of Congress, November 2023. URL <https://blogs.loc.gov/thesignal/2023/11/why-experiment-machine-learning-at-the-library-of-congress/>.
- George Alter, Flavio Rizzolo, and Kathi Schleidt. View points on data points: A shared vocabulary for cross-domain conversations on data and metadata. *IASSIST Quarterly*, 47(1):1–39, March 2023. ISSN 2331-4141. doi:10.29173/iq1051. URL <https://iassistquarterly.com/index.php/iassist/article/view/1051>.
- Michael L. Anderson. *After Phrenology: Neural Reuse and the Interactive Brain*. The MIT Press, 2014. ISBN 978-0-262-32067-2. doi:10.7551/mitpress/10111.001.0001. URL <https://direct.mit.edu/books/book/3691/after-phrenologyneural-reuse-and-the-interactive>.
- S.L. Andresen. John McCarthy: father of AI. *IEEE Intelligent Systems*, 17(5):84–85, September 2002. ISSN 1941-1294. doi:10.1109/MIS.2002.1039837. URL <https://ieeexplore.ieee.org/abstract/document/1039837>.
- Jane Andresoo. The hundred-year-old national library’s message to future generations. *Bosniaca*, pages 90–94, December 2018. ISSN 2303-8888, 1512-5033. doi:10.37083/bosn.2018.23.90. URL <http://bosniaca.nub.ba/index.php/bosniaca/article/view/438>.
- Nicola Angius, Giuseppe Primiero, and Raymond Turner. The Philosophy of Computer Science. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition, 2021. URL <https://plato.stanford.edu/archives/spr2021/entries/computer-science/>.
- Michael Appleby, Tom Crane, Robert Sanderson, Jon Stroop, and Simeon Warner. IIIF Design Principles, March 2018. URL https://iiif.io/api/annex/notes/design_principles/.
- Michael Appleby, Tom Crane, Robert Sanderson, Jon Stroop, and Simeon Warner. IIIF Image API 3.0, June 2020a. URL <https://iiif.io/api/image/3.0/>.

- Michael Appleby, Tom Crane, Robert Sanderson, Jon Stroop, and Simeon Warner. IIIF Presentation API 3.0, June 2020b. URL <https://iiif.io/api/presentation/3.0/>.
- Michael Appleby, Tom Crane, Robert Sanderson, Jon Stroop, and Simeon Warner. IIIF Change Discovery API 1.0.0, June 2021. URL <https://iiif.io/api/discovery/1.0/>.
- Michael Appleby, Dawn Childress, Tom Crane, Jeff Mixter, Robert Sanderson, and Simeon Warner. IIIF Content Search API 2.0, November 2022a. URL <https://iiif.io/api/search/2.0/>.
- Michael Appleby, Dawn Childress, Tom Crane, Jeff Mixter, Robert Sanderson, Simeon Warner, and Maria Whitaker. IIIF Content State API 1.0, February 2022b. URL <https://iiif.io/api/content-state/1.0/>.
- Michael Appleby, Dawn Childress, Tom Crane, Jeff Mixter, Robert Sanderson, and Simeon Warner. IIIF Authorization Flow API 2.0, June 2023. URL <https://iiif.io/api/auth/2.0/>.
- Zaki Aslan. Protective Structures for the Conservation and Presentation of Archaeological Sites. *Journal of Conservation and Museum Studies*, 3(0):16, November 1997. ISSN 1364-0429. doi:10.5334/jcms.3974. URL <https://jcms-journal.com/articles/10.5334/jcms.3974>.
- Henriette D. Avram. The MARC Pilot Project. Final Report. Technical Report ED029663, Library of Congress, Washington DC, USA, 1968. URL <https://eric.ed.gov/?id=ED029663>.
- Elaine Azzopardi, Jasper O. Kenter, Juliette Young, Chris Leakey, Seb O'Connor, Simone Martino, Wesley Flannery, Lisa P. Sousa, Dimitra Mylona, Katia Frangoudes, Irène Béguier, Maria Pafi, Arturo Rey da Silva, Jacob Ainscough, Manos Koutrakis, Margarida Ferreira da Silva, and Cristina Pita. What are heritage values? Integrating natural and cultural heritage into environmental valuation. *People and Nature*, 5(2):368–383, 2023. ISSN 2575-8314. doi:10.1002/pan3.10386. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/pan3.10386>.
- Franz Baader and Carsten Lutz. 13 Description logic. In Patrick Blackburn, Johan Van Benthem, and Frank Wolter, editors, *Studies in Logic and Practical Reasoning*, volume 3 of *Handbook of Modal Logic*, pages 757–819. Elsevier, January 2007. doi:10.1016/S1570-2464(07)80016-4. URL <https://www.sciencedirect.com/science/article/pii/S1570246407800164>.
- Murtha Baca and Patricia Harpring. Categories for the description of works of art. Report, Getty Research Institute, October 2017. URL <https://apo.org.au/node/14985>.
- Bruno Bachimont. Archive et mémoire : le numérique et les mnémophores. *Signata. Annales des sémiotiques / Annals of Semiotics*, 1(12), May 2021. ISSN 2032-9806. doi:10.4000/signata.2980. URL <https://journals.openedition.org/signata/2980>.
- Phil Bainbridge. Extract text from multiple Google Docs into a Sheet, July 2021. URL <https://www.pbainbridge.co.uk/2021/07/extract-text-from-multiple-google-docs.html>.
- Kyle Banerjee. The Linked Data Myth, August 2020. URL <https://www.libraryjournal.com/story/the-linked-data-myth>.
- Vincenzo Barrile and Ernesto Bernardo. Big Data and Cultural Heritage. In Francesco Calabrò, Lucia Della Spina, and María José Piñeira Mantiñán, editors, *New Metropolitan Perspectives*, volume 482 of *Lecture Notes in Networks and Systems*, pages 2708–2716, Cham, 2022. Springer International Publishing. ISBN 978-3-031-06825-6. doi:10.1007/978-3-031-06825-6_259. URL https://link.springer.com/chapter/10.1007/978-3-031-06825-6_259.
- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1):361–362, March 2009. ISSN 2334-0770. doi:10.1609/icwsm.v3i1.13937. URL <https://ojs.aaai.org/index.php/ICWSM/article/view/13937>.
- Robert Battle and Edward Benson. Bridging the semantic Web and Web 2.0 with Representational State Transfer (REST). *Journal of Web Semantics*, 6(1):61–69, February 2008. ISSN 1570-8268. doi:10.1016/j.websem.2007.11.002. URL <https://www.sciencedirect.com/science/article/pii/S1570826807000510>.
- Florian Bauer and Martin Kaltenböck. *Linked Open Data: The Essentials*. edition mono, Vienna, Austria, 2012. ISBN 978-3-902796-05-9.
- Federico Becattini, Pietro Bongini, Luana Bulla, Alberto Del Bimbo, Ludovica Marinucci, Misael Mongiovi, and Valentina Presutti. VISCONT: A Large-scale Multilingual Visual Question Answering Dataset for Cultural Heritage. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 19(6):193:1–193:20, July 2023. ISSN 1551-6857. doi:10.1145/3590773. URL <https://dl.acm.org/doi/10.1145/3590773>.

- Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2):599–611, February 2013. ISSN 0167739X. doi:10.1016/j.future.2011.08.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167739X11001439>.
- Dave Beckett. RDF/XML Syntax Specification (Revised), February 2004. URL <https://www.w3.org/TR/REC-rdf-syntax/>.
- Davia Beckett, Tim Berners-Lee, Eric Prud'hommeaux, and Gavin Carothers. RDF 1.1 Turtle, February 2014. URL <https://www.w3.org/TR/turtle/>.
- Chryssoula Bekiari, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. CIDOC Conceptual Reference Model 7.1.1, April 2021. URL <https://doi.org/10.26225/FDZH-X261>.
- Chryssoula Bekiari, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. CIDOC Conceptual Reference Model 7.1.3, February 2024. URL <https://www.cidoc-crm.org/Version/version-7.1.3>.
- Mohamed Benabdelkrim, Clément Levallois, Jean Savinien, and Céline Robardet. Opening Fields: A Methodological Contribution to the Identification of Heterogeneous Actors in Unbounded Relational Orders. *M@n@gement*, 23(1):4–18, 2020. ISSN 1286-4692. doi:10.3917/mana.231.0004. URL <https://www.cairn.info/revue-management-2020-1-page-4.htm>.
- Francesco Beretta. A challenge for historical research: Making data FAIR using a collaborative ontology management environment (OntoME). *Semantic Web*, 12(2):279–294, January 2021. ISSN 1570-0844. doi:10.3233/SW-200416. URL <https://content.iospress.com/articles/semantic-web/sw200416>.
- Francesco Beretta. Interopérabilité des données de la recherche et ontologies fondationnelles : un éco-système d'extensions du CIDOC CRM pour les sciences humaines et sociales. In Nicolas Lasolle, Olivier Bruneau, and Jean Lieber, editors, *Actes des journées humanités numériques et Web sémantique*, pages 2–22, Nancy, France, 2022. Les Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les Technologies (AHP-PreST); Laboratoire lorrain de recherche en informatique et ses applications (LORIA). doi:10.5281/zenodo.7014341.
- Emmanuelle Bermès. Modélisons un peu : le choix d'un type de bases de données, December 2023. URL <https://figoblog.org/2023/12/13/modelisons-un-peu-le-choix-dun-type-de-bases-de-donnees/>.
- Eleonora Bernasconi, Miguel Ceriani, and Massimo Mecella. Linked Data interfaces: a survey. In Bardi Alessia, Falcon Alex, Ferilli Stefano, Marchesin Stefano, and Redavid Domenico, editors, *Proceedings of the 19th Conference on Information and Research Science Connecting to Digital and Library Science*, volume 3365 of *CEUR Workshop Proceedings*, pages 1–16, Bari, Italy, March 2023. CEUR. URL <https://ceur-ws.org/Vol-3365/#paper1>. ISSN: 1613-0073.
- Tim Berners-Lee. WorldWideWeb - executive summary, August 1991. URL <https://archive.md/Lfopj>.
- Tim Berners-Lee. Realising the Full Potential of the Web. *Technical Communication*, 46(1):79–82, 1999. ISSN 0049-3155. URL <https://www.jstor.org/stable/43088605>.
- Tim Berners-Lee. Linked Data, July 2006. URL <https://www.w3.org/DesignIssues/LinkedData.html>.
- Tim Berners-Lee. Long Live the Web. *Scientific American*, 303(6):80–85, 2010. ISSN 0036-8733. URL <https://www.jstor.org/stable/26002308>.
- Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The World-Wide Web. *Communications of the ACM*, 37(8):76–82, August 1994. ISSN 0001-0782. doi:10.1145/179606.179671. URL <https://dl.acm.org/doi/10.1145/179606.179671>.
- Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001. ISSN 0036-8733. URL <https://www.jstor.org/stable/26059207>.
- Tim Berners-Lee, Roy T. Fielding, and Larry M. Masinter. Uniform Resource Identifier (URI): Generic Syntax. Request for Comments RFC 3986, Internet Engineering Task Force, January 2005. URL <https://doi.org/10.17487/RFC3986>.
- Hanjo Berressem. Déjà Vu: Serres after Latour, Deleuze after Harman, "Nature Writing" after "Network Theory". *Amerikastudien / American Studies*, 60(1):59–79, 2015. ISSN 0340-2827. URL <https://www.jstor.org/stable/44071895>.
- Sonja Bezjak, Philipp Conzett, Pedro L. Fernandes, Edit Görögh, Kerstin Helbig, Bianca Kramer, Ignasi Labastida, Kyle Niemeyer, Fotis Psomopoulos, Tony Ross-Hellauer, René Schneider, Jon Tennant, Ellen Verbakel, and April Clyburne-Sherin. The Open Science Training Handbook. *FOSTER*, March 2019. doi:10.5281/zenodo.2587951. URL <https://book.fosteropenscience.eu>.

- Praphulla Ms Bhawsar, Erich Bremer, Máire A Duggan, Stephen Chanock, Montserrat Garcia-Closas, Joel Saltz, and Jonas S Almeida. ImageBox3: No-Server Tile Serving to Traverse Whole Slide Images on the Web, May 2023. URL <https://www.researchsquare.com/article/rs-2864977/v1>.
- Ceri Binding, Tim Evans, Jo Gilham, Douglas Tudhope, and Holly Wright. Linked Data for the Historic Environment. *Internet Archaeology*, 2022(59), July 2022. ISSN 1363-5387. doi:10.11141/ia.59.7. URL <http://intarch.ac.uk/journal/issue59/7/index.html>.
- LuAnn Bishop. Making Connections, December 2023. URL <https://your.yale.edu/youryale/making-connections-0>.
- Christian Bizer, Tom Heath, Kingsley Idehen, and Tim Berners-Lee. Linked data on the web (LDOW2008). In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 1265–1266, New York, NY, USA, April 2008. Association for Computing Machinery. ISBN 978-1-60558-085-2. doi:10.1145/1367497.1367760. URL <https://doi.org/10.1145/1367497.1367760>.
- Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, July 2009. ISSN 1552-6283. doi:10.4018/jswis.2009081901. URL <https://www.igi-global.com/article/linked-data-story-far/37496>.
- Blue Shield. Blue Shield Statutes (Articles of Association), April 2016. URL <https://web.archive.org/web/20230802104458/https://theblueshield.org/wp-content/uploads/2021/12/statute-Amendments.BSI.2016.pdf>.
- J. David Bolter and Richard A. Grusin. *Remediation: understanding new media*. MIT Press, Cambridge, Mass, 1999. ISBN 978-0-262-02452-5.
- Rick Bonney. Citizen Science: A Lab Tradition. *Living Bird*, 15(4):7–15, 1996. ISSN 1059-521X. URL <https://www.biodiversitylibrary.org/page/59717773>.
- Stefano Borgo, Roberta Ferrario, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Daniele Porello, Emilio M. Sanfilippo, and Laure Vieu. DOLCE: A descriptive ontology for linguistic and cognitive engineering. *Applied Ontology*, 17(1):45–69, January 2022. ISSN 1570-5838. doi:10.3233/AO-210259. URL <https://content.iospress.com/articles/applied-ontology/ao210259>. Publisher: IOS Press.
- Dominique Boullier. Médialab stories: How to align actor network theory and digital methods. *Big Data & Society*, 5(2):2053951718816722, July 2018. ISSN 2053-9517. doi:10.1177/2053951718816722. URL <https://doi.org/10.1177/2053951718816722>. Publisher: SAGE Publications Ltd.
- Pierre Bourdieu. *Homo Academicus*. Stanford University Press, Stanford, 1988. ISBN 978-0-8047-1798-4.
- Blythe A. Bowman. Transnational Crimes Against Culture: Looting at Archaeological Sites and the “Grey” Market in Antiquities. *Journal of Contemporary Criminal Justice*, 24(3):225–242, August 2008. ISSN 1043-9862. doi:10.1177/1043986208318210. URL <https://doi.org/10.1177/1043986208318210>. Publisher: SAGE Publications Inc.
- Scott O. Bradner. Key words for use in RFCs to Indicate Requirement Levels. Request for Comments RFC 2119, Internet Engineering Task Force, March 1997. URL <https://doi.org/10.17487/RFC2119>.
- Dan Brickley and R.V. Guha. RDF Schema 1.1, February 2014. URL <https://www.w3.org/TR/rdf11-schema/>.
- Karen Brown, Alissandra Cummins, and Ana S. González Rueda, editors. *Communities and Museums in the 21st Century: Shared Histories and Climate Action*. ICOM Advances in Museum Research. Routledge, London, UK, 1 edition, August 2023. ISBN 978-1-00-328813-8. URL <https://doi.org/10.4324/9781003288138>.
- Susan Brown and Kimberley Martin. Towards a Linked Infrastructure for Networked Cultural Scholarship. In *CSDH-SCHN 2020*, Virtual, June 2020. doi:10.17613/nwnq-tm04. URL <https://hcommons.org/deposits/item/hc:30219/>.
- Susan Brown, Erin Canning, Kim Martin, Sarah Roger, and Zach Schoenberger. Linking Communities of Practice. In *CSDH-SCHN 2021*, Virtual event, 2021. doi:10.17613/w6r2-n763. URL <https://hcommons.org/deposits/item/hc:39049/>.
- Sasha Bruns, Tabea Tietz, and Harald Sack. 2.6 Logical Inference with RDF(S), 2023. URL <https://open.hpi.de/courses/knowledgegraphs2023/items/1bKJLIFAZfOcNGm3ZTjVvk>.
- George Bruseker, Nicola Carboni, and Anaïs Guillem. Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. In Matthew L. Vincent, Víctor Manuel López-Menchero Bendicho, Marinos Ioannides, and Thomas E. Levy, editors, *Heritage and Archaeology in the Digital Age: Acquisition, Curation, and Dissemination of Spatial Cultural Heritage Data*, Quantitative Methods in the Humanities and Social Sciences, pages 93–131. Springer International Publishing, Cham, 2017. ISBN 978-3-319-65370-9. doi:10.1007/978-3-319-65370-9_6. URL https://doi.org/10.1007/978-3-319-65370-9_6.

- Christian Bueger. Praxiography. In Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug, and Richard A. Williams, editors, *SAGE Research Methods Foundations*. SAGE Publications Ltd, London, UK, September 2019. ISBN 978-1-5264-2103-6. URL <https://doi.org/10.4135/9781526421036807154>.
- Jørgen Burchardt. Researchers Outside APC-Financed Open Access: Implications for Scholars Without a Paying Institution. *Sage Open*, 4(4):2158244014551714, October 2014. ISSN 2158-2440. doi:10.1177/2158244014551714. URL <https://doi.org/10.1177/2158244014551714>.
- James Cahill. Open-Source Code for LUX Cross-Collection Discovery Platform, May 2024. URL <https://www.progress.com/blogs/yale-university-shares-open-source-code-for-lux-cross-collection-discovery-platform>.
- Michel Callon. Actor-Network Theory—The Market Test. *The Sociological Review*, 47(1_suppl):181–195, May 1999. ISSN 0038-0261. doi:10.1111/j.1467-954X.1999.tb03488.x. URL <https://doi.org/10.1111/j.1467-954X.1999.tb03488.x>. Publisher: SAGE Publications Ltd.
- Michel Callon. Actor Network Theory. In Neil J. Smelser and Paul B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 62–66. Pergamon, Oxford, January 2001. ISBN 978-0-08-043076-8. doi:10.1016/B0-08-043076-7/03168-5. URL <https://www.sciencedirect.com/science/article/pii/B0080430767031685>.
- Fiona Cameron. Beyond the Cult of the Replicant: Museums and Historical Digital Objects—Traditional Concerns, New Discourses. In Fiona Cameron and Sarah Kenderdine, editors, *Theorizing Digital Cultural Heritage: A Critical Discourse*. The MIT Press, March 2007. ISBN 978-0-262-03353-4. doi:10.7551/mitpress/9780262033534.003.0004. URL <https://doi.org/10.7551/mitpress/9780262033534.003.0004>.
- Alexandre Camus, Dessislava Paneva-Marinova, and Detelin Luchev. Digital Humanities: Challenges of the Transformation of Tools and Objects of Knowledge in Contemporary Humanities. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 3:109–118, September 2013. ISSN 2535-0366, 1314-4006. doi:10.55630/dipp.2013.3.11. URL <https://dipp.math.bas.bg/dipp/article/view/dipp.2013.3.11>.
- Gustavo Candela, Pilar Escobar, Rafael C. Carrasco, and Manuel Marco-Such. Migration of a library catalogue into RDA linked open data. *Semantic Web*, 9(4):481–491, June 2018. ISSN 22104968, 15700844. doi:10.3233/SW-170274. URL <https://content.iospress.com/articles/semantic-web/sw274>.
- Gustavo Candela, Nele Gabriëls, Sally Chambers, Milena Dobрева, Sarah Ames, Meghan Ferriter, Neil Fitzgerald, Victor Harbo, Katrine Hofmann, Olga Holownia, Alba Irollo, Mahendra Mahey, Eileen Manchester, Thuy-An Pham, Abigail Potter, and Ellen Van Keer. A checklist to publish collections as data in GLAM institutions. *Global Knowledge, Memory and Communication*, ahead-of-print(ahead-of-print), January 2023a. ISSN 2514-9342. doi:10.1108/GKMC-06-2023-0195. URL <https://doi.org/10.1108/GKMC-06-2023-0195>.
- Gustavo Candela, Javier Pereda, Dolores Sáez, Pilar Escobar, Alexander Sánchez, Andrés Villa Torres, Albert A. Palacios, Kelly McDonough, and Patricia Murrieta-Flores. An ontological approach for unlocking the Colonial Archive. *Journal on Computing and Cultural Heritage*, page 3594727, April 2023b. ISSN 1556-4673, 1556-4711. doi:10.1145/3594727. URL <https://dl.acm.org/doi/10.1145/3594727>.
- Erin Canning. What is a museum object according to a museum database?, October 2023. URL <https://vimeo.com/878850227>.
- Erin Canning, Susan Brown, Sarah Roger, and Kimberley Martin. The Power to Structure : Making Meaning from Metadata Through Ontologies. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 6(3):1–15, July 2022. ISSN 2398-4112. doi:10.18357/kula.169. URL <https://kula.uvic.ca/index.php/kula/article/view/169>.
- Linda Cantara. METS: The Metadata Encoding and Transmission Standard. *Cataloging & Classification Quarterly*, 40(3-4):237–253, September 2005. ISSN 0163-9374. doi:10.1300/J104v40n03_11. URL https://doi.org/10.1300/J104v40n03_11.
- Sarven Capadisli. *Linked Research on the Decentralised Web*. PhD Thesis, Rheinischen Friedrich-Wilhelms-Universität Bonn, May 2020. URL <https://hdl.handle.net/20.500.11811/8352>.
- Sarven Capadisli. Social and Technical Decisions — Inherently Social, Decentralised, and for Everyone, September 2023. URL <https://csarven.ca/presentations/inherently-social-decentralised-and-for-everyone>.
- Priscilla Caplan and Rebecca S Guenther. Practical Preservation: The PREMIS Experience. *Library Trends*, 54(1): 111–124, 2005. ISSN 1559-0682. URL <https://muse.jhu.edu/pub/1/article/193223>.
- Carlotta Capurro and Gertjan Plets. Europeana, EDM, and the Europeanisation of Cultural Heritage Institutions. *Digital Culture & Society*, 6(2):163–190, December 2020. ISSN 2364-2122, 2364-2114. doi:10.14361/dcs-2020-0209. URL <https://www.degruyter.com/document/doi/10.14361/dcs-2020-0209/html>.

- Costanza Caraffa and Tiziana Serena. Introduction: Photography, Archives and the Discourse of Nation. In Costanza Caraffa and Tiziana Serena, editors, *Photo Archives and the Idea of Nation*, pages 3–16. De Gruyter, Berlin, Germany, January 2015. ISBN 978-3-11-033181-3 978-3-11-033183-7. doi:[10.1515/9783110331837](https://doi.org/10.1515/9783110331837). URL <https://www.degruyter.com/view/title/304249>.
- John Carman. Where the Value Lies: the importance of materiality to the immaterial aspects of heritage. In Emma Waterton and Laurajane Smith, editors, *Taking Archaeology Out of Heritage*, pages 192–208. Cambridge Scholars Publishing, Cambridge, UK, 2009. ISBN 978-1-4438-1442-3.
- Stephanie Russo Carroll, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, Kay Raseroka, Desi Rodriguez-Lonebear, Robyn Rowe, Rodrigo Sara, Jennifer D. Walker, Jane Anderson, and Maui Hudson. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1):43, November 2020. ISSN 1683-1470. doi:[10.5334/dsj-2020-043](https://doi.org/10.5334/dsj-2020-043). URL <https://datascience.codata.org/articles/10.5334/dsj-2020-043>.
- Stephanie Russo Carroll, Edit Herczog, Maui Hudson, Keith Russell, and Shelley Stall. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, 8(1):108, April 2021. ISSN 2052-4463. doi:[10.1038/s41597-021-00892-0](https://doi.org/10.1038/s41597-021-00892-0). URL <https://www.nature.com/articles/s41597-021-00892-0>.
- V. Cerf and R. Kahn. A Protocol for Packet Network Intercommunication. *IEEE Transactions on Communications*, 22(5):637–648, May 1974. ISSN 1558-0857. doi:[10.1109/TCOM.1974.1092259](https://doi.org/10.1109/TCOM.1974.1092259). URL <https://ieeexplore.ieee.org/abstract/document/1092259>.
- Matthieu Cevey and Julien Antoine Raemy. *Swiss PID Hub : Création d'un Hub pour la gestion des identifiants pérennes en Suisse*. Mémoire de recherche, Haute école de gestion de Genève, Carouge, Suisse, January 2020. URL <https://sonar.ch/hesso/documents/315101>.
- Sally Chambers, Frédéric Lemmers, Thuy-An Pham, Julie M. Birkholz, Vincent Ducatteeuw, Antoine Jacquet, Wout Dillen, Dilawar Ali, Kenzo Milleville, and Steven Verstockt. Collections as Data: interdisciplinary experiments with KBR's digitised historical newspapers : a Belgian case study. In *DH Benelux 2021, Abstracts*, Leiden, The Netherlands, 2021. URL <http://hdl.handle.net/1854/LU-8712404>.
- Liang Chang, Uli Sattler, and Tianlong Gu. An ABox Revision Algorithm for the Description Logic EL_{bot}. In Meghyn Bienvenu, Magdalena Ortiz, Riccardo Rosati, and Mantas Simkus, editors, *Informal Proceedings of the 27th International Workshop on Description Logics*, volume 1193 of *CEUR Workshop Proceedings*, pages 459–470, Vienna, Austria, July 2014. CEUR. URL https://ceur-ws.org/Vol-1193/#paper_64. ISSN: 1613-0073.
- Anne Chardonens. *La gestion des données d'autorité archivistiques dans le cadre du Web de données*. Thèse de doctorat, Université libre de Bruxelles, Bruxelles, Belgique, December 2020. URL <https://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/315804>. Publisher: Université libre de Bruxelles.
- David Charles. Aristotle on Agency. In Oxford Handbooks Editorial Board, editor, *The Oxford Handbook of Topics in Philosophy*, page 0. Oxford University Press, Oxford, UK, May 2017. ISBN 978-0-19-993531-4. URL <https://doi.org/10.1093/oxfordhb/9780199935314.013.6>.
- Valentine Charles and Antoine Isaac. Enhancing the Europeana Data Model (EDM). White paper, Europeana Foundation, May 2015. URL <http://pro.europeana.eu/files/Europeana.Professional/Publications/EDM.WhitePaper.17062015.pdf>.
- Vera Chiquet. How to digitally preserve UNESCO intangible cultural heritage? A web-archive for ephemeral events at the Basler Carnival. In *Archiving Conference*, volume 20, pages 105–108, Oslo, Norway, June 2023. Society for Imaging Science and Technology. doi:[10.2352/issn.2168-3204.2023.20.1.22](https://doi.org/10.2352/issn.2168-3204.2023.20.1.22). URL <https://library.imaging.org/archiving/articles/20/1/22>.
- Vera Chiquet, Ulrike Felsing, and Peter Fornaro. A Participatory Interface for a Photo Archives. In *Archiving Conference*, volume 20, pages 109–111, Oslo, Norway, June 2023. Society for Imaging Science and Technology. doi:[10.2352/issn.2168-3204.2023.20.1.23](https://doi.org/10.2352/issn.2168-3204.2023.20.1.23). URL <https://library.imaging.org/archiving/articles/20/1/23>.
- Ryan Chute and Herbert Van De Sompel. Introducing djabatoka: A Reuse Friendly, Open Source JPEG 2000 Image Server. *D-Lib Magazine*, 14(9/10), September 2008. ISSN 1082-9873. doi:[10.1045/september2008-chute](https://doi.org/10.1045/september2008-chute). URL <https://mirror.dlib.org/dlib/september08/chute/09chute.html>.
- Arianna Ciula and Øyvind Eide. Modelling in digital humanities: Signs in context. *Digital Scholarship in the Humanities*, page fqw045, September 2016. ISSN 2055-7671, 2055-768X. doi:[10.1093/llc/fqw045](https://doi.org/10.1093/llc/fqw045). URL <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqw045>.

- Florence Clavaud and Tobias Wildi. ICA Records in Contexts-Ontology (RiC-O): a Semantic Framework for Describing Archival Resources. In *Linked Archives International Workshop 2021*, volume 3019, pages 79–92, Online, 2021. CEUR-WS.org. URL <https://enc.hal.science/hal-03965776>.
- Joan Cobb. The Journey to Linked Open Data: The Getty Vocabularies. *Journal of Library Metadata*, 15(3-4):142–156, October 2015. ISSN 1938-6389. doi:10.1080/19386389.2015.1103081. URL <https://www.tandfonline.com/doi/full/10.1080/19386389.2015.1103081>.
- Erin Coburn, Elisa Lanzi, Elizabeth O’Keefe, Regine Stein, and Ann Whiteside. The Cataloging Cultural Objects experience: Codifying practice for the cultural heritage community. *IFLA Journal*, 36(1):16–29, March 2010a. ISSN 0340-0352. doi:10.1177/0340035209359561. URL <https://doi.org/10.1177/0340035209359561>.
- Erin Coburn, Richard Light, Gordon McKenna, Regine Stein, and Axel Vitzthum. LIDO - Lightweight Information Describing Objects Version 1.0, November 2010b. URL <https://lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>.
- Catherine Nicole Coleman. Managing Bias When Library Collections Become Data. *International Journal of Librarianship*, 5(1):8–19, July 2020. ISSN 2474-3542. doi:10.23974/ijol.2020.vol5.1.162. URL <https://journal.calaijol.org/index.php/ijol/article/view/162>.
- Tracy Commock and Dionne Newell. Connecting Museums through Citizen Science: Jamaica/US partnership in environmental preservation. In Karen Brown, Alissandra Cummins, and Ana S. González Rueda, editors, *Communities and Museums in the 21st Century*, ICOM Advances in Museum Research, pages 137–159. Routledge, London, UK, 2023. ISBN 978-1-00-328813-8. URL <https://doi.org/10.4324/9781003288138-9>.
- Panos Constantopoulos and Costis Dallas. Aspects of a digital curation agenda for cultural heritage. In *2008 IEEE International Conference on Distributed Human-Machine Systems. Athens, Greece: IEEE*, pages 1–6, 2008.
- Paul Conway. Digital transformations and the archival nature of surrogates. *Archival Science*, 15(1):51–69, March 2015. ISSN 1389-0166, 1573-7519. doi:10.1007/s10502-014-9219-z. URL <http://link.springer.com/10.1007/s10502-014-9219-z>.
- Murielle Cornut. Open, edit, save: über die performative Materialität privater Fotoalben. In Ulrich Hägele, editor, *Kuratierte Erinnerungen: das Fotoalbum*, number 15 in Visuelle Kultur. Studien und Materialien, pages 157–170. Waxmann, Münster, Germany; New York, NY, USA, 2023. ISBN 978-3-8309-4763-9.
- Murielle Cornut, Julien Antoine Raemy, and Florian Spiess. Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning. *Journal on Computing and Cultural Heritage*, 16(4): 1–19, November 2023. ISSN 1556-4673. doi:10.1145/3625301. URL <https://dl.acm.org/doi/10.1145/3625301>.
- Manuel Corpas, Nadezda V. Kovalevskaia, Amanda McMurray, and Fiona G. G. Nielsen. A FAIR guide for data providers to maximise sharing of human genomic data. *PLOS Computational Biology*, 14(3):e1005873, March 2018. ISSN 1553-7358. doi:10.1371/journal.pcbi.1005873. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005873>.
- Amanda Frances Cossham. *Models of the bibliographic universe*. PhD Thesis, Monash University, Melbourne, Australia, July 2017. URL <https://doi.org/10.4225/03/596e9bc6c1d09>.
- Stefano Cossu. Getty Common Image Service. Research & Design Report, J. Paul Getty Trust, Los Angeles, CA, USA, 2019a. URL https://www.getty.edu/project-files/iiif_getty_research_report.pdf.
- Stefano Cossu. Labours of Love and Convenience: Dealing with Community-Supported Knowledge in Museums. *Publications*, 7(1):19, March 2019b. ISSN 2304-6775. doi:10.3390/publications7010019. URL <https://www.mdpi.com/2304-6775/7/1/19>.
- Karen Coyle and Diane Hillmann. Resource Description and Access (RDA): Cataloging Rules for the 20th Century. *D-Lib Magazine*, 13(1/2), January 2007. ISSN 1082-9873. doi:10.1045/january2007-coyle. URL <http://www.dlib.org/dlib/january07/coyle/01coyle.html>.
- Dave Cramer. W3C and IDPF: Better Together?, May 2016. URL <https://medium.com/@dauwhe/w3c-and-idpf-better-together-c92988674444>.
- Tom Cramer. IIIF Consortium Formed, June 2015. URL <https://iiif.io/news/2015/06/17/iiif-consortium/>.
- Tom Cramer. IIIF: A Brief History & Looking Forward, June 2019. URL https://youtu.be/J_NgQhlfW44.
- Tom Crane. What is IIIF Content State?, November 2020. URL <https://tom-crane.medium.com/what-is-iiif-content-state-dd15a543939f>.

- Robert Crystal-Ornelas, Charuleka Varadharajan, Dylan O’Ryan, Kathleen Beilsmith, Benjamin Bond-Lamberty, Kristin Boye, Madison Burrus, Shreyas Cholia, Danielle S. Christianson, Michael Crow, Joan Damerow, Kim S. Ely, Amy E. Goldman, Susan L. Heinz, Valerie C. Hendrix, Zarine Kakalia, Kayla Mathes, Fianna O’Brien, Stephanie C. Pennington, Emily Robles, Alistair Rogers, Maegen Simmonds, Terri Velliquette, Pamela Weisenhorn, Jessica Nicole Welch, Karen Whitenack, and Deborah A. Agarwal. Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Scientific Data*, 9(1):700, November 2022. ISSN 2052-4463. doi:10.1038/s41597-022-01606-w. URL <https://www.nature.com/articles/s41597-022-01606-w>.
- Richard Cyganiak, David Wood, and Markus Lanthaler. RDF 1.1 Concepts and Abstract Syntax, February 2014. URL <https://www.w3.org/TR/rdf11-concepts/>.
- Radosław Czahajda, Neda Čairović, and Mitja Černko. Live Online Education Efficiency Mediators From the Actor Network Theory Perspective. *Frontiers in Education*, 7, 2022. ISSN 2504-284X. URL <https://www.frontiersin.org/articles/10.3389/educ.2022.859783>.
- Enrico Daga, Luigi Asprino, Rossana Damiano, Marilena Daquino, Belen Diaz Agudo, Aldo Gangemi, Tsvi Kuflik, Antonio Lieto, Mark Maguire, Anna Maria Marras, Delfina Martinez Pandiani, Paul Mulholland, Silvio Peroni, Sofia Pescarin, and Alan Wecker. Integrating Citizen Experiences in Cultural Heritage Archives: Requirements, State of the Art, and Challenges. *Journal on Computing and Cultural Heritage*, 15(1):11:1–11:35, January 2022. ISSN 1556-4673. doi:10.1145/3477599. URL <https://doi.org/10.1145/3477599>.
- Anna Dahlgren and Karin Hansson. The Diversity Paradox: Conflicting Demands on Metadata Production in Cultural Heritage Collections. *Digital Culture & Society*, 6(2):239–256, December 2020. ISSN 2364-2122, 2364-2114. doi:10.14361/dcs-2020-0212. URL <https://www.degruyter.com/document/doi/10.14361/dcs-2020-0212/html>.
- Jérôme Darmont, Cécile Favre, Sabine Loudcher, and Camille Noûs. Data lakes for digital humanities. In *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, DTUC ’20, pages 1–4, New York, NY, USA, December 2020. Association for Computing Machinery. ISBN 978-1-4053-7753-9. doi:10.1145/3423603.3424004. URL <https://dl.acm.org/doi/10.1145/3423603.3424004>.
- Souripriya Das, Seema Sundara, and Richard Cyganiak. R2RML: RDB to RDF Mapping Language, September 2012. URL <https://www.w3.org/TR/r2rml/>.
- Edie Davis and Bahareh Heravi. Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation. *Journal on Computing and Cultural Heritage*, 14(2):21:1–21:18, May 2021. ISSN 1556-4673. doi:10.1145/3429458. URL <https://dl.acm.org/doi/10.1145/3429458>.
- Julien De Muynke, Marie Baltazar, Martin Monferran, Claudie Voisenat, and Brian F. G. Katz. Ears of the past, an inquiry into the sonic memory of the acoustics of Notre-Dame before the fire of 2019. *Journal of Cultural Heritage*, September 2022. ISSN 1296-2074. doi:10.1016/j.culher.2022.09.006. URL <https://www.sciencedirect.com/science/article/pii/S1296207422001455>.
- Jeremy Debattista, Christoph Lange, Simon Scerri, and Sören Auer. Linked ‘Big’ Data: Towards a Manifold Increase in Big Data Value and Veracity. In *2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC)*, pages 92–98, Limassol, Cyprus, December 2015. IEEE. doi:10.1109/BDC.2015.34. URL <https://ieeexplore.ieee.org/abstract/document/7406335>.
- Duane Degler, Charlie Butcosk, and Rana Chan. The Georgia O’Keeffe Museum Collections Website, January 2020. URL <https://www.designforcontext.com/insights/okeeffe-museum-collections-browser>.
- Emmanuelle Delmas-Glass. Lessons learned from digital collaborations: Standards and descriptive practices. In Lorraine A. Stuart, Thomas F.R. Clareson, and Joyce Ray, editors, *Economic Considerations for Libraries, Archives and Museums*, pages 61–76. Routledge, Abingdon, Oxon, UK; New York, NY, USA, 2021. ISBN 978-1-00-303710-1.
- Emmanuelle Delmas-Glass and Robert Sanderson. Fostering a community of PHAROS scholars through the adoption of open standards. *Art Libraries Journal*, 45(1):19–23, January 2020. ISSN 0307-4722, 2059-7525. doi:10.1017/alj.2019.32. URL <https://doi.org/10.1017/alj.2019.32>.
- Edward Demenchonok. Michel Foucault’s Theory of Practices of the Self and the Quest for a New Philosophical Anthropology. In *Peace, Culture, and Violence*, pages 218–247. Brill, February 2018. ISBN 978-90-04-36191-1. URL https://doi.org/10.1163/9789004361911_013.
- William Denton. Functional Requirements for Bibliographic Records (FRBR): Hype or Cure-All? *portal: Libraries and the Academy*, 6(2):231–232, 2006. ISSN 1530-7131. doi:10.1353/pla.2006.0018. URL <https://muse.jhu.edu/article/197083>.
- Jacques Derrida. *Mal d’archive: une impression freudienne*. Collection Incises. Galilée, Paris, nouv. éd. edition, 2008. ISBN 978-2-7186-0771-9. OCLC: 263438164.

- Digital Preservation Coalition. Persistent Identifiers, 2017. URL <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>.
- Chris Dijkshoorn. Building Collection Data Infrastructure at the Rijksmuseum. In *EuropeanaTech 2023*, The Hague, Netherlands, October 2023.
- Anastasia Dimou and Miel Vander Sande. RDF Mapping Language (RML), November 2022. URL <https://rml.io/specs/rml/>.
- Martin Doerr. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, September 2003. ISSN 2371-9621. doi:10.1609/aimag.v24i3.1720. URL <https://ojs.aaai.org/index.php/aimagazine/article/view/1720>. Number: 3.
- Georgios Drakopoulos, Evaggelos Spyrou, Yorghos Voutos, and Phivos Mylonas. A semantically annotated JSON metadata structure for open linked cultural data in Neo4j. In *Proceedings of the 23rd Pan-Hellenic Conference on Informatics, PCI '19*, pages 81–88, New York, NY, USA, November 2019. Association for Computing Machinery. ISBN 978-1-4503-7292-3. doi:10.1145/3368640.3368659. URL <https://dl.acm.org/doi/10.1145/3368640.3368659>.
- Marcel Duin. WebAssembly: Beyond the Browser, March 2022. URL <https://engineering.q42.nl/webassembly-beyond-the-browser/>.
- Alastair Dunning, Madeleine de Smaele, and Jasmin Böhmer. Are the FAIR Data Principles fair? *International Journal of Digital Curation*, 12(2):177–195, December 2017. ISSN 1746-8256. doi:10.2218/ijdc.v12i2.567. URL <http://www.ijdc.net/article/view/567>.
- Jeff Edmunds. BIBFRAME Must Die. *ScholarSphere*, pages 1–7, October 2023. doi:10.26207/V18M-0G05. URL <https://scholarsphere.psu.edu/resources/fc19faee-70b9-44b3-9346-18e40a2cd990>.
- Elizabeth Edwards. Photography and the Material Performance of the Past. *History and Theory*, 48(4):130–150, 2009. ISSN 0018-2656. URL <https://www.jstor.org/stable/25621444>. Publisher: [Wesleyan University, Wiley].
- Elizabeth Edwards. Photographs: Material Form and the Dynamic Archive*. In Costanza Caraffa, Courtauld Institute of Art, and Kunsthistorisches Institut in Florenz, editors, *Photo archives and the photographic memory of art history*, number Bd. 14 in Mandorli, pages 47–56. Deutscher Kunstverlag, Berlin, 2011. ISBN 978-3-422-07029-5.
- Elizabeth Edwards and Janice Hart, editors. *Photographs Objects Histories*. Routledge, London, UK, 1st edition edition, June 2004. ISBN 978-1-134-52357-3. doi:10.4324/9780203506493. URL <https://www.taylorfrancis.com/books/9781134523573>.
- Sabine Eggmann and Christine Bischoff. Interview : zwischen "think big" und pragmatischem Realismus - zur konzeptuellen Ausrichtung der SGV. *Schweizer Volkskunde : Korrespondenzblatt der Schweizerischen Gesellschaft für Volkskunde = Folklore suisse : bulletin de la Société suisse des traditions populaires = Folclore svizzero : bollettino della Società svizzera per le tradizioni popolari*, 100(4):157–160, 2010. doi:10.5169/SEALS-1003866. URL <https://www.e-periodica.ch/digbib/view?pid=sgv-003:2010:100:220>.
- Lisa Ehrlinger and Wolfram Wöß. Towards a Definition of Knowledge Graphs. In Michael Martin, Martí Cuquet, and Erwin Folmer, editors, *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCESS'16)*, volume 1695 of *CEUR Workshop Proceedings*, Leipzig, Germany, September 2016. CEUR. URL <https://ceur-ws.org/Vol-1695/#paper4>.
- Jeffrey P. Emanuel. Chapter 9 - Stitching Together Technology for the Digital Humanities With the International Image Interoperability Framework (IIIF). In Robin Kear and Kate Joranson, editors, *Digital Humanities, Libraries, and Partnerships*, pages 125–135. Chandos Publishing, Witney, Oxfordshire, UK, January 2018. ISBN 978-0-08-102023-4. URL <https://doi.org/10.1016/B978-0-08-102023-4.00009-4>.
- Isitor Emmanuel and Clare Stanier. Defining Big Data. In *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies, BDAW '16*, pages 1–6, New York, NY, USA, November 2016. Association for Computing Machinery. ISBN 978-1-4503-4779-2. doi:10.1145/3010089.3010090. URL <https://dl.acm.org/doi/10.1145/3010089.3010090>.
- Bill Endres. Digitizing Medieval Manuscripts: The St. Chad Gospels, Materiality, Recoveries, and Representation in 2D & 3D. In *Digitizing Medieval Manuscripts*. ARC, Amsterdam University Press, August 2019. ISBN 978-1-942401-80-3. URL <https://doi.org/10.1515/9781942401803>.

- Emily Escamilla, Lamia Salsabil, Martin Klein, Jian Wu, Michele C. Weigle, and Michael L. Nelson. It's Not Just GitHub: Identifying Data and Software Sources Included in Publications. In Omar Alonso, Helena Cousijn, Gianmaria Silvello, Mónica Marrero, Carla Teixeira Lopes, and Stefano Marchesin, editors, *Linking Theory and Practice of Digital Libraries*, Lecture Notes in Computer Science, pages 195–206, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43849-3. doi:10.1007/978-3-031-43849-3_17.
- Elika J. Etemad and Florian Rivoal. W3C Process Document, June 2023. URL <https://www.w3.org/2023/Process-20230612/>.
- European Commission. Directorate General for Research and Innovation. *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*. Publications Office, Brussels, Belgium, 2018. URL <https://data.europa.eu/doi/10.2777/1524>.
- Pavlos Fafalios, Yannis Marketakis, Anastasia Axaridou, Yannis Tzitzikas, and Martin Doerr. A workflow model for holistic data management and semantic interoperability in quantitative archival research. *Digital Scholarship in the Humanities*, 38(3):1049–1066, April 2023. ISSN 2055-7671, 2055-768X. doi:10.1093/llc/fqad018. URL <https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqad018/7110232>.
- Ulrike Felsing and Murielle Cornut. Re-Imagining the Collection of the Kreis Family. *Research in Arts and Education*, 2024(1):41–53, May 2024. ISSN 2670-2142. doi:10.54916/rae.142567. URL <https://journal.f/rae/article/view/142567>.
- Ulrike Felsing and Max Frischknecht. Critical Map Visualizations. In Christine Schranz, editor, *Shifts in Mapping*, pages 95–124. transcript Verlag, Bielefeld, Germany, December 2021. ISBN 978-3-8394-6041-2. doi:10.1515/9783839460412-008. URL <https://www.degruyter.com/document/doi/10.1515/9783839460412-008/html>.
- Ulrike Felsing, Peter Fornaro, Max Frischknecht, and Julien Antoine Raemy. Community and Interoperability at the Core of Sustaining Image Archives. *Digital Humanities in the Nordic and Baltic Countries Publications*, 5(1):40–54, October 2023. ISSN 2704-1441. doi:10.5617/dhnpub.10649. URL <https://journals.uio.no/dhnpub/article/view/10649>.
- Sabrina Ferrazzi. The Notion of “Cultural Heritage” in the International Field: Behind Origin and Evolution of a Concept. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 34(3):743–768, July 2021. ISSN 1572-8722. doi:10.1007/s11196-020-09739-0. URL <https://doi.org/10.1007/s11196-020-09739-0>.
- R. Fielding, M. Nottingham, and J. Reschke. HTTP Semantics. Request for Comments RFC9110, Internet Engineering Task Force, June 2022. URL <https://doi.org/10.17487/RFC9110>.
- Roy Thomas Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD Thesis, University of California, Irvine, CA, USA, 2000. URL <https://ics.uci.edu/~fielding/pubs/dissertation/top.htm>.
- Eleanor E. Fink. American Art Collaborative (AAC) Linked Open Data (LOD) Initiative. Overview and Recommendations for Good Practices, American Art Collaborative, 2018. URL <https://hdl.handle.net/10088/106410>.
- Sara Fiorentino and Tania Chinni. The Persistence of Memory. Exploring the Significance of Glass from Materiality to Intangible Values. *Heritage*, 6(6):4834–4842, June 2023. ISSN 2571-9408. doi:10.3390/heritage6060257. URL <https://www.mdpi.com/2571-9408/6/6/257>.
- Kathleen Fitzpatrick. Reporting from the Digital Humanities 2010 Conference, July 2010. URL <https://web.archive.org/web/20190829004943/https://www.chronicle.com/blogs/profhacker/reporting-from-the-digital-humanities-2010-conference/25473>.
- Kathleen Fitzpatrick. *Generous Thinking: A Radical Approach to Saving the University*. Johns Hopkins University Press, Baltimore, 2019. ISBN 978-1-4214-2947-2. URL <https://doi.org/10.1353/book.99579>.
- Darja Fišer and Ulrike Wuttke. Boost your eHumanities and eHeritage research with Research Infrastructures (PARTHENOS Webinar), April 2018. URL <https://hal.science/cel-01784097>.
- Luciano Floridi. Is Semantic Information Meaningful Data? *Philosophy and Phenomenological Research*, 70(2):351–370, 2005. ISSN 0031-8205. URL <https://www.jstor.org/stable/40040796>.
- Luciano Floridi. The Method of Levels of Abstraction. *Minds and Machines*, 18(3):303–329, September 2008. ISSN 1572-8641. doi:10.1007/s11023-008-9113-7. URL <https://doi.org/10.1007/s11023-008-9113-7>.
- Luciano Floridi. *Information: a very short introduction*. Number 225 in Very short introductions. Oxford University Press, Oxford ; New York, 2010. ISBN 978-0-19-955137-8.
- Luciano Floridi. *The philosophy of information*. Oxford University Press, Oxford ; New York, 2011. ISBN 978-0-19-923238-3.

- Luciano Floridi. On good and evil, the mistaken idea that technology is ever neutral, and the importance of the double-charge thesis, August 2023. URL <https://doi.org/10.2139/ssrn.4551487>.
- Donald C. Force and Randy Smith. Context Lost: Digital Surrogates, Their Physical Counterparts, and the Metadata that Is Keeping Them Apart. *The American Archivist*, 84(1):91–118, June 2021. ISSN 0360-9081. doi:10.17723/0360-9081-84.1.91. URL <https://doi.org/10.17723/0360-9081-84.1.91>.
- Peter Fornaro and Vera Chiquet. Das Digital Humanities Lab der Universität Basel und die interdisziplinäre Zusammenarbeit mit der Kunstgeschichte und der Archäologie. *Zeitschrift für schweizerische Archäologie und Kunstgeschichte (ZAK)*, 77 (2020)(2-3):111–124, 2020. doi:10.5169/SEALS-882469. URL <https://www.e-periodica.ch/digbib/view?pid=zak-003:2020:77::223>.
- FOSTER. Open Science, 2019. URL <https://www.fosteropenscience.eu/taxonomy/term/100>.
- Michel Foucault. The Subject and Power. *Critical Inquiry*, 8(4):777–795, 1982. ISSN 0093-1896. URL <https://www.jstor.org/stable/1343197>.
- Fenella G. France and Andrew Forsberg. Linked Open and Annotated Science and Heritage Data. *Archiving Conference*, 16(1):151–155, May 2019. ISSN 2161-8798. doi:10.2352/issn.2168-3204.2019.1.0.35. URL <https://library.imaging.org/archiving/articles/16/1/art00035>.
- Fenella G. France and Andrew Forsberg. Addressing the Challenges of Interoperability and Cultural Heritage Data. *Archiving Conference*, 18:33–37, June 2021. ISSN 2161-8798. doi:10.2352/issn.2168-3204.2021.1.0.8. URL <https://library.imaging.org/archiving/articles/18/1/art00008>.
- Fenella G. France and Michael B. Toth. Integrating science and art: the scriptospatial visualization interface. In *IFLA WLIC 2014 Proceedings*, pages 1–12, Lyon, France, 2014. URL <https://library.ifa.org/id/eprint/763/>.
- Nuno Freire and Antoine Isaac. Technical Usability of Wikidata’s Linked Data. In Witold Abramowicz and Rafael Corchuelo, editors, *Business Information Systems Workshops*, Lecture Notes in Business Information Processing, pages 556–567, Cham, 2019. Springer International Publishing. ISBN 9783030366919. doi:10.1007/978-3-030-36691-9_47.
- Nuno Freire, Antoine Isaac, Glen Robson, John Brooks, and Hugo Manguinhas. A survey of Web technology for metadata aggregation in cultural heritage. *Information Services & Use*, 37(4):425–436, October 2017. ISSN 01675265. doi:10.3233/ISU-170859. URL <https://content.iospress.com/articles/information-services-and-use/isu859>.
- Nuno Freire, Pável Calado, and Bruno Martins. Availability of Cultural Heritage Structured Metadata in the World Wide Web. In Leslie Chan and Pierre Mounier, editors, *ELPUB 2018*, Toronto, Canada, June 2018a. doi:10.4000/proceedings.elpub.2018.20. URL <https://hal.archives-ouvertes.fr/hal-01816696>.
- Nuno Freire, Enno Meijers, Sjors de Valk, René Voorburg, Antoine Isaac, and Roland Cornelissen. Aggregation of Linked Data : A case study in the cultural heritage domain. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 522–527, December 2018b. doi:10.1109/BigData.2018.8622348. URL <https://doi.org/10.1109/BigData.2018.8622348>.
- Nuno Freire, Glen Robson, John B. Howard, Hugo Manguinhas, and Antoine Isaac. Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org. *International Journal on Digital Libraries*, October 2018c. ISSN 1432-1300. doi:10.1007/s00799-018-0259-5. URL <https://doi.org/10.1007/s00799-018-0259-5>.
- Nuno Freire, Enno Meijers, Sjors de Valk, Julien A. Raemy, and Antoine Isaac. Metadata Aggregation via Linked Data: Results of the Europeana Common Culture Project. In Emmanouel Garoufallou and Maria-Antonia Ovalle-Perandones, editors, *Metadata and Semantic Research*, Communications in Computer and Information Science, pages 383–394, Cham, 2021. Springer International Publishing. ISBN 978-3-030-71903-6. doi:10.1007/978-3-030-71903-6_35.
- Antonella Fresa. A Data Infrastructure for Digital Cultural Heritage: Characteristics, Requirements and Priority Services. *International Journal of Humanities and Arts Computing*, 7(supplement):29–46, March 2013. ISSN 1753-8548. doi:10.3366/ijhac.2013.0058. URL <https://www.eupublishing.com/doi/abs/10.3366/ijhac.2013.0058>.
- Max Frischknecht. Generating Perspectives: Applying Generative Design to critically explore the Atlas of Swiss Folklore. In *DARIAH-CH Study Day 2022 Posters*, Mendrisio, Switzerland, 2022. DARIAH-CH. doi:10.24451/arbor.17911. URL <https://arbor.bfh.ch/17911/>.
- Ramona Fritschi. e-codices–Virtual Manuscript Library of Switzerland: Great Achievements and Even Greater Challenges to Come. *Digital Philology: A Journal of Medieval Cultures*, 6(2):244–256, 2017. ISSN 2162-9552. doi:10.1353/dph.2017.0013. URL <https://muse.jhu.edu/article/689453>.

- Jean-Gabriel Ganascia. Abstraction of levels of abstraction. *Journal of Experimental & Theoretical Artificial Intelligence*, 27(1):23–35, January 2015. ISSN 0952-813X. doi:10.1080/0952813X.2014.940685. URL <https://www.tandfonline.com/doi/full/10.1080/0952813X.2014.940685>.
- Fabien Gandon. Pour tout le monde : Tim Berners-Lee, lauréat du prix Turing 2016 pour avoir inventé... le Web. *Bulletin 1024*, 2017(11):129–154, September 2017. ISSN 2270-1419. doi:10.48556/SIF.1024.11.129. URL <https://www.societe-informatique-de-france.fr/wp-content/uploads/2017/09/1024.11.2017.129.html>.
- Fabien Gandon. Web Science, Artificial Intelligence and Intelligence Augmentation. Seminar Dagstuhl Perspectives Workshop 18262. 10 Years of Web Science: Closing The Loop, Schloss Dagstuhl — Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2019a. URL <https://inria.hal.science/hal-01976768>.
- Fabien Gandon. The Web We Mix: Benevolent AIs for a Resilient Web. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, pages 115–116, New York, NY, USA, June 2019b. Association for Computing Machinery. ISBN 978-1-4503-6202-3. doi:10.1145/3292522.3329406. URL <https://dl.acm.org/doi/10.1145/3292522.3329406>.
- Fabien Gandon and Guus Schreiber. RDF 1.1 XML Syntax, February 2014. URL <https://www.w3.org/TR/rdf-syntax-grammar/>.
- Rita Gautschy. Swiss National Data and Service Center for the Humanities (DaSCH), October 2022. URL <https://doi.org/10.5281/zenodo.7251759>.
- Andreas Georgopoulos. CIPA's Perspectives on Cultural Heritage. In Sander Münster, Kristina Friedrichs, Florian Niebling, and Agnieszka Seidel-Grzebińska, editors, *Digital Research and Education in Architectural Heritage*, Communications in Computer and Information Science, pages 215–245, Cham, 2018. Springer International Publishing. ISBN 978-3-319-76992-9. doi:10.1007/978-3-319-76992-9_13.
- Giuseppe De Giacomo and Maurizio Lenzerini. TBox and ABox reasoning in expressive description logics. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning*, KR'96, pages 316–327, San Francisco, CA, USA, November 1996. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-421-6. URL <https://dl.acm.org/doi/10.5555/3087368.3087406>.
- Anne J. Gilliland. Setting the Stage. In Murtha Baca, editor, *Introduction to metadata*. Getty Research Institute, Los Angeles, CA, third edition edition, 2016. ISBN 978-1-60606-479-5. URL <https://www.getty.edu/publications/intrometadata/setting-the-stage/>.
- GO FAIR. FAIRification Process, 2016. URL <https://www.go-fair.org/fair-principles/fairification-process/>.
- Federico Gobbo and Marco Benini. What Can We Know of Computational Information? Measuring, Quantity, and Quality at Work in Programmable Artifacts. *Topoi*, 35(1):203–212, April 2016. ISSN 1572-8749. doi:10.1007/s11245-014-9248-5. URL <https://doi.org/10.1007/s11245-014-9248-5>.
- Daniel Gomes, João Miranda, and Miguel Costa. A Survey on Web Archiving Initiatives. In Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt, editors, *Research and Advanced Technology for Digital Libraries*, volume 6966 of *Lecture Notes in Computer Science*, pages 408–420, Berlin, Heidelberg, 2011. Springer. ISBN 978-3-642-24469-8. doi:10.1007/978-3-642-24469-8_41.
- Joshua Gomez, Kevin S. Clarke, and Anthony Vuong. IIF by the Numbers. *The Code4Lib Journal*, 2020(48), May 2020. ISSN 1940-5758. URL <https://journal.code4lib.org/articles/15217>.
- David Gough. Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2):213–228, June 2007. ISSN 0267-1522. doi:10.1080/02671520701296189. URL <https://doi.org/10.1080/02671520701296189>.
- Yash Goyal, Tejas Khot, Douglas Summers-Stay, Dhruv Batra, and Devi Parikh. Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering. *arXiv*, May 2017. doi:10.48550/arXiv.1612.00837. URL <http://arxiv.org/abs/1612.00837>.
- Stephan Graf, Birgit Huber, Aila Özvegyi, and Nicole Peduzzi. Das Fotoarchiv der Schweizerischen Gesellschaft für Volkskunde: Ein Ort der Forschung, der Begegnungen und Affekte. *Rundbrief Fotografie*, 26(1):31–40, 2019.
- Jane Greenberg. Understanding Metadata and Metadata Schemes. *Cataloging & Classification Quarterly*, 40(3-4):17–36, September 2005. ISSN 0163-9374. doi:10.1300/J104v40n03_02. URL https://doi.org/10.1300/J104v40n03_02.
- Algirdas Julien Greimas. *Sémantique structurale*. Larousse, Paris, France, 1966.

- Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, June 1993. ISSN 1042-8143. doi:10.1006/knac.1993.1008. URL <https://www.sciencedirect.com/science/article/pii/S1042814383710083>.
- Bianca Gualandi, Luca Pareschi, and Silvio Peroni. What do we mean by “data”? A proposed classification of data types in the arts and humanities. *Journal of Documentation*, 79(7):51–71, January 2022. ISSN 0022-0418. doi:10.1108/JD-07-2022-0146. URL <https://doi.org/10.1108/JD-07-2022-0146>.
- Rebecca S. Guenther. MODS: The Metadata Object Description Schema. *portal: Libraries and the Academy*, 3(1): 137–150, 2003. ISSN 1530-7131. doi:10.1353/pla.2003.0006. URL <https://muse.jhu.edu/article/38558>.
- Anaïs Guillem, Antoine Gros, and Livio De Luca. Faire parler les claveaux effondrés de la cathédrale Notre-Dame de Paris. In *Recueil des communications du 4e colloque Humanistica*, Genève, Suisse, June 2023a. URL <https://hal.science/hal-04106101>.
- Anaïs Guillem, Antoine Gros, Kevin Reby, Violette Abergel, and Livio De Luca. RCC8 for CIDOC CRM: Semantic Modeling of Mereological and Topological Spatial Relations in Notre-Dame de Paris. In Antonis Bikakis, Roberta Ferrario, Stéphane Jean, Béatrice Markhoff, Alessandro Mosca, and Marianna Nicolosi Asmundo, editors, *Proceedings of the International Workshop on Semantic Web and Ontology Design for Cultural Heritage*, volume 3540 of *CEUR Workshop Proceedings*, Athens, Greece, November 2023b. CEUR. URL <https://ceur-ws.org/Vol-3540/#paper2>.
- Piraye Hacıgüzeller, James Stuart Taylor, and Sara Perry. On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind. *Open Archaeology*, 7(1): 1709–1730, January 2021. ISSN 2300-6560. doi:10.1515/opar-2020-0220. URL <https://www.degruyter.com/document/doi/10.1515/opar-2020-0220/html>.
- Josh Hadro. Introduction to IIIF, November 2019. URL <https://youtu.be/l8kc8nH5f8I>.
- Josh Hadro. Newcomers Session: Introduction to the Community and Consortium, December 2022. URL <https://youtu.be/ojOy9fW1BRk>.
- Aric Hagberg, Pieter J. Swart, and Daniel A. Schult. Exploring network structure, dynamics, and function using NetworkX. Technical Report LA-UR-08-05495; LA-UR-08-5495, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), January 2008. URL <https://www.osti.gov/biblio/960616>.
- Mark Hahnel and Dan Valen. How to (Easily) Extend the FAIRness of Existing Repositories. *Data Intelligence*, 2(1-2): 192–198, January 2020. ISSN 2641-435X. doi:10.1162/dint.a.00041. URL <https://doi.org/10.1162/dint.a.00041>.
- Muki Haklay. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In Daniel Sui, Sarah Elwood, and Michael Goodchild, editors, *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, pages 105–122. Springer Netherlands, Dordrecht, 2013. ISBN 978-94-007-4587-2. URL https://doi.org/10.1007/978-94-007-4587-2_7.
- Michael Hammond. What is an online community? A new definition based around commitment, connection, reciprocity, interaction, agency, and consequences. *International Journal of Web Based Communities*, 13(1):118–136, January 2017. ISSN 1477-8394. doi:10.1504/IJWBC.2017.082717. URL <https://www.inderscienceonline.com/doi/abs/10.1504/IJWBC.2017.082717>.
- Donna Haraway. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3):575–599, 1988. ISSN 0046-3663. doi:10.2307/3178066. URL <https://www.jstor.org/stable/3178066>.
- Donna Haraway. Encounters with Companion Species: Entangling Dogs, Baboons, Philosophers, and Biologists. *Configurations*, 14(1):97–114, 2008. ISSN 1080-6520. doi:10.1353/con.0.0002. URL <http://muse.jhu.edu/content/crossref/journals/configurations/v014/14.1.haraway.html>.
- Donna Haraway. Tentacular Thinking: Anthropocene, Capitalocene, Chthulucene. *e-flux*, 2016(75), September 2016a. URL <https://www.e-flux.com/journal/75/67125/tentacular-thinking-anthropocene-capitalocene-chthulucene/>.
- Donna Jeanne Haraway. *The companion species manifesto: dogs, people, and significant otherness*. Number 8 in Paradigm. Prickly Paradigm Press, Chicago, 2003. ISBN 978-0-9717575-8-5.
- Donna Jeanne Haraway. *Staying with the trouble: making kin in the Chthulucene*. Experimental futures: technological lives, scientific arts, anthropological voices. Duke University Press, Durham, 2016b. ISBN 978-0-8223-7378-0.
- Juliet Hardesty and Allison Nolan. Mitigating Bias in Metadata: A Use Case Using Homosaurus Linked Data. *Information Technology and Libraries*, 40(3), September 2021. ISSN 2163-5226. doi:10.6017/ital.v40i3.13053. URL <https://ejournals.bc.edu/index.php/ital/article/view/13053>. Number: 3.

- Garrett Hardin. The Tragedy of the Commons. *Science*, 162(3859):1243–1248, December 1968. doi:10.1126/science.162.3859.1243. URL <https://www.science.org/doi/10.1126/science.162.3859.1243>.
- Patricia Harpring. Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1):67–72, April 2010. ISSN 0730-7187. doi:10.1086/adx.29.1.27949541. URL <https://www.journals.uchicago.edu/doi/abs/10.1086/adx.29.1.27949541>. Publisher: The University of Chicago Press.
- Patricia Harpring. Linking the Getty Vocabularies: The Content Perspective, Including an Update on CONA. In *2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, pages 1–8, San Francisco, CA, October 2018. IEEE. ISBN 978-986-95317-1-9. doi:10.23919/PNC.2018.8579460. URL <https://ieeexplore.ieee.org/document/8579460/>.
- Olaf Hartig, Pierre-Antoine Champin, and Gregg Kellogg. RDF 1.2 Concepts and Abstract Syntax, July 2023. URL <https://www.w3.org/TR/rdf12-concepts/>.
- Bernhard Haslhofer, Rainer Simon, Robert Sanderson, and Herbert Van de Sompel. The Open Annotation Collaboration (OAC) Model. In *2011 Workshop on Multimedia on the Web*, pages 5–9, September 2011. doi:10.1109/MMWeb.2011.21. URL <https://ieeexplore.ieee.org/abstract/document/6167821>.
- Ali Hasnain and Dietrich Rebholz-Schuhmann. Assessing FAIR Data Principles Against the 5-Star Open Data Principles. In Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, Sebastian Rudolph, Maria Maleshkova, Heiko Paulheim, Jeff Z. Pan, and Mehwish Alam, editors, *The Semantic Web: ESWC 2018 Satellite Events*, Lecture Notes in Computer Science, pages 469–477, Cham, 2018. Springer International Publishing. ISBN 978-3-319-98192-5. doi:10.1007/978-3-319-98192-5.60.
- Mark Haugaard. Foucault and Power: A Critique and Retheorization. *Critical Review*, 34(3-4):341–371, October 2022. ISSN 0891-3811. doi:10.1080/08913811.2022.2133803. URL <https://doi.org/10.1080/08913811.2022.2133803>.
- Claudie Haxaire. The Power of Ambiguity: The Nature and Efficacy of the Zamble Masks Revealed by ‘Disease Masks’ Among the Gouro People (Côte d’Ivoire). *Africa*, 79(4):543–569, November 2009. ISSN 1750-0184, 0001-9720. doi:10.3366/E0001972009001065.
- Ronald Haynes. Evolving Standards in Digital Cultural Heritage – Developing a IIF 3D Technical Specification. In Marinou Ioannides and Petros Patias, editors, *3D Research Challenges in Cultural Heritage III: Complexity and Quality in Digitisation*, Lecture Notes in Computer Science, pages 50–64. Springer International Publishing, Cham, 2023. ISBN 978-3-031-35593-6. URL https://doi.org/10.1007/978-3-031-35593-6_3.
- Ronald Haynes, Edward Silvertown, and Julie Winchester. IIF and 3D: The IIF 3D Technical Specification Group, October 2023. URL <https://ark.dasch.swiss/ark:/72163/1/0810/frGzxfL-QgW9fGt8VUuOIwG>.
- Y. He, Y. H. Ma, and X. R. Zhang. “Digital Heritage” Theory and Innovative Practice. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5:335–342, August 2017. ISSN 2194-9034. doi:10.5194/isprs-archives-XLII-2-W5-335-2017. URL <https://isprs-archives.copernicus.org/articles/XLII-2-W5/335/2017/>.
- Tom Heath and Christian Bizer. Principles of Linked Data. In Tom Heath and Christian Bizer, editors, *Linked Data: Evolving the Web into a Global Data Space*, Synthesis Lectures on Data, Semantics, and Knowledge, pages 7–27. Springer International Publishing, Cham, 2011. ISBN 978-3-031-79432-2. URL https://doi.org/10.1007/978-3-031-79432-2_2.
- Ellen Hertz, Florence Graezer Bideau, Walter Leimgruber, and Hervé Munz. *Politiques de la tradition. le patrimoine culturel immatériel*, volume 131. Presses polytechniques et universitaires romandes, Lausanne, April 2018. ISBN 978-2-88915-224-7. URL <https://edoc.unibas.ch/68569/>.
- Linda Hill, Olha Buchel, Greg Janée, and Marcia Lei Zeng. Integration of Knowledge Organization Systems into Digital Library Architectures: Position Paper for 13th ASIST SIGICR Workshop, Reconceptualizing Classification Research. *Advances in Classification Research Online*, 13(1):46–52, November 2002. ISSN 2324-9773. doi:10.7152/acro.v13i1.13835. URL <http://journals.lib.washington.edu/index.php/acro/article/view/13835>.
- Diane I. Hillmann, Rhonda Marker, and Chris Brady. Metadata Standards and Applications. *The Serials Librarian*, 54(1-2):7–21, May 2008. ISSN 0361-526X. doi:10.1080/03615260801973364. URL <https://doi.org/10.1080/03615260801973364>.
- HIMSS. Interoperability in Healthcare, August 2020. URL <https://www.himss.org/resources/interoperability-healthcare>.

- Günter Hipler, Nicolas Prongué, and René Schneider. Swissbib und linked.swissbib.ch: Leistung und Potenziale einer offenen Plattform für Schweizer Bibliotheksdaten. In Alice Keller and Susanne Uhl, editors, *Bibliotheken der Schweiz: Innovation durch Kooperation*, pages 160–172. De Gruyter, Berlin, Boston, June 2018. ISBN 978-3-11-055379-6. URL <https://doi.org/10.1515/9783110553796-008>.
- Gail M. Hodge. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Digital Library Federation, Council on Library and Information Resources, Washington, DC, USA, 2000. ISBN 978-1-933645-06-3.
- Anna Lauren Hoffmann. Terms of inclusion: Data, discourse, violence. *New Media & Society*, 23(12):3539–3556, December 2021. ISSN 1461-4448. doi:[10.1177/1461444820958725](https://doi.org/10.1177/1461444820958725). URL <https://doi.org/10.1177/1461444820958725>.
- Kerstin P. Hofmann, Susanne Grunwald, Franziska Lang, Ulrike Peter, Katja Rösler, Louise Rokohl, Stefan Schreiber, Karsten Tolle, and David Wigg-Wolf. Ding-Editionen. Vom archäologischen (Be-)Fund übers Corpus ins Netz. *e-Forschungsberichte des DAI*, pages 1–12, October 2019. ISSN 2198-7734. doi:[10.34780/s7a5-71aj](https://doi.org/10.34780/s7a5-71aj). URL <https://publications.dainst.org/journals/efb/article/view/2236>.
- Andrew Hopkins. Glossary of acronyms used in Digital Humanities. *Journal of Art Historiography*, 2022(27s), 2022. ISSN 2042-4752. doi:[10.48352/UOBXJAH.00004194](https://doi.org/10.48352/UOBXJAH.00004194). URL https://arthistoriography.files.wordpress.com/2022/11/000_glossary.pdf.
- Yumeng Hou and Sarah Kenderdine. Ontology-based knowledge representation for traditional martial arts. *Digital Scholarship in the Humanities*, pages 1–18, February 2024. ISSN 2055-7671. doi:[10.1093/llc/fqae005](https://doi.org/10.1093/llc/fqae005). URL <https://doi.org/10.1093/llc/fqae005>.
- Yumeng Hou, Sarah Kenderdine, Davide Picca, Mattia Egloff, and Alessandro Adamou. Digitizing Intangible Cultural Heritage Embodied: State of the Art. *Journal on Computing and Cultural Heritage*, 15(3):55:1–55:20, September 2022. ISSN 1556-4673. doi:[10.1145/3494837](https://doi.org/10.1145/3494837). URL <https://doi.org/10.1145/3494837>.
- Birgit Huber. Die Entdeckung der «Brünig-Napf-Reuss-Linie», October 2023. URL <https://blog.nationalmuseum.ch/2023/10/die-entdeckung-der-bruenig-napf-reuss-linie/>.
- Birgit Huber and Max Frischknecht. Digitalisierung und (De-)Konstruktion. Überlegungen zur Entwicklung eines Prototyps für die digitale Zugänglichkeit des «Atlas der Schweizerischen Volkskunde». In Sabine Eggmann and Konrad J. Kuhn, editors, *Schweizerisches Archiv für Volkskunde | Archives suisses des traditions populaires*, volume 2024/1 of *Schweizerisches Archiv für Volkskunde / Archives suisses des traditions populaires*, pages 27–52. Chronos, Zürich, August 2024a. ISBN 978-3-0340-1785-5. URL <https://doi.org/10.33057/CHRONOS.1785/27-51>.
- Birgit Huber and Max Frischknecht. Ein Prototyp entsteht: Der “Atlas der Schweizerischen Volkskunde” wird digital, April 2024b.
- Kyle Huynh, Natkeeran Ledchumykanthan, Kirsta Stapelfeldt, and Irfan Rahman. Simplifying ARK ID management for persistent access to digital objects. *The Code4Lib Journal*, 2022(54), August 2022. ISSN 1940-5758. URL <https://journal.code4lib.org/articles/16774>.
- Eero Hyvönen. Cultural Heritage on the Semantic Web. In *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*, Synthesis Lectures on Data, Semantics, and Knowledge, pages 1–11. Springer International Publishing, Cham, 2012. ISBN 978-3-031-79438-4. URL https://doi.org/10.1007/978-3-031-79438-4_1.
- Eero Hyvönen. Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web*, 11(1):187–193, January 2020. ISSN 1570-0844. doi:[10.3233/SW-190386](https://doi.org/10.3233/SW-190386). URL <https://content.iospress.com/articles/semantic-web/sw190386>.
- Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In Valentina Presutti, Eva Blomqvist, Raphael Troncy, Harald Sack, Ioannis Papadakis, and Anna Tordai, editors, *The Semantic Web: ESWC 2014 Satellite Events*, volume 8798 of *Lecture Notes in Computer Science*, pages 226–230, Cham, 2014. Springer International Publishing. ISBN 978-3-319-11955-7. doi:[10.1007/978-3-319-11955-7_24](https://doi.org/10.1007/978-3-319-11955-7_24).
- ICA Expert Group on Archival Description. Records in Context Conceptual Model 1.0, November 2023. URL <https://www.ica.org/sites/default/files/ric-cm-1.0-0.pdf>.
- Kingsley Uyi Idehen. Semantic Web Layer Cake Tweak, Explained, July 2017. URL <https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>.
- Marinos Ioannides and Robert Davies. Towards a Holistic Documentation and Wider Use of Digital Cultural Heritage. In Emmanouel Garoufallou, Fabio Sartori, Rania Siatri, and Marios Zervas, editors, *Metadata and Semantic Research*, Communications in Computer and Information Science, pages 76–88, Cham, 2019. Springer International Publishing. ISBN 978-3-030-14401-2. doi:[10.1007/978-3-030-14401-2_7](https://doi.org/10.1007/978-3-030-14401-2_7).

- Kathryn Irish and Jessica Saba. Bots are the new fraud: A post-hoc exploration of statistical methods to identify bot-generated responses in a corrupt data set. *Personality and Individual Differences*, 213, October 2023. ISSN 0191-8869. doi:10.1016/j.paid.2023.112289. URL <https://www.sciencedirect.com/science/article/pii/S019188692300212X>.
- Alan Irwin. *Citizen Science: A Study of People, Expertise, and Sustainable Development*. Routledge, New York, NY, USA, 1995. ISBN 0-415-11548-5.
- Leif Isaksen, Rainer Simon, Elton T.E. Barker, and Pau de Soto Cañamares. Pelagios and the emerging graph of ancient world data. In *Proceedings of the 2014 ACM conference on Web science, WebSci '14*, pages 197–201, New York, NY, USA, June 2014. Association for Computing Machinery. ISBN 978-1-4503-2622-3. doi:10.1145/2615569.2615693. URL <https://dl.acm.org/doi/10.1145/2615569.2615693>.
- Benjamin Obeghare Izu. The Sociocultural Significance of the Emedjo (Masquerade) Dance Among the Abraka People in Delta State, Nigeria. *E-Journal of Humanities, Arts and Social Sciences*, pages 413–423, September 2022. ISSN 2720-7722. doi:10.38159/ehass.2022394. URL <https://noyam.org/ehass2022394/>.
- Steven J. Jackson, Paul N. Edwards, Geoffrey C. Bowker, and Cory P. Knobel. Understanding infrastructure: History, heuristics and cyberinfrastructure policy. *First Monday*, 12(6), June 2007. ISSN 1396-0466. doi:10.5210/fm.v12i6.1904. URL <https://firstmonday.org/ojs/index.php/fm/article/view/1904>.
- Ian Jacobs and Norman Walsh. Architecture of the World Wide Web, Volume One, December 2004. URL <https://www.w3.org/TR/webarch/>.
- Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagna, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Philippe Rocca-Serra, Robert Pergl, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn, Mark Thompson, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Egon L. Willighagen, Peter Wittenburg, Marco Roos, Barend Mons, and Erik Schultes. FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, 2(1-2):10–29, January 2020. ISSN 2641-435X. doi:10.1162/dint.r.00024. URL <https://doi.org/10.1162/dint.r.00024>.
- Lise Jaillant and Annalina Caputo. Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data. *AI & SOCIETY*, 37(3):823–835, September 2022. ISSN 1435-5655. doi:10.1007/s00146-021-01367-x. URL <https://doi.org/10.1007/s00146-021-01367-x>.
- Lise Jaillant and Arran Rees. Applying AI to digital archives: trust, collaboration and shared professional ethics. *Digital Scholarship in the Humanities*, 38(2):571–585, June 2023. ISSN 2055-7671. doi:10.1093/llc/fqac073. URL <https://doi.org/10.1093/llc/fqac073>.
- Sikke R. Jansma, Anne M. Dijkstra, and Menno D.T. de Jong. Co-creation in support of responsible research and innovation: an analysis of three stakeholder workshops on nanotechnology for health. *Journal of Responsible Innovation*, 9(1): 28–48, January 2022. ISSN 2329-9460. doi:10.1080/23299460.2021.1994195. URL <https://doi.org/10.1080/23299460.2021.1994195>. Publisher: Routledge eprint: <https://doi.org/10.1080/23299460.2021.1994195>.
- Florian Jatón. We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science*, 47(6):811–840, December 2017. ISSN 0306-3127. doi:10.1177/0306312717730428. URL <https://doi.org/10.1177/0306312717730428>.
- Florian Jatón. Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society*, 8(1):1–15, January 2021. ISSN 2053-9517. doi:10.1177/20539517211013569. URL <https://doi.org/10.1177/20539517211013569>.
- Florian Jatón. Groundwork for AI: Enforcing a benchmark for neoantigen prediction in personalized cancer immunotherapy. *Social Studies of Science*, 53(5):787–810, October 2023. ISSN 0306-3127. doi:10.1177/03063127231192857. URL <https://doi.org/10.1177/03063127231192857>.
- Rachel Jewkes and Anne Murcott. Meanings of community. *Social Science & Medicine*, 43(4):555–563, August 1996. ISSN 0277-9536. doi:10.1016/0277-9536(95)00439-4. URL <https://www.sciencedirect.com/science/article/pii/S0277953695004394>.
- Sébastien Jodogne. On the Use of DICOM as a Storage Layer for IIIF. In *DaSCHCon 2023.2*, Basel, Switzerland, October 2023. URL <http://hdl.handle.net/2078.1/279235>.
- Simon Josefsson. The Base16, Base32, and Base64 Data Encodings. Request for Comments RFC4648, Internet Engineering Task Force, October 2006. URL <https://doi.org/10.17487/RFC4648>.

- Pauline Junginger and Marian Dörk. Categorizing Queer Identities: An Analysis of Archival Practices Using the Concept of Boundary Objects. *Journal of Feminist Scholarship*, 19(19), 2021. ISSN 21586179. doi:10.23860/jfs.2021.19.05. URL <https://digitalcommons.uri.edu/jfs/vol19/iss19/>.
- Lambert Kansy and Martin Lüthi. Going digital – Ein digitaler Lesesaal für die Staatsarchive Basel-Stadt und St.Gallen. *ABI Technik*, 42(3):144–156, August 2022. ISSN 2191-4664. doi:10.1515/abitech-2022-0028. URL <https://www.degruyter.com/document/doi/10.1515/abitech-2022-0028/html?lang=de>. Publisher: De Gruyter.
- Dick Kasperowski, Karl-Magnus Johansson, and Olof Karsvall. Temporalities and Values in an Epistemic Culture: Citizen Humanities, Local Knowledge, and AI-supported Transcription of Archives. *Archives & Manuscripts*, pages e10937–e10937, February 2024. ISSN 2164-6058. doi:10.37683/asa.v51.10937. URL <https://publications.archivists.org.au/index.php/asa/article/view/10937>.
- Brian F. G. Katz. Digitally exploring the acoustic history of Notre-Dame Cathedral. In *EuropeanaTech 2023*, The Hague, Netherlands, October 2023. URL https://youtu.be/JDcNV_X54oQ.
- Mike Kelly. JSON Hypertext Application Language. Internet Draft draft-kelly-json-hal-10, Internet Engineering Task Force, October 2023. URL <https://datatracker.ietf.org/doc/draft-kelly-json-hal-10>.
- Gérald Kembellec. Je viens d’avoir une fulgurance et relisant un texte de B. Bachimont et en le croisant avec les recherches d’@azaroth42 sur #IIIF. L’”Hypotypose sémantique” qui décrirait finement tout ou partie d’une œuvre d’art par les schémas et les autorités. Reste le navigateur à créer... <https://t.co/Aj1BzQb0x8>, April 2023. URL <https://web.archive.org/web/20230411062913/https://twitter.com/kembellec/status/1643244527124135936>.
- Robert Kiley and Tom Crane. Publishing scientific images using the IIIF, May 2016. URL <https://elifesciences.org/labs/aabe94cd/publishing-scientific-images-using-the-iiif>.
- Anne Klammt. LOUD Discourse: >Linked Open Usable Data< (LOUD)-basierte Dokumentation zur Herstellung und Entwicklung typologischer Ordnungen am Beispiel archäologischer Keramikforschung / Documentation basée sur les « linked Open Usable Data » (LOUD) pour l’établissement et le développement de classements typologiques à partir de l’exemple de la céramologie. *Deutsches Forum für Kunstgeschichte (DFK Paris): Jahresbericht / Rapport annuel*, pages 77–79, 2019. ISSN 2554-9863. doi:10.11588/dfkjb.2019.1.79405. URL <https://journals.ub.uni-heidelberg.de/index.php/dfkjb/article/view/79405>.
- Lukas Klic. *Digital publishing and research infrastructure for cultural heritage : an institutional roadmap*. PhD Thesis, Università Ca’ Foscari Venezia, July 2019. URL <http://dspace.unive.it/handle/10579/15587>.
- Craig A. Knoblock, Pedro Szekely, Eleanor Fink, Duane Degler, David Newbury, Robert Sanderson, Kate Blanch, Sara Snyder, Nilay Chheda, Nimesh Jain, Ravi Raju Krishna, Nikhila Begur Sreekanth, and Yixiang Yao. Lessons Learned in Building Linked Data for the American Art Collaborative. In Claudia d’Amato, Miriam Fernandez, Valentina Tamma, Freddy Lecue, Philippe Cudré-Mauroux, Juan Sequeda, Christoph Lange, and Jeff Heflin, editors, *The Semantic Web – ISWC 2017*, volume 10588 of *Lecture Notes in Computer Science*, pages 263–279, Cham, 2017. Springer International Publishing. ISBN 978-3-319-68204-4. doi:10.1007/978-3-319-68204-4_26.
- Holger Knublauch and Dimitris Kontokostas. Shapes Constraint Language (SHACL), July 2017. URL <https://www.w3.org/TR/shacl/>.
- Marcel Knöchelmann. Open Science in the Humanities, or: Open Humanities? *Publications*, 7(4):65, December 2019. ISSN 2304-6775. doi:10.3390/publications7040065. URL <https://www.mdpi.com/2304-6775/7/4/65>.
- Inês Koch, Cristina Ribeiro, and Carla Teixeira Lopes. ArchOnto, a CIDOC-CRM-Based Linked Data Model for the Portuguese Archives. In Mark Hall, Tanja Merčun, Thomas Risse, and Fabien Duchateau, editors, *Digital Libraries for Open Knowledge*, volume 12246 of *Lecture Notes in Computer Science*, pages 133–146. Springer International Publishing, Cham, 2020. ISBN 978-3-030-54955-8 978-3-030-54956-5. URL https://doi.org/10.1007/978-3-030-54956-5_10.
- Lukas Koster. Persistent identifiers for heritage objects. *The Code4Lib Journal*, 2020(47), February 2020. ISSN 1940-5758. URL <https://journal.code4lib.org/articles/14978>.
- Markus Krötzsch, Frantisek Simancik, and Ian Horrocks. A Description Logic Primer, June 2013. URL <http://arxiv.org/abs/1201.4089>.
- Thomas S. Kuhn. *The structure of scientific revolutions*. Number 2,2 in International encyclopedia of unified science. Univ. of Chicago Press, Chicago, IL, USA, 2. ed., enlarged, 21. print edition, 1994. ISBN 978-0-226-45804-5 978-0-226-45803-8.
- John Kunze and Emmanuelle Bermès. The ARK Identifier Scheme, May 2024. URL <https://datatracker.ietf.org/doc/html/draft-kunze-ark-39>.

- Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. The Open Archives Initiative Protocol for Metadata Harvesting - v.2.0, June 2002. URL <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
- Carl Lagoze, Herbert Van de Sompel, Michael Nelson, Simeon Warner, Robert Sanderson, and Pete Johnston. A Web-based resource model for scholarship 2.0: object reuse & exchange. *Concurrency and Computation: Practice and Experience*, 24(18):2221–2240, 2012. ISSN 1532-0634. doi:10.1002/cpe.1594. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.1594>.
- Doug Laney. 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70):1, 2001.
- Kathryn B. Laskey and Kenneth Laskey. Service oriented architecture. *WIREs Computational Statistics*, 1(1):101–105, 2009. ISSN 1939-0068. doi:10.1002/wics.8. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.8>.
- Ora Lassila and Ralph R. Swick. Resource Description Framework (RDF) Model and Syntax Specification, February 1999. URL <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- Bruno Latour. Postmodern? No, simply amodern! Steps towards an anthropology of science. *Studies in History and Philosophy of Science Part A*, 21(1):145–171, March 1990. ISSN 0039-3681. doi:10.1016/0039-3681(90)90018-4. URL <https://www.sciencedirect.com/science/article/abs/pii/0039368190900184>.
- Bruno Latour. *We have never been modern*. Harvard University Press, Cambridge, MA, USA, 1993. ISBN 978-0-674-94839-6.
- Bruno Latour. On actor-network theory: A few clarifications. *Soziale Welt*, 47(4):369–381, 1996. ISSN 0038-6073. URL <https://www.jstor.org/stable/40878163>.
- Bruno Latour. *Reassembling the social: an introduction to actor-network-theory*. Clarendon lectures in management studies. Oxford University Press, Oxford ; New York, 2005. ISBN 978-0-19-925604-4.
- Bruno Latour. *Habiter la Terre : Entretiens avec Nicolas Truong*. Éditions Les Liens qui libèrent ; Arte éditions, Paris, France, 2022. ISBN 979-10-209-1153-7.
- Bruno Latour and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, Princeton, NJ, United States, 1986. ISBN 978-0-691-02832-3. URL <https://doi.org/10.2307/j.ctt32bbxc>.
- Bruno Latour, Pablo Jensen, Tommaso Venturini, Sébastien Grauwin, and Dominique Boullier. ‘The whole is always smaller than its parts’ - a digital test of Gabriel Tarde’s monads: ‘The whole is always smaller than its parts’. *The British Journal of Sociology*, 63(4):590–615, December 2012. ISSN 00071315. doi:10.1111/j.1468-4446.2012.01428.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1468-4446.2012.01428.x>.
- Jean Lave and Etienne Wenger. *Situated learning: legitimate peripheral participation*. Learning in doing. Cambridge University Press, Cambridge, UK ; New York, USA, 1991. ISBN 978-0-521-41308-4 978-0-521-42374-8.
- Christopher A. Lee. Open Archival Information System (OAIS) Reference Model. In Marcia J. Bates and Mary Niles Maack, editors, *Encyclopedia of Library and Information Sciences, Third Edition*, pages 4020–4030. CRC Press, Boca Raton, FL, USA, 3rd edition, December 2009. ISBN 978-0-203-75763-5. URL <https://doi.org/10.1081/E-ELIS3-120044377>.
- Yong-Ju Lee and Chang-Su Kim. A Learning Ontology Method for RESTful Semantic Web Services. In *2011 IEEE International Conference on Web Services*, pages 251–258, Washington, DC, USA, July 2011. IEEE. doi:10.1109/ICWS.2011.59. URL <https://ieeexplore.ieee.org/abstract/document/6009396>.
- Susan Leigh Star, Geoffrey C. Bowker, and Laura J. Neumann. Transparency beyond the Individual Level of Scale: Convergence between Information Artifacts and Communities of Practice. In Ann P. Bishop, Nancy A. Van House, and Barbara Pfeil Buttenfield, editors, *Digital library use: social practice in design and evaluation*, Digital libraries and electronic publishing, pages 241–269. MIT Press, Cambridge, MA, USA; London, England, 2003. ISBN 978-0-262-02544-7.
- Walter Leimgruber. Was ist immaterielles Kulturerbe? *Bulletin / Schweizerische Akademie der Geistes- und Sozialwissenschaften*, 2008, H. 2:24–25, 2008. ISSN 1420-6560. URL <http://edoc.unibas.ch/dok/A5251330>.
- Walter Leimgruber. Switzerland and the UNESCO Convention on Intangible Cultural Heritage. *Journal of Folklore Research*, 47(1-2):161–196, 2010. ISSN 0737-7037. doi:10.2979/JFR.2010.47.1-2.161. URL <https://www.jstor.org/stable/10.2979/jfr.2010.47.1-2.161>.

- Walter Leimgruber, Silke Andris, and Christine Bischoff. Von der Volkskunde zur Kulturwissenschaft. *Schweizerisches Archiv für Volkskunde = Archives suisses des traditions populaires*, 107(1):77–88, 2011. doi:10.5169/SEALS-154340. URL <https://www.e-periodica.ch/digbib/view?pid=sav-001:2011:107::135>.
- Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Lawrence G. Roberts, and Stephen S. Wolff. The past and future history of the Internet. *Communications of the ACM*, 40(2):102–108, February 1997. ISSN 0001-0782, 1557-7317. doi:10.1145/253671.253741. URL <https://dl.acm.org/doi/10.1145/253671.253741>.
- Christine Lemmer-Webber and Jessica Tallon. ActivityPub, January 2018. URL <https://www.w3.org/TR/activitypub/>.
- Daniela Lucas Da Silva Lemos, Dalton Lopes Martins, Asla Medeiros E Sá, Luciana Conrado Martins, and Danielle Do Carmo. A Proposal in Creating a Semantic Repository for Digital 3D Replicas: The Case of Modernist Sculptures in Public Spaces of Rio De Janeiro. *KNOWLEDGE ORGANIZATION*, 49(3):151–171, 2022. ISSN 0943-7444. doi:10.5771/0943-7444-2022-3-151. URL <https://www.nomos-elibrary.de/index.php?doi=10.5771/0943-7444-2022-3-151>.
- Federico Lenzerini. Intangible Cultural Heritage: The Living Culture of Peoples. *European Journal of International Law*, 22(1):101–120, February 2011. ISSN 0938-5428. doi:10.1093/ejil/chr006. URL <https://doi.org/10.1093/ejil/chr006>.
- Bruce V. Lewenstein. Is Citizen Science a Remedy for Inequality? *The ANNALS of the American Academy of Political and Social Science*, 700(1):183–194, March 2022. ISSN 0002-7162. doi:10.1177/00027162221092697. URL <https://doi.org/10.1177/00027162221092697>.
- Linda C. Li, Jeremy M. Grimshaw, Camilla Nielsen, Maria Judd, Peter C. Coyte, and Ian D. Graham. Evolution of Wenger’s concept of community of practice. *Implementation Science*, 4(1):11, March 2009. ISSN 1748-5908. doi:10.1186/1748-5908-4-11. URL <https://doi.org/10.1186/1748-5908-4-11>.
- Xigao Li, Babak Amin Azad, Amir Rahmati, and Nick Nikiforakis. Good Bot, Bad Bot: Characterizing Automated Browsing Activity. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1589–1605, May 2021. doi:10.1109/SP40001.2021.00079. URL https://ieeexplore.ieee.org/abstract/document/9519384?casa_token=zIc-SHEJ9qgAAAAA:lc5sm2zXZ2_Q-XCWu2CdLkMu4ieTMLiqU3v0cpO_UV_M4xPycfp03qTc7_PlkusqmLjLt5ViVg. ISSN: 2375-1207.
- Shirley Lim and Chern Li Liew. Metadata quality and interoperability of GLAM digital images. *Aslib Proceedings*, 63(5): 484–498, September 2011. ISSN 0001-253X. doi:10.1108/00012531111164978. URL <https://www.emerald.com/insight/content/doi/10.1108/00012531111164978/full/html>.
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8693, pages 740–755. Springer International Publishing, Cham, 2014. ISBN 978-3-319-10601-4 978-3-319-10602-1. URL https://doi.org/10.1007/978-3-319-10602-1_48.
- Alexander Linden. Hype Cycle for Advanced Analytics and Data Science, 2015. Strategic Analysis Report, Gartner, Stamford, CT, USA, July 2015. URL <https://www.gartner.com/en/documents/3087721>.
- Alexander Linden and Jackie Fenn. Understanding Gartner’s hype cycles. Strategic Analysis Report R-20-1971, Gartner, Stamford, CT, USA, 2003. URL <https://web.archive.org/web/20231108132356/http://ask-force.org/web/Discourse/Linden-HypeCycle-2003.pdf>.
- Jutta Lindenthal, Hanna-Lena Meiners, and Detlev Balzer. LIDO Primer, September 2023. URL <https://lido-schema.org/documents/primer/latest/lido-primer.html>.
- Clare Llewellyn, Robert Sanderson, Kevin R. Page, Aruna Bhauggerutty, Andrew Shapland, Kelly Davis, Emmanuelle Delmas-Glass, and Bonnet Tyler. Enriching Exhibition Scholarship. In Anne Baillet, Walter Scholger, Toma Tasovac, and Georg Vogeler, editors, *Digital Humanities 2023 Book of Abstracts*, volume 2023, pages 516–517, Graz, Austria, June 2023. Alliance of Digital Humanities Organizations (ADHO). doi:10.5281/zenodo.8107930. URL <https://zenodo.org/records/8107930>.
- Giorgia Lodi, Luigi Asprino, Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini. Semantic Web for Cultural Heritage Valorisation. In Shalin Hai-Jew, editor, *Data Analytics in Digital Humanities*, Multimedia Systems and Applications, pages 3–37. Springer International Publishing, Cham, 2017. ISBN 978-3-319-54499-1. URL https://doi.org/10.1007/978-3-319-54499-1_1.

- Tolina Loulanski. Revising the Concept for Cultural Heritage: The Argument for a Functional Approach. *International Journal of Cultural Property*, 13(2):207–233, May 2006. ISSN 1465-7317, 0940-7391. doi:10.1017/S0940739106060085. URL <https://www.cambridge.org/core/journals/international-journal-of-cultural-property/article/abs/revising-the-concept-for-cultural-heritage-the-argument-for-a-functional-approach/E392CC8A3C7E7DE7102F8CF33AE389B6>.
- David Lowenthal. Natural and cultural heritage. *International Journal of Heritage Studies*, 11(1):81–92, January 2005. ISSN 1352-7258. doi:10.1080/13527250500037088. URL <https://doi.org/10.1080/13527250500037088>.
- Christopher S. Lowry and Kristine F. Stepenuck. Is Citizen Science Dead? *Environmental Science & Technology*, 55(8):4194–4196, April 2021. ISSN 0013-936X. doi:10.1021/acs.est.0c07873. URL <https://doi.org/10.1021/acs.est.0c07873>.
- Yue Ma and Pascal Hitzler. Paraconsistent Reasoning for OWL 2. In Axel Polleres and Terrance Swift, editors, *Web Reasoning and Rule Systems*, Lecture Notes in Computer Science, pages 197–211, Berlin, Heidelberg, 2009. Springer. ISBN 978-3-642-05082-4. doi:10.1007/978-3-642-05082-4_14.
- Daichi Machiya, Michiko Takahashi, and Yuka Koguchi. Improvement of metadata interoperability for promoting distribution and utilization of Japanese digital cultural resources. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pages 1–3, Virtual, March 2023. Dublin Core Metadata Initiative. doi:10.23106/dcmi.953168702. URL <https://dcpapers.dublincore.org/article/953168702>.
- Simon Mahony. Cultural Diversity and the Digital Humanities. *Fudan Journal of the Humanities and Social Sciences*, 11(3):371–388, September 2018. ISSN 2198-2600. doi:10.1007/s40647-018-0216-0. URL <https://doi.org/10.1007/s40647-018-0216-0>.
- Hugo Manguinhas, Ariadna Matas, and Maša Škrinjar. Enrichments policy for the common European data space for cultural heritage. Enrichments Policy CNECT/LUX/2021/OP/0070, Europeana, The Hague, Netherlands, August 2023. URL <https://pro.europeana.eu/post/enrichments-policy-for-the-common-european-data-space-for-cultural-heritage>.
- Lev Manovich. Cultural Analytics, Social Computing and Digital Humanities. In Mirko Tobias Schäfer and Karin van Es, editors, *The Datafied Society*, Studying Culture through Data, pages 55–68. Amsterdam University Press, 2017. ISBN 978-94-6298-136-2. URL <https://www.jstor.org/stable/j.ctt1v2xsqn.8>.
- Marian Clemens Manz, Julien Antoine Raemy, and Peter Fornaro. Recommended 3D Workflow for Digital Heritage Practices. In *Archiving Conference*, volume 20, pages 23–28, Oslo, Norway, June 2023. Society for Imaging Science and Technology. doi:10.2352/issn.2168-3204.2023.20.1.5. URL <https://library.imaging.org/archiving/articles/20/1/5>.
- Carlos Henrique Marcondes. Integrated classification schemas to interlink cultural heritage collections over the web using LOD technologies. *International Journal of Metadata, Semantics and Ontologies*, 15(3):170, 2021. ISSN 1744-2621, 1744-263X. doi:10.1504/IJMSO.2021.123040. URL <http://www.inderscience.com/link.php?id=123040>.
- Sol Martinez Demarco. Empowering women through digital skills in Argentina: A tale of two stories. *TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis*, 28(2):23–28, July 2019. ISSN 2567-8833. doi:10.14512/tatup.28.2.s23. URL <https://www.tatup.de/index.php/tatup/article/view/215>.
- Sol Martinez Demarco. From digital inclusion to IT appropriation: gendered aspects of appropriation imaginary and practices. *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft*, 15(1):72–86, April 2023. ISSN 2196-4467. doi:10.3224/gender.v15i1.06. URL <https://www.budrich-journals.de/index.php/gender/article/view/41780>.
- Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. Wonder Web Deliverable D18: Ontology Library. Ontology Infrastructure for the Semantic Web Del 18, Laboratory For Applied Ontology - ISTC-CNR, Trento, Italy, December 2003. URL <http://www.loa.istc.cnr.it/old/Papers/D18.pdf>.
- Arcangelo Massari, Silvio Peroni, Francesca Tomasi, and Ivan Heibi. Representing provenance and track changes of cultural heritage metadata in RDF: a survey of existing approaches. In *DH2023 Book of Abstracts*, Graz, Austria, June 2023. Alliance of Digital Humanities Organizations (ADHO). doi:10.5281/zenodo.8108101. URL <https://zenodo.org/record/8108101>.
- Mathieuaexhache. OAIS Functional Model, January 2021. URL [https://commons.wikimedia.org/wiki/File:OAIS_Functional_Model_\(en\).svg](https://commons.wikimedia.org/wiki/File:OAIS_Functional_Model_(en).svg).
- Fulvio Mazzocchi. Knowledge organization system (KOS). *Knowledge Organization*, 45(1):54–78, 2018. ISSN 0943-7444. doi:10.5771/0943-7444-2018-1-54. URL <https://www.nomos-elibrary.de/10.5771/0943-7444-2018-1-54/>.
- Willard McCarty. Pursuing a combinatorial habit of mind and machine. In Julianne Nyhan, Geoffrey Rockwell, Stéfan Sinclair, and Alexandra Ortolja-Baird, editors, *On Making in the Digital Humanities*, pages 251–266. UCL Press, London, UK, March 2023. ISBN 978-1-80008-420-9. URL <https://doi.org/10.14324/111.9781800084209>.

- Barbara McGillivray, Thierry Poibeau, and Pablo Ruiz Fabo. Digital Humanities and Natural Language Processing: Je t'aime... Moi non plus. *Digital Humanities Quarterly*, 014(2), June 2020. ISSN 1938-4122. URL <https://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html>.
- David W. McMillan and David M. Chavis. Sense of community: A definition and theory. *Journal of Community Psychology*, 14(1):6–23, 1986. ISSN 1520-6629. doi:10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/1520-6629%28198601%2914%3A1%3C6%3A%3AAID-JCOP2290140103%3E3.0.CO%3B2-I>.
- Ashley Rose Mehlenbacher. On Expertise: Cultivating Character, Goodwill, and Practical Wisdom. In *On Expertise*. Penn State University Press, April 2022. ISBN 978-0-271-09313-0. URL <https://doi.org/10.1515/9780271093130>.
- Emma Metcalfe Hurst. LUX: Yale Collections Discovery. *ARLIS/NA Multimedia & Technology Reviews*, 2023(4):1–4, December 2023. ISSN 2474-6673. doi:10.17613/3hy1-pv45.
- Allison Metz, Annette Boaz, and Glenn Robert. Co-creative approaches to knowledge production: what next for bridging the research to practice gap? *Evidence & Policy*, 15(3):331–337, August 2019. ISSN 1744-2648, 1744-2656. doi:10.1332/174426419X15623193264226. URL <https://bristoluniversitypressdigital.com/view/journals/evp/15/3/article-p331.xml>. Publisher: Policy Press Section: Evidence & Policy.
- Jean-Guy Meunier. Humanités numériques et modélisation scientifique. *Questions de communication*, 2017(31):19–48, September 2017. ISSN 1633-5961, 2259-8901. doi:10.4000/questionsdecommunication.11040. URL <http://journals.openedition.org/questionsdecommunication/11040>.
- Dorel Micle. Archaeological Heritage Between Natural Hazard and Anthropic Destruction: The Negative Impact of Social Non-involvement in the Protection of Archaeological Sites. *Procedia - Social and Behavioral Sciences*, 163:269–278, December 2014. ISSN 1877-0428. doi:10.1016/j.sbspro.2014.12.316. URL <https://www.sciencedirect.com/science/article/pii/S1877042814064179>.
- Sarah Middle. *Investigating Linked Data Usability for Ancient World Research*. PhD Thesis, The Open University, Milton Keynes, England, August 2022. URL <https://doi.org/10.21954/ou.ro.00014b1f>.
- Daria Mikhaylova and Daniele Metilli. Extending RiC-O to Model Historical Architectural Archives: The ITDT Ontology. *Journal on Computing and Cultural Heritage*, 16(4):67:1–67:15, August 2023. ISSN 1556-4673. doi:10.1145/3606706. URL <https://dl.acm.org/doi/10.1145/3606706>.
- Jeff Mixer. Using a Common Model: Mapping VRA Core 4.0 Into an RDF Ontology. *Journal of Library Metadata*, 14(1):1–23, January 2014. ISSN 1938-6389. doi:10.1080/19386389.2014.891890. URL <https://doi.org/10.1080/19386389.2014.891890>.
- Annemarie Mol. *The Body Multiple: Ontology in Medical Practice*. Science and Cultural Theory. Duke University Press, Durham, NC, USA, 2002. ISBN 978-0-8223-2917-6.
- Susana Morales. La apropiación de TIC: una perspectiva. In Susana Morales and M.I. Loyola, editors, *Los jóvenes y las TIC. Apropiación y uso en educación*, pages 99–120. Edición de las autoras, Córdoba, Argentina, 2009.
- Susana Morales. *Imaginación y software: aportes para la construcción del paradigma de la apropiación*. Del Gato Gris, Córdoba, Argentina, 2017. ISBN 978-987-45354-9-8. URL <http://hdl.handle.net/11086/27405>.
- Susana Morales. La apropiación de tecnologías. Ideas para un paradigma en construcción. In *Acerca de la apropiación de tecnologías. Teoría, estudios y debates*, pages 23–33. Del Gato Gris, Rada Tilly, Argentina:, 2018.
- Robbie Morrison. Redrawn slide from presentation of Ana Persic, Division of Science Policy and Capacity-Building (SC/PCB), UNESCO (France) presentation to Open Science Conference 2021, ZBW — Leibniz Information Centre for Economics, Germany., February 2021. URL <https://commons.wikimedia.org/wiki/File:Osc2021-unesco-open-science-no-gray.png>.
- Ioannis Moutsatsos. DevOps and Life Sciences Informatics: Librarians and Triple-Eye-Eff to the rescue of scientific imaging, June 2017. URL <https://imoutsatsos.blogspot.com/2017/04/librarians-and-triple-eye-eff-to-rescue.html>.
- Mr Gee. Day 2 Closing – A multitude of tools. In *EuropeanaTech 2023*, The Hague, Netherlands, October 2023. URL <https://youtu.be/pOX9CrvAG7I>.
- Dawson Munjeri. Tangible and Intangible Heritage: from difference to convergence. *Museum International*, 56(1-2):12–20, May 2004. ISSN 1350-0775, 1468-0033. doi:10.1111/j.1350-0775.2004.00453.x. URL <https://www.tandfonline.com/doi/full/10.1111/j.1350-0775.2004.00453.x>.

- Brad A. Myers and Jeffrey Stylos. Improving API usability. *Communications of the ACM*, 59(6):62–69, May 2016. ISSN 0001-0782, 1557-7317. doi:[10.1145/2896587](https://doi.org/10.1145/2896587). URL <https://dl.acm.org/doi/10.1145/2896587>.
- Katja Müller. Digitale Objekte - subjektive Materie. Zur Materialität digitalisierter Objekte in Museum und Archiv. In Hans Peter Hahn and Friedemann Neumann, editors, *Edition Kulturwissenschaft*, volume 182, pages 49–66. transcript Verlag, Bielefeld, Germany, 1 edition, October 2018. ISBN 978-3-8376-4513-2 978-3-8394-4513-6. doi:[10.14361/9783839445136-004](https://doi.org/10.14361/9783839445136-004). URL <https://www.transcript-open.de/doi/10.14361/9783839445136-004>.
- S. Münster, F. I. Apollonio, P. Bell, P. Kuroczynski, I. Di Lenardo, F. Rinaudo, and R. Tamborrino. Digital Cultural Heritage meets Digital Humanities. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W15:813–820, August 2019. ISSN 2194-9034. doi:[10.5194/isprs-archives-XLII-2-W15-813-2019](https://doi.org/10.5194/isprs-archives-XLII-2-W15-813-2019). URL <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W15/813/2019/>.
- Sander Münster, Ronja Utescher, and Selda Ulutas Aydogan. Digital topics on cultural heritage investigated: how can data-driven and data-guided methods support to identify current topics and trends in digital heritage? *Built Heritage*, 5(1):25, December 2021. ISSN 2662-6802. doi:[10.1186/s43238-021-00045-7](https://doi.org/10.1186/s43238-021-00045-7). URL <https://doi.org/10.1186/s43238-021-00045-7>.
- Satoru Nakamura. Open Data Initiatives at the University of Tokyo’s Digital Archives Construction Project, Kyushu University Collection, and Kyushu University Library. In *Open Data and Universities*, Fukuoka, Japan, January 2019. Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University. URL <https://hdl.handle.net/2324/2197530>.
- Fatemeh Nargesian, Erkang Zhu, Renée J. Miller, Ken Q. Pu, and Patricia C. Arocena. Data lake management: challenges and opportunities. *Proceedings of the VLDB Endowment*, 12(12):1986–1989, August 2019. ISSN 2150-8097. doi:[10.14778/3352063.3352116](https://doi.org/10.14778/3352063.3352116). URL <https://dl.acm.org/doi/10.14778/3352063.3352116>.
- Robert Nasarek. Virtuelle Forschungsumgebungen und Sammlungsräume: Objekte digital modellieren und miteinander vernetzen. In Udo Andraschke and Sarah Wagner, editors, *Digitale Gesellschaft*, volume 33, pages 131–146. transcript Verlag, Bielefeld, Germany, 1 edition, December 2020. ISBN 978-3-8376-5571-1 978-3-8394-5571-5. doi:[10.14361/9783839455715-010](https://doi.org/10.14361/9783839455715-010). URL <https://www.transcript-open.de/doi/10.14361/9783839455715-010>.
- Martha Nell Smith. Frozen Social Relations and Time for a Thaw: Visibility, Exclusions, and Considerations for Postcolonial Digital Archives. *Journal of Victorian Culture*, 19(3):403–410, September 2014. ISSN 1355-5502. doi:[10.1080/13555502.2014.947189](https://doi.org/10.1080/13555502.2014.947189). URL <https://doi.org/10.1080/13555502.2014.947189>.
- Michael L. Nelson and Herbert Van de Sompel. A 25 Year Retrospective on D-Lib Magazine. *arXiv*, August 2020. doi:[10.48550/arXiv.2008.11680](https://doi.org/10.48550/arXiv.2008.11680). URL <https://arxiv.org/abs/2008.11680>.
- Michael L. Nelson and Herbert Van de Sompel. D-Lib Magazine pioneered Web-based Scholarly Communication. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries, JCDL '22*, pages 1–12, New York, NY, USA, June 2022. Association for Computing Machinery. ISBN 978-1-4503-9345-4. doi:[10.1145/3529372.3530929](https://doi.org/10.1145/3529372.3530929). URL <https://dl.acm.org/doi/10.1145/3529372.3530929>.
- Peter A. Nelson. The Role of GPR in Community-Driven Compliance Archaeology with Tribal and Non-tribal Communities in Central California. *Advances in Archaeological Practice*, 9(3):215–225, August 2021. ISSN 2326-3768. doi:[10.1017/aap.2021.14](https://doi.org/10.1017/aap.2021.14). URL https://www.cambridge.org/core/product/identifier/S2326376821000140/type/journal_article.
- Clemens Neudecker. Cultural Heritage as Data: Digital Curation and Artificial Intelligence in Libraries. In Adrian Paschke, Georg Rehm, Clemens Neudecker, and Lydia Pintscher, editors, *Proceedings of the Third Conference on Digital Curation Technologies (Curator 2022)*, volume 3234 of *CEUR Workshop Proceedings*, Berlin, Germany, September 2022. CEUR. URL <https://ceur-ws.org/Vol-3234/#paper2>.
- David Newbury. LOUD: Linked Open Usable Data and linked.art. In *2018 CIDOC Conference*, pages 1–11, Heraklion, Greece, September 2018. International Council of Museums. URL https://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2021/03/CIDOC2018_paper_153.pdf.
- David Newbury. Linked Data in Production: Moving Beyond Ontologies, March 2024. URL <https://www.slideshare.net/slideshow/linked-data-in-production-moving-beyond-ontologies/266976602>.
- Erland Kolding Nielsen. Digitisation of Library Material in Europe: Problems, Obstacles and Perspectives anno 2007. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 18(1):20–27, June 2008. ISSN 2213-056X. doi:[10.18352/lq.7901](https://doi.org/10.18352/lq.7901). URL <https://liberquarterly.eu/article/view/10498>.

- NISO. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. American National Standard ANSI/NISO Z39.19-2005 (R2010), National Information Standards Organization, Baltimore, MD, USA, May 2010. URL <https://groups.niso.org/higherlogic/ws/public/download/12591/z39-19-2005r2010.pdf>.
- Bethany Nowviskie. Digital Humanities in the Anthropocene. *Digital Scholarship in the Humanities*, 30(suppl_1):i4–i15, December 2015. ISSN 2055-7671. doi:10.1093/llc/fqv015. URL <https://doi.org/10.1093/llc/fqv015>.
- Charles Kay Ogden and Ivor Armstrong Richards. *The Meaning Of Meaning*. Kegan Paul, Trench, Trubner & Co, London, 1930. URL <https://n2t.net/ark:/13960/t14n48t6b>.
- Haron Shakirat Oluwatosin. Client-Server Model. *IOSR Journal of Computer Engineering*, 16(1):67–71, 2014. ISSN 2278-8727. doi:10.9790/0661-16195771. URL <https://www.iosrjournals.org/iosr-jce/papers/Vol16-issue1/Version-9/J016195771.pdf>.
- Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1 edition, November 1990. ISBN 978-0-521-37101-8 978-0-521-40599-7 978-0-511-80776-3. URL <https://doi.org/10.1017/CBO9780511807763>.
- Simon O’Toole and James Tocknell. FAIR standards for astronomical data, March 2022. URL <http://arxiv.org/abs/2203.10710>.
- A. M. Ouksel and A. Sheth. Semantic interoperability in global information systems. *ACM SIGMOD Record*, 28(1):5–12, March 1999. ISSN 0163-5808. doi:10.1145/309844.309849. URL <https://doi.org/10.1145/309844.309849>.
- Trevor Owens. Defining Data for Humanists: Text, Artifact, Information or Evidence? *Journal of Digital Humanities*, 1(1), 2011. ISSN 2165-6673. URL <https://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/>.
- Trevor Owens. Digital Cultural Heritage and the Crowd. *Curator: The Museum Journal*, 56(1):121–130, 2013. ISSN 2151-6952. doi:10.1111/cura.12012. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/cura.12012>.
- Oxford English Dictionary. Agency, September 2023a. URL <https://doi.org/10.1093/OED/1249589150>.
- Oxford English Dictionary. Artificial Intelligence, December 2023b. URL <https://doi.org/10.1093/OED/3194963277>.
- Oxford English Dictionary. Citizen Science, September 2023c. URL <https://doi.org/10.1093/OED/6784505301>.
- Oxford English Dictionary. Community, September 2023d. URL <https://doi.org/10.1093/OED/1005093760>.
- Oxford English Dictionary. Phronesis, July 2023e. URL <https://doi.org/10.1093/OED/8835195828>.
- Oxford English Dictionary. Practice, July 2023f. URL <https://doi.org/10.1093/OED/7409132020>.
- Oxford English Dictionary. Semantics, July 2023g. URL <https://doi.org/10.1093/OED/2881032460>.
- Joseph Padfield, Charlotte Bolland, Neil Fitzgerald, Anne McLaughlin, Glen Robson, and Melissa Terras. Practical applications of IIF as a building block towards a digital National Collection. Arts and Humanities Research Council Final Report, Towards a National Collection, July 2022. URL <https://doi.org/10.5281/zenodo.6884885>.
- Thomas Padilla, Laurie Allen, Hannah Frost, Sarah Potvin, Elisabeth Russey Roke, and Stewart Varner. Always Already Computational: Collections as Data. *Collections as Data*, March 2017. doi:10.17605/OSF.IO/MX6UK. URL <https://osf.io/mx6uk/>.
- Thomas Padilla, Hannah Scates Kettler, and Yasmeen Shorish. Collections as Data: Part to Whole. Final Report, Always Already Computational - Collections as Data, November 2023a. URL <https://doi.org/10.5281/zenodo.10161976>.
- Thomas Padilla, Hannah Scates Kettler, Stewart Varner, and Yasmeen Shorish. Vancouver Statement on Collections as Data. White paper, Internet Archive Canada, Vancouver, Canada, September 2023b. URL <https://doi.org/10.5281/zenodo.8341519>.
- Kevin R. Page, Emmanuelle Delmas-Glass, David Beaudet, Samantha Norling, Lynn Rother, and Thomas Hänslı. Linked Art: Networking Digital Collections and Scholarship. In *Digital Humanities 2020 Book of Abstracts*, pages 504–509, Online, 2020. Alliance of Digital Humanities Organizations (ADHO). URL https://dh2020.adho.org/wp-content/uploads/2020/07/139_LinkedArtNetworkingDigitalCollectionsandScholarship.html.
- Cord Pagenstecher. Private Fotoalben als historische Quelle. *Zeithistorische Forschungen/Studies in Contemporary History*, 6(3):449–463, 2009. doi:10.14765/ZZF.DOK-1803. URL <https://zeitgeschichte-digital.de/doks/1803>.

- Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhanian, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. Large Language Models and Knowledge Graphs: Opportunities and Challenges, August 2023. URL <http://arxiv.org/abs/2308.06374>. arXiv:2308.06374 [cs].
- Ioannis Papadakis, Konstantinos Kyprianos, and Michalis Stefanidakis. Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine*, 21(5/6), May 2015. ISSN 1082-9873. doi:10.1045/may2015-papadakis. URL <http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>.
- Anna P. Paquet. Linked Data and Linked Open Data Projects for Libraries, Archives and Museums: Constructing Pathways to Information Discovery and Cultural Heritage Sector Collaboration. *Johns Hopkins Libraries*, December 2020. URL <http://jhir.library.jhu.edu/handle/1774.2/63875>.
- Pedro Patrón, Emilio Miguelañez, Yvan R. Petillot, Pedro Patrón, Emilio Miguelañez, and Yvan R. Petillot. Embedded Knowledge and Autonomous Planning: The Path Towards Permanent Presence of Underwater Networks. In *Autonomous Underwater Vehicles*, pages 199–224. IntechOpen, London, UK, October 2011. ISBN 978-953-307-432-0. URL <https://doi.org/10.5772/24649>.
- Urszula Pawlicka-Deger. Infrastructuring digital humanities: On relational infrastructure and global reconfiguration of the field. *Digital Scholarship in the Humanities*, 37(2):534–550, June 2022. ISSN 2055-7671. doi:10.1093/llc/fqab086. URL <https://doi.org/10.1093/llc/fqab086>.
- Nicole Peduzzi. un archivio tra intimità e rigore scientifico. *Memoriav Bulletin*, 25:38–39, October 2018. URL <https://memoriav.ch/wp-content/uploads/2018/10/Bulletin25-Web.pdf>.
- Maite Pelacho, Hannot Rodríguez, Fernando Broncano, Renata Kubus, Francisco Sanz García, Beatriz Gavete, and Antonio Lafuente. Science as a Commons: Improving the Governance of Knowledge Through Citizen Science. In Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht, editors, *The Science of Citizen Science*, pages 57–78. Springer International Publishing, Cham, 2021. ISBN 978-3-030-58277-7 978-3-030-58278-4. doi:10.1007/978-3-030-58278-4_4.
- Billy Perrigo. Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. *Time*, January 2023. ISSN 0040-781X. URL <https://time.com/6247678/openai-chatgpt-kenya-workers/>.
- Ana Persic. Building a Global Consensus on Open Science – the future UNESCO Recommendation on Open Science, February 2021. URL <https://doi.org/10.5446/53434>.
- Simon Peterhans, Loris Sauter, Florian Spiess, and Heiko Schuldt. Automatic Generation of Coherent Image Galleries in Virtual Reality. In Gianmaria Silvello, Oscar Corcho, Paolo Manghi, Giorgio Maria Di Nunzio, Koraljka Golub, Nicola Ferro, and Antonella Poggi, editors, *Linking Theory and Practice of Digital Libraries*, volume 13541 of *Lecture Notes in Computer Science*, pages 282–288, Cham, 2022. Springer International Publishing. ISBN 978-3-031-16802-4. doi:10.1007/978-3-031-16802-4_23.
- Georg Petz. Linked Open Data. Zukunftsweisende Strategien. *Bibliothek Forschung und Praxis*, June 2023. ISSN 1865-7648. doi:10.1515/bfp-2023-0006. URL <https://www.degruyter.com/document/doi/10.1515/bfp-2023-0006/html>. Publisher: De Gruyter.
- Sébastien Peyrard, John A. Kunze, and Jean-Philippe Traroni. The ARK Identifier Scheme: Lessons Learnt at the BnF and Questions Yet Unanswered. *International Conference on Dublin Core and Metadata Applications*, pages 83–94, October 2014. ISSN 1939-1366. URL <http://dcpapers.dublincore.org/pubs/article/view/3704>.
- Peter Pfrunder. *Ernst Brunner: Photographien, 1937-1962*. Schweizerische Gesellschaft für Volkskunde ; Offizin, Basel : Zürich, 2. aufl edition, 1995. ISBN 978-3-908122-56-2 978-3-907495-65-0.
- Kavita Philip. The Internet Will Be Decolonized. In Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip, editors, *Your Computer Is on Fire*, pages 91–116. The MIT Press, March 2021. ISBN 978-0-262-36077-7. URL <https://doi.org/10.7551/mitpress/10993.003.0002>.
- Robert B. Pippin. Idealism and Agency in Kant and Hegel. *The Journal of Philosophy*, 88(10):532–541, 1991. ISSN 0022-362X. URL <https://www.jstor.org/stable/2027098>. Publisher: Journal of Philosophy, Inc.
- Luba Pirgova-Morgan. Looking towards a brighter future: the potentiality of AI and digital transformations to library spaces. Digital Futures Research Report, University of Leeds Libraries, Leeds, UK, February 2023. URL <https://library.leeds.ac.uk/downloads/download/196/artificial-intelligence-ai-in-libraries>.
- Daniel V. Pitti. Encoded archival description: An introduction and overview. *New Review of Information Networking*, 5(1):61–69, January 1999. ISSN 1361-4576. doi:10.1080/13614579909516936. URL <https://doi.org/10.1080/13614579909516936>. Publisher: Routledge eprint: <https://doi.org/10.1080/13614579909516936>.

- Plato. *The Republic*. Internet Classics Archive, Cambridge, MA, USA, 360 B.C.E. URL <https://classics.mit.edu/Plato/republic.html>.
- Adrian Pohl. How We Built a Spatial Subject Classification Based on Wikidata. *The Code4Lib Journal*, 2021(51), June 2021. ISSN 1940-5758. URL <https://journal.code4lib.org/articles/15875>.
- Adrian Pohl and Adrian Ostrowski. Presenting the SkoHub Vocab Prototypes, September 2019. URL <https://blog.skohub.io/2019-09-27-skohub-vocabs/>.
- Adrian Pohl, Fabian Steeg, and Pascal Christoph. lobid – Dateninfrastruktur für Bibliotheken. *Informationspraxis*, 4 (1), December 2018. ISSN 2297-3249. doi:10.11588/ip.2018.1.52445. URL <https://journals.ub.uni-heidelberg.de/index.php/ip/article/view/52445>. Number: 1.
- Cristina Portalés, João M. F. Rodrigues, Alexandra Rodrigues Gonçalves, Ester Alba, and Jorge Sebastián. Digital Cultural Heritage. *Multimodal Technologies and Interaction*, 2(3):58, September 2018. ISSN 2414-4088. doi:10.3390/mti2030058. URL <https://www.mdpi.com/2414-4088/2/3/58>.
- Vassilis Pouloupoulos and Manolis Wallace. Digital Technologies and the Role of Data in Cultural Heritage: The Past, the Present, and the Future. *Big Data and Cognitive Computing*, 6(3):181–199, September 2022. ISSN 2504-2289. doi:10.3390/bdcc6030073. URL <https://www.mdpi.com/2504-2289/6/3/73>.
- Gautier Poupeau. Réflexions et questions autour du Web sémantique, October 2018. URL <https://www.lespetitescases.net/reflexions-et-questions-autour-du-web-semantique>.
- Nicolas Prongué and René Schneider. Modelling library linked data in practice: three Swiss case studies. In Franjo Pehar, Christian Schlögl, and Christian Wolff, editors, *Re:inventing Information Science in the Networked Society: Proceedings of the 14th International Symposium on Information Science*, volume 66 of *Schriften zur Informationswissenschaft*, pages 118–128, Glückstadt, Germany, 2015. Verlag Werner Hülsbusch. ISBN 978-3-86488-081-0. URL <https://arodes.hes-so.ch/record/988>.
- Jon Purday. Think culture: Europeana.eu from concept to construction. *Bibliothek Forschung und Praxis*, 33(2):170–180, September 2009. ISSN 1865-7648. doi:10.1515/bfup.2009.018. URL <https://www.degruyter.com/document/doi/10.1515/bfup.2009.018/html>.
- Sheila Rabun. Scoping the “IIIF universe”: First Steps to Discovery. Research Proposal, Information School, University of Washington, Seattle, WA, December 2016.
- Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust Speech Recognition via Large-Scale Weak Supervision, December 2022. URL <http://arxiv.org/abs/2212.04356>. arXiv:2212.04356 [cs, eess].
- Julien A. Raemy and René Martin Schneider. Towards Trusted Identities for Swiss Researchers and their Data. *International Journal of Digital Curation*, 14(1):303–314, September 2019a. ISSN 1746-8256. doi:10.2218/ijdc.v14i1.596. URL <http://www.ijdc.net/article/view/596>.
- Julien Antoine Raemy. *The International Image Interoperability Framework (IIIF): raising awareness of the user benefits for scholarly editions*. Bachelor’s thesis, HES-SO University of Applied Sciences and Arts, Haute école de gestion de Genève, Geneva, Switzerland, July 2017. URL <https://sonar.ch/hesso/documents/314853>.
- Julien Antoine Raemy. Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions. Master’s thesis, HES-SO University of Applied Sciences and Arts, Haute école de gestion de Genève, Geneva, Switzerland, August 2020. URL <https://sonar.ch/hesso/documents/315109>.
- Julien Antoine Raemy. Applying Effective Data Modelling Approaches for the Creation of a Participatory Archive Platform. In Yumeng Hou, editor, *Human Factors in Digital Humanities*, pages 1–5, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, December 2021. Institut des humanités digitales. doi:10.5451/unibas-ep87517. URL <https://infoscience.epfl.ch/handle/20.500.14299/184412>.
- Julien Antoine Raemy. Améliorer la valorisation des données du patrimoine culturel grâce au Linked Open Usable Data (LOUD). In Nicolas Lasolle, Olivier Bruneau, and Jean Lieber, editors, *Actes des journées humanités numériques et Web sémantique*, pages 132–149, Nancy, France, 2022a. Les Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les Technologies (AHP-PreST); Laboratoire lorrain de recherche en informatique et ses applications (LORIA). doi:10.5451/unibas-ep89725.
- Julien Antoine Raemy. Back and Forth from Boundary Objects to IIIF Resources: The Recipes of a Community-driven Initiative Specifying Standards. In *DARIAH-CH Study Day 2022 Posters*, Mendrisio, Switzerland, October 2022b. DARIAH-CH. doi:10.5281/zenodo.7015256. URL <https://zenodo.org/record/7015256>.

- Julien Antoine Raemy. Characterising the IIF and Linked Art Communities: survey report. Report, University of Basel, Basel, Switzerland, July 2023a. URL <https://hal.science/hal-04162572>. doi:10.5451/unibas-ep95340.
- Julien Antoine Raemy. DaSCHCon 2023.2 Wrap-up, October 2023b. URL https://ark.dasch.swiss/ark:/72163/1/0810/J3fJY80lSsWkooqot76n8w_.
- Julien Antoine Raemy. Dear Mastodon, #IIF and #LinkedArt enthusiasts or newcomers (please boost!), Are you involved or have you already been in contact with the... , March 2023c. URL <https://hcommons.social/@julsraemy/110077871549376350>.
- Julien Antoine Raemy. Linked Open Usable Data for Cultural Heritage, February 2023d. URL <https://julsraemy.ch/prezi/loud-midway.html>.
- Julien Antoine Raemy. Pseudonymised Dataset of the Characterising the IIF and Linked Art communities survey, July 2023e. URL <https://doi.org/10.5281/zenodo.8143828>.
- Julien Antoine Raemy. Interlinking Cultural Heritage Data with Community-driven Principles and Standards, May 2024a. URL <https://julsraemy.ch/prezi/pia-ringvorlesung-2024.html>.
- Julien Antoine Raemy. Some notes from the 2024 IIF Conference held in Los Angeles, June 2024b. URL <https://julsraemy.ch/posts/2024/06/26/iif-conference-la/>.
- Julien Antoine Raemy and Adrian Demleitner. Implementation of the IIF Presentation API 3.0 based on Software Support: Use Case of an Incremental IIF Deployment within a Citizen Science Project, 2023. URL <https://arxiv.org/abs/2312.11116>.
- Julien Antoine Raemy and Rita Gautschy. Élaboration d'un processus pour les images 3D reposant sur IIF. In Simon Gabay, Elina Leblanc, Nicola Carboni, Radu Suciuc, Gabriella Lini, Marie Barras, and Fatiha Idmhand, editors, *Recueil des communications du 4e colloque Humanistica*, pages 1–3, Genève, Suisse, June 2023. Humanistica, l'association francophone des humanités numériques/digitales. doi:10.5451/unibas-ep94862. URL <https://hal.science/hal-04099506>.
- Julien Antoine Raemy and Robert Sanderson. Analysis of the Usability of Automatically Enriched Cultural Heritage Data, September 2023. URL <http://arxiv.org/abs/2309.16635>.
- Julien Antoine Raemy and Robert Sanderson. Analysis of the Usability of Automatically Enriched Cultural Heritage Data. In Fernando Moral-Andrés, Elena Merino-Gómez, and Pedro Reviriego, editors, *Decoding Cultural Heritage: A Critical Dissection and Taxonomy of Human Creativity through Digital Tools*, pages 69–93. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-57675-1. URL https://doi.org/10.1007/978-3-031-57675-1_4.
- Julien Antoine Raemy and René Schneider. Suggested measures for deploying IIF in Swiss cultural heritage institutions. White paper, Haute école de gestion de Genève, Geneva, Switzerland, May 2019b. URL <https://doi.org/10.5281/zenodo.2640416>.
- Julien Antoine Raemy, Tanya Gray, Alwyn Collinson, and Kevin R. Page. Enabling Participatory Data Perspectives for Image Archives through a Linked Art Workflow (Poster). In *Digital Humanities 2023 Posters*, Graz, Austria, July 2023a. Alliance of Digital Humanities Organizations (ADHO). doi:10.5281/zenodo.7878358. URL <https://zenodo.org/records/7878358>.
- Julien Antoine Raemy, Tanya Gray, Alwyn Collinson, and Kevin R. Page. Enabling Participatory Data Perspectives for Image Archives through a Linked Art Workflow. In Anne Baillet, Walter Scholger, Toma Tasovac, and Georg Vogeler, editors, *Digital Humanities 2023 Book of Abstracts*, volume 2023, pages 515–516, Graz, Austria, June 2023b. Alliance of Digital Humanities Organizations (ADHO). doi:10.5451/unibas-ep95099. URL <https://edoc.unibas.ch/95099/>.
- Sudha Ram and Jun Liu. A New Perspective on Semantics of Data Provenance. In *Proceedings of the First International Conference on Semantic Web in Provenance Management*, volume 526, pages 35–40, Aachen, Germany, October 2009. CEUR-WS.org. URL https://ceur-ws.org/Vol-526/InvitedPaper_1.pdf.
- Michel Rautenberg. L'émergence patrimoniale de l'ethnologie : entre mémoire et politiques publiques. In *Patrimoine et modernité*, Collection Chemins de la mémoire, pages 279–289. L'Harmattan, Paris, France, 1998. ISBN 978-2-7384-6287-9.
- Zahid Raza, Khalid Mahmood, and Nosheen Fatima Warraich. Application of linked data technologies in digital libraries: a review of literature. *Library Hi Tech News*, 36(3):9–12, January 2019. ISSN 0741-9058. doi:10.1108/LHTN-10-2018-0067. URL <https://doi.org/10.1108/LHTN-10-2018-0067>. Publisher: Emerald Publishing Limited.
- María Aránzazu Respaldiza Hidalgo, Monica Wachowicz, and Antonio Vázquez Hoehne. Metadata Visualization of Cultural Heritage Information within a Collaborative Environment. In *Proceedings of XXXIIIrd International CIPA Symposium*, Prague, Czech Republic, 2011. URL <https://oa.upm.es/11636/>.

- David Ribes and Charlotte P. Lee. Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. *Computer Supported Cooperative Work (CSCW)*, 19(3):231–244, August 2010. ISSN 1573-7551. doi:10.1007/s10606-010-9120-0. URL <https://doi.org/10.1007/s10606-010-9120-0>.
- Mia Ridge, editor. *Crowdsourcing our cultural heritage*. Digital research in the arts and humanities. Routledge, London New York, first issued in paperback edition, 2017. ISBN 978-1-138-70617-0.
- Mia Ridge. Enriching lives: connecting communities and culture with the help of machines. In *EuropeanaTech 2023*, The Hague, Netherlands, October 2023. doi:10.5281/zenodo.8429858. URL <https://zenodo.org/record/8429858>.
- Mia Ridge, Samantha Blickhan, Meghan Ferriter, Austin Mast, Ben Brumfield, Brendon Wilkins, Daria Cybulska, Denise Burgher, Jim Casey, Kurt Luther, Michael Haley Goldman, Nick White, Pip Willcox, Sara Carlstead Brumfield, Sonya J. Coleman, and Ylva Berglund Prytz. 12. Connecting with communities. In *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage - community review version*. British Library, London, UK, 1st edition edition, April 2021a. URL <https://doi.org/10.21428/a5d7554f.1b80974b>.
- Mia Ridge, Samantha Blickhan, Meghan Ferriter, Austin Mast, Ben Brumfield, Brendon Wilkins, Daria Cybulska, Denise Burgher, Jim Casey, Kurt Luther, Michael Haley Goldman, Nick White, Pip Willcox, Sara Carlstead Brumfield, Sonya J. Coleman, and Ylva Berglund Prytz. 5. Designing cultural heritage crowdsourcing projects. In *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage - community review version*. British Library, London, UK, 1st edition edition, April 2021b. URL <https://doi.org/10.21428/a5d7554f.1b80974b>.
- Mia Ridge, Meghan Ferriter, and Samantha Blickhan. Recommendations, Challenges and Opportunities for the Future of Crowdsourcing in Cultural Heritage: a White Paper. Technical report, British Library, London, UK, July 2023. URL <https://doi.org/10.21428/a5d7554f.2a84f94b>.
- Jenn Riley. Seeing Standards: A Visualization of the Metadata Universe, 2009. URL <https://jennriley.com/metadatamap/>.
- Jenn Riley. *Understanding metadata. What is metadata, and what is it for?* NISO Primer Series. National Information Standards Organization (NISO), Baltimore, MD, USA, 2017. ISBN 978-1-937522-72-8.
- Pat Riva, Patrick Le Boeuf, and Maja Žumer. IFLA Library Reference Model: A Conceptual Model for Bibliographic Information. IFLA-LRM, International Federation of Library Associations and Institutions, The Hague, Netherlands, December 2017. URL <https://repository.ifa.org/handle/123456789/40>.
- Rebecca C. Roberts, Junaid Abdul Jabbar, Huw Jones, Hector A. Orenge, Marco Madella, and Cameron A. Petrie. Paper and pixels: historic maps as a multifaceted resource. *Abstracts of the ICA*, 3:1–2, December 2021. doi:10.5194/ica-abs-3-250-2021. URL <https://ica-abs.copernicus.org/articles/3/250/2021/>.
- Régis Robineau. Introduction aux protocoles IIIF, January 2019. URL <https://doi.org/10.5281/zenodo.3760306>.
- Cathy J. Robinson, Taryn Kong, Rebecca Coates, Ian Watson, Chris Stokes, Petina Pert, Andrew McConnell, and Caron Chen. Caring for Indigenous Data to Evaluate the Benefits of Indigenous Environmental Programs. *Environmental Management*, 68(2):160–169, August 2021. ISSN 1432-1009. doi:10.1007/s00267-021-01485-8.
- Glen Robson. BL Training - IIIF Image API, November 2021. URL <https://youtu.be/28tkpIdnh4g>.
- Glen Robson, Stefano Cossu, Ruven Pillay, and Michael D. Smith. Evaluating HTJ2K as a Drop-In Replacement for JPEG2000 with IIIF. *The Code4Lib Journal*, 2023(57), August 2023. ISSN 1940-5758. URL <https://journal.code4lib.org/articles/17596>.
- Dario Rodighiero. *Mapping Affinities: Democratizing Data Visualization*. Métis Presses, 2021. ISBN 978-2-940563-99-9. URL <https://dash.harvard.edu/handle/1/37368046>.
- Elizabeth Russey Roke and Ruth Kitchin Tillman. Pragmatic Principles for Archival Linked Data. *The American Archivist*, 85(1):173–201, July 2022. ISSN 0360-9081. doi:10.17723/2327-9702-85.1.173. URL <https://doi.org/10.17723/2327-9702-85.1.173>.
- C. Annemieke Romein, Max Kemman, Julie M. Birkholz, James Baker, Michel De Gruijter, Albert Meroño-Peñuela, Thorsten Ries, Ruben Ros, and Stefania Scagliola. State of the Field: Digital History. *History*, 105(365):291–312, 2020. ISSN 1468-229X. doi:10.1111/1468-229X.12969. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-229X.12969>.
- Lukas Rosenthaler. Long-Term Archiving of Digital Assets in Arts and Humanities: IIIF and Beyond, October 2023. URL <https://ark.dasch.swiss/ark:/72163/1/0810/hEOJNSNeQi6GzuumLps6Lgm>.

- Lukas Rosenthaler and Peter Fornaro. The "International Image Interoperability Framework" and its Implication to Preservation. *Archiving Conference*, 13(1):95–99, April 2016. ISSN 2161-8798. doi:10.2352/issn.2168-3204.2016.1.0.95. URL <https://library.imaging.org/archiving/articles/13/1/art00021>.
- Lukas Rosenthaler, Peter Fornaro, and Claire Clivaz. DASCH: Data and Service Center for the Humanities. *Digital Scholarship in the Humanities*, 30(Supl.1):i43–i49, October 2015. ISSN 2055-7671, 2055-768X. doi:10.1093/llc/fqv051. URL <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqv051>.
- Lukas Rosenthaler, Peter Fornaro, Andrea Bianco, and Benjamin Geer. Simple Image Presentation Interface (SIPI): an IIIF-based Image-Server. *Archiving Conference*, 14(1):28–33, May 2017. ISSN 2161-8798. doi:10.2352/issn.2168-3204.2017.1.0.28. URL <https://library.imaging.org/archiving/articles/14/1/art00008>.
- Lozana Rossenova. IIIF for 3D – making web interoperability multi-dimensional, November 2023. URL <https://pro.europeana.eu/post/iiif-for-3d-making-web-interoperability-multi-dimensional>.
- Lozana Rossenova and Karen Di Franco. Iterative Pasts and Linked Futures: A Feminist Approach to Modeling Data in Archives and Collections of Artists' Publishing. *Perspectives on Data*, 2, December 2022. doi:10.53269/9780865593152/05. URL <https://www.artic.edu/digital-publications/37/perspectives-on-data/28/iterative-pasts-and-linked-futures>.
- Lozana Rossenova, Paul Duchesne, and Ina Blümel. Wikidata and Wikibase as complementary research data management services for cultural heritage data. In Lucie-Aimée Kaffee, Simon Razniewski, Gabriel Amaral, and Kholoud Saad Alghamdi, editors, *Proceedings of the 3rd Wikidata Workshop 2022*, volume 3262 of *CEUR Workshop Proceedings*, Virtual Event, Hangzhou, October 2022. CEUR. URL <https://ceur-ws.org/Vol-3262/#paper15>.
- Luca Rossetto, Ivan Giangreco, Claudiu Tanase, and Heiko Schuldt. vitivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1183–1186, Amsterdam The Netherlands, October 2016. ACM. ISBN 978-1-4503-3603-1. doi:10.1145/2964284.2973797. URL <https://dl.acm.org/doi/10.1145/2964284.2973797>.
- Maria Rousi, Georgios Meditskos, Stefanos Vrochidis, and Ioannis Kompatsiaris. Supporting the Discovery and Reuse of Digital Content in Creative Industries using Linked Data. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pages 100–103, Laguna Hills, CA, USA, January 2021. IEEE. ISBN 978-1-72818-899-7. doi:10.1109/ICSC50631.2021.00025. URL <https://ieeexplore.ieee.org/abstract/document/9364378>.
- Stuart J. Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Prentice Hall series in artificial intelligence. Prentice Hall, Upper Saddle River, 3rd ed edition, 2010. ISBN 978-0-13-604259-4.
- SAA Dictionary. Taxonomy, 2023. URL <https://dictionary.archivists.org/entry/taxonomy.html>.
- Arjun Sabharwal. 2 - Archives and special collections in the digital humanities. In Arjun Sabharwal, editor, *Digital Curation in the Digital Humanities*, pages 27–47. Chandos Publishing, January 2015. ISBN 978-0-08-100143-1. doi:10.1016/B978-0-08-100143-1.00002-7. URL <https://www.sciencedirect.com/science/article/pii/B9780081001431000027>.
- Eveline R. Sacramento, Susana Sardo, Ana Flávia Miguel, Hélder J. M. Caixinha, and Cristina Cortés. Considering the Question of the interoperability. *Páginas a&b Arquivos & Bibliotecas*, 3(18):120–133, 2022. ISSN 08735670, 21836671. doi:10.21747/21836671/pag18a7. URL <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/12533/11786>.
- Barna Saha and Divesh Srivastava. Data quality: The other face of Big Data. In *2014 IEEE 30th International Conference on Data Engineering*, pages 1294–1297, Chicago, IL, USA, March 2014. IEEE. ISBN 978-1-4799-2555-1. doi:10.1109/ICDE.2014.6816764. URL <https://ieeexplore.ieee.org/document/6816764>. ISSN: 2375-026X.
- Robert Sanderson. *Linking Past and Future: An Application of Dynamic HTML for Medieval Manuscript Editions*. PhD Thesis, University of Liverpool, Liverpool, UK, October 2003.
- Robert Sanderson. RDF: Resource Description Failures and Linked Data Letdowns. *Journal of Digital Humanities*, 2(3), 2013. ISSN 2165-6673. URL <https://journalofdigitalhumanities.org/2-3/rdf-resource-description-failures-and-linked-data-letdowns/>.
- Robert Sanderson. Linked Data Best Practices and BibFrame, December 2015. URL <https://www.slideshare.net/azaro42/linked-data-best-practices-and-bibframe>.
- Robert Sanderson. Community Challenges for Practical Linked Open Data, December 2016. URL <https://www.slideshare.net/azaro42/community-challenges-for-practical-linked-open-data-linked-pasts-keynote>.
- Robert Sanderson. Shout it Out: LOUD, May 2018. URL <https://www.slideshare.net/Europeana/shout-it-out-loud-by-rob-sanderson-europeantech-conference-2018>.

- Robert Sanderson. Introduction to Linked Art Model, November 2019a. URL <https://www.slideshare.net/azaroth42/introduction-to-linked-art-model>.
- Robert Sanderson. Keynote: Standards and Communities: Connected People, Consistent Data, Usable Applications. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, page 28, Urbana-Champaign, Illinois, USA, June 2019b. IEEE. doi:10.1109/JCDL.2019.00009.
- Robert Sanderson. Cultural Heritage Research Data Ecosystem, March 2020a. URL <https://www.slideshare.net/azaroth42/sanderson-cni-2020-keynote-cultural-heritage-research-data-ecosystem>.
- Robert Sanderson. The Importance of being LOUD, February 2020b. URL <https://www.slideshare.net/azaroth42/the-importance-of-being-loud>.
- Robert Sanderson. Linked Art: Sustainable Cultural Knowledge through Linked Open Usable Data, July 2021. URL <https://www.slideshare.net/azaroth42/introduction-to-linked-art-model>.
- Robert Sanderson. Linking and enriching data & user experience in LUX: Yale's cross-collection discovery platform. In *EuropeanaTech 2023*, The Hague, Netherlands, October 2023a. URL <https://youtu.be/kkVXEjt8F7c>.
- Robert Sanderson. Semantic Completeness vs Data Usability in Cultural Heritage Collections: Lessons Learnt from LUX, Linked Art and IIF., June 2023b. URL <https://youtu.be/87Igs5t82U>.
- Robert Sanderson. Understanding Linked Art, October 2023c. URL <https://www.slideshare.net/azaroth42/understanding-linked-art>.
- Robert Sanderson. Implementing Linked Art in a Multi-Modal Database for Cross-Collection Discovery. *Open Library of Humanities*, 10(2), July 2024a. ISSN 2056-6700. doi:10.16995/olh.15407. URL <https://olh.openlibhums.org/article/id/15407/>.
- Robert Sanderson. Linked Data Enlightenment: Lessons Learned through LUX, March 2024b. URL <https://doi.org/10.5281/zenodo.10759740>.
- Robert Sanderson and Benjamin Albritton. Shared Canvas Data Model 1.0, February 2013. URL <https://iif.io/api/model/shared-canvas/1.0/>.
- Robert Sanderson and Heather White. LUX - Cross Collections Cultural Heritage at Yale, February 2023. URL <https://www.slideshare.net/azaroth42/lux-cross-collections-cultural-heritage-at-yale>.
- Robert Sanderson, Benjamin Albritton, Rafael Schwemmer, and Herbert Van de Sompel. SharedCanvas: a collaborative model for medieval manuscript layout dissemination. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, JCDL '11, pages 175–184, New York, NY, USA, June 2011. Association for Computing Machinery. ISBN 978-1-4503-0744-4. doi:10.1145/1998076.1998111. URL <https://dl.acm.org/doi/10.1145/1998076.1998111>.
- Robert Sanderson, Paolo Ciccarese, and Herbert Van de Sompel. Designing the W3C open annotation data model. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 366–375, New York, NY, USA, May 2013. Association for Computing Machinery. ISBN 978-1-4503-1889-1. doi:10.1145/2464464.2464474. URL <https://doi.org/10.1145/2464464.2464474>.
- Robert Sanderson, Paolo Ciccarese, and Benjamin Young. Web Annotation Data Model, February 2017. URL <https://www.w3.org/TR/annotation-model/>.
- Andrea Scharnhorst, Pascal Flohr, Vyacheslav Tykhonov, Jerry De Vries, Hella Hollander, Jetze Touber, Wim Hugo, Richard Smiraglia, Yann Le Franc, Ronald Siebes, and Enno Meijers. Knowledge Organisation Systems in the Humanities—Semantic Interoperability in Practice. *Proceedings of the Association for Information Science and Technology*, 60(1):1113–1115, October 2023. ISSN 2373-9231, 2373-9231. doi:10.1002/pra2.962. URL <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/pra2.962>.
- Flint Schier. Hume and the Aesthetics of Agency. *Proceedings of the Aristotelian Society*, 87:121–135, 1986. ISSN 0066-7374. URL <https://www.jstor.org/stable/4545059>. Publisher: [Aristotelian Society, Wiley].
- Sophie C. Schmidt and Florian Thiery. SPARQLing Ogham Stones: New Options for Analyzing Analog Editions by Digitization in Wikidata. In Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera, and Joris van Zundert, editors, *Graph Technologies in the Humanities - Proceedings 2020*, volume 3110 of *CEUR Workshop Proceedings*, page 211, Vienna, Austria, February 2020. CEUR. URL <https://ceur-ws.org/Vol-3110/#paper11>.

- Sophie C. Schmidt, Florian Thiery, and Martina Trognitz. Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata. *Digital*, 2(3):333–364, September 2022. ISSN 2673-6470. doi:10.3390/digital2030019. URL <https://www.mdpi.com/2673-6470/2/3/19>.
- Friedemann Schmoll. Richard Weiss : Skizzen zum internationalen Wirken des Schweizer Volkskundlers. *Schweizerisches Archiv für Volkskunde/ Archives suisses des traditions populaires*, 2009(105):15–32, 2009a. doi:10.5169/SEALS-118266. URL <https://www.e-periodica.ch/digbib/view?pid=sav-001:2009:105::26>.
- Friedemann Schmoll. *Die Vermessung der Kultur: der "Atlas der deutschen Volkskunde" und die Deutsche Forschungsgemeinschaft, 1928-1980*. Number Bd. 5 in Studien zur Geschichte der Deutschen Forschungsgemeinschaft. Steiner, Stuttgart, 2009b. ISBN 978-3-515-09298-2.
- Erik Schultes and Peter Wittenburg. FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure. In Yannis Manolopoulos and Sergey Stupnikov, editors, *Data Analytics and Management in Data Intensive Domains*, Communications in Computer and Information Science, pages 3–16, Cham, 2019. Springer International Publishing. ISBN 978-3-030-23584-0. doi:10.1007/978-3-030-23584-0-1.
- Tobias Schweizer and Benjamin Geer. Gravsearch: Transforming SPARQL to query humanities data. *Semantic Web*, 12(2):379–400, January 2021. ISSN 1570-0844. doi:10.3233/SW-200386. URL <https://dl.acm.org/doi/abs/10.3233/SW-200386>.
- Christof Schöch. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2(3), November 2013. ISSN 2165-6673. URL <https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- Kymberley R. Scroggie, Klementine J. Burrell-Sander, Peter J. Rutledge, and Alice Motion. GitHub as an open electronic laboratory notebook for real-time sharing of knowledge and collaboration. *Digital Discovery*, 2(4), July 2023. ISSN 2635-098X. doi:10.1039/D3DD00032J. URL <https://pubs.rsc.org/en/content/articlelanding/2023/dd/d3dd00032j>.
- Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, pages 59–68, New York, NY, USA, January 2019. Association for Computing Machinery. ISBN 978-1-4503-6125-5. doi:10.1145/3287560.3287598. URL <https://dl.acm.org/doi/10.1145/3287560.3287598>.
- Concetta Semeraro, Mario Lezoche, Hervé Panetto, and Michele Dassisti. Digital twin paradigm: A systematic literature review. *Computers in Industry*, 130:103469, September 2021. ISSN 0166-3615. doi:10.1016/j.compind.2021.103469. URL <https://www.sciencedirect.com/science/article/pii/S0166361521000762>.
- Michel Serres. *Le parasite*. Pluriel, Paris, France, 2014. ISBN 978-2-8185-0430-7.
- SGV, editor. *125 Jahre Alltagskultur: das Jubiläumsmagazin der Schweizerischen Gesellschaft für Volkskunde (SGV): 1896-2021*. Schweizerische Gesellschaft für Volkskunde, Basel, 2021. ISBN 978-3-908123-05-7.
- Guodong Shao and Deogratias Kibira. DIGITAL MANUFACTURING: REQUIREMENTS AND CHALLENGES FOR IMPLEMENTING DIGITAL SURROGATES. In *2018 Winter Simulation Conference (WSC)*, pages 1226–1237, December 2018. doi:10.1109/WSC.2018.8632242. ISSN: 1558-4305.
- Robert Sheldon. What is internationalization (i18n)?, June 2023. URL <https://www.techtarget.com/whatis/definition/internationalization-I18N>.
- Elizabeth Shepherd and Charlotte Smith. The Application of ISAD(G) to the Description of Archival Datasets. *Journal of the Society of Archivists*, 21(1):55–86, April 2000. ISSN 0037-9816. doi:10.1080/00379810050006911. URL <https://doi.org/10.1080/00379810050006911>.
- Ray Siemens. Communities of practice, the methodological commons, and digital self-determination in the Humanities. *Digital Studies / Le champ numérique*, 5(3), June 2016. ISSN 1918-3666. doi:10.16995/dscn.31. URL <https://www.digitalstudies.org/article/id/7291/>.
- Anna Simandiraki-Grimshaw. What is a museum object according to a museum database?, September 2023. URL <https://www.tetrarchs.org/index.php/2023/09/19/what-is-a-museum-object-according-to-a-museum-database/>.
- Agnès Simon, Romain Wenz, Vincent Michel, and Adrien Di Mascio. Publishing Bibliographic Records on the Web of Data: Opportunities for the BnF (French National Library). In Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, *The Semantic Web: Semantics and Big Data*, Lecture Notes in Computer Science, pages 563–577, Berlin, Heidelberg, 2013. Springer. ISBN 978-3-642-38288-8. doi:10.1007/978-3-642-38288-8_38.

- Robert Simpson, Kevin R. Page, and David De Roure. Zooniverse: observing the world's largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1049–1054, Seoul Korea, April 2014. ACM. ISBN 978-1-4503-2745-9. doi:10.1145/2567948.2579215. URL <https://dl.acm.org/doi/10.1145/2567948.2579215>.
- Marcus Smith. Linked open data and aggregation infrastructure in the cultural heritage sector: A case study of SOCH, a linked data aggregator for Swedish open cultural heritage. In *Information and Knowledge Organisation in Digital Humanities*, pages 64–85. Routledge, Thames, Oxfordshire, UK, 2021. ISBN 978-1-00-313181-6. URL <https://doi.org/10.4324/9781003131816-4>.
- James M. Snell and Evan Prodromou. Activity Streams 2.0, May 2017. URL <https://www.w3.org/TR/activitystreams-core/>.
- Stuart Snyderman, Robert Sanderson, and Tom Cramer. The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. In *Archiving Conference*, volume 12, pages 16–21, Los Angeles, CA, USA, May 2015. Society for Imaging Science and Technology. doi:10.2352/issn.2168-3204.2015.12.1.art00005. URL <https://purl.stanford.edu/df650pk4327>.
- Stian Soiland-Reyes, Leyla Jael Castro, Daniel Garijo, Marc Portier, Carole Goble, and Paul Groth. Updating Linked Data practices for FAIR Digital Object principles. *Research Ideas and Outcomes*, 8:e94501, December 2022. ISSN 2367-7163. doi:10.3897/rio.8.e94501. URL <https://riojournal.com/article/94501/>.
- Florian Spiess and Heiko Schuldt. Multimodal Interactive Lifelog Retrieval with vitrivr-VR. In *Proceedings of the 5th Annual on Lifelog Search Challenge, LSC '22*, pages 38–42, New York, NY, USA, June 2022. Association for Computing Machinery. ISBN 978-1-4503-9239-6. doi:10.1145/3512729.3533008. URL <https://dl.acm.org/doi/10.1145/3512729.3533008>.
- Florian Spiess and Markus Stauffiger. Forschung und Archive: Erschließung und Zugänglichkeit neu gedacht. *arbido*, 2023(1), 2023. URL <https://arbido.ch/de/ausgaben-artikel/2023/archiv-der-zukunft/forschung-und-archive-erschliessung-und-zugaenglichkeit-neu-gedacht>.
- Florian Spiess, Luca Rossetto, and Heiko Schuldt. Exploring Multimedia Vector Spaces with vitrivr-VR. In Stevan Rudinac, Alan Hanjalic, Cynthia Liem, Marcel Worring, Björn Thor Jónsson, Bei Liu, and Yoko Yamakata, editors, *MultiMedia Modeling*, volume 14557 of *Lecture Notes in Computer Science*, pages 317–323, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-53302-0. doi:10.1007/978-3-031-53302-0_27. URL https://link.springer.com/chapter/10.1007/978-3-031-53302-0_27.
- Caroline Sporleder. Natural Language Processing for Cultural Heritage Domains. *Language and Linguistics Compass*, 4(9):750–768, 2010. ISSN 1749-818X. doi:10.1111/j.1749-818X.2010.00230.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-818X.2010.00230.x>.
- Amanda Sprochi. Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model. *The International Information & Library Review*, 48(2):129–136, April 2016. ISSN 1057-2317. doi:10.1080/10572317.2016.1176455. URL <https://doi.org/10.1080/10572317.2016.1176455>.
- Susan Leigh Star. The Ethnography of Infrastructure. *American Behavioral Scientist*, 43(3):377–391, 1999. doi:10.1177/00027649921955326. URL <https://journals.sagepub.com/doi/10.1177/00027649921955326>.
- Susan Leigh Star. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, 35(5):601–617, September 2010. ISSN 0162-2439. doi:10.1177/0162243910377624.
- Susan Leigh Star and James R. Griesemer. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3):387–420, 1989. ISSN 0306-3127. URL <https://www.jstor.org/stable/285080>. Publisher: Sage Publications, Ltd.
- Susan Leigh Star and Karen Ruhleder. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work, CSCW '94*, pages 253–264, New York, NY, USA, October 1994. Association for Computing Machinery. ISBN 978-0-89791-689-9. doi:10.1145/192844.193021. URL <https://dl.acm.org/doi/10.1145/192844.193021>.
- Regine Stein and Oguzhan Balandi. Using LIDO for Evolving Object Documentation into CIDOC CRM. *Heritage*, 2(1):1023–1031, March 2019. ISSN 2571-9408. doi:10.3390/heritage2010066. URL <https://www.mdpi.com/2571-9408/2/1/66>.
- Isabelle Stengers and Stephen Muecke. *Another science is possible: a manifesto for slow science*. Polity, Cambridge ; Medford, MA, 2018. ISBN 978-1-5095-2180-7 978-1-5095-2181-4.

- Christian Stocker. Use LinkedDataGPT to query Open Linked Data from the City of Zurich, August 2023. URL <https://www.liip.ch/en/blog/use-linkeddatagpt-to-query-open-linked-data-from-the-city-of-zurich>.
- Bruno J. Strasser and Muki Haklay. Citizen Science:Expertise, Demokratie und öffentliche Partizipation. Empfehlungen des Schweizerischen Wissenschaftsrates SWR, 2018. URL https://www.swir.ch/images/stories/pdf/de/Policy_Analysis_SSC_1.2018.Citizen_Science_WEB.pdf.
- Daniel van Strien, Mark Bell, Nora Rose McGregor, and Michael Trizna. An Introduction to AI for GLAM. In *Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop*, pages 20–24. PMLR, March 2022. URL <https://proceedings.mlr.press/v170/strien22a.html>. ISSN: 2640-3498.
- Teo Susnjak. Applying BERT and ChatGPT for Sentiment Analysis of Lyme Disease in Scientific Literature, February 2023. URL <http://arxiv.org/abs/2302.06474>. arXiv:2302.06474 [cs].
- Gabriel Tarde. *Social Laws: An Outline of Sociology*. Batoche Books, Kitchener, Ontario, Canada, 2000.
- Sinclair Target. Whatever Happened to the Semantic Web?, May 2018. URL <https://twobithistory.org/2018/05/27/semantic-web.html>.
- Toma Tasovac, Sally Chambers, and Erzsébet Tóth-Czifra. Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper. *DARIAH-EU*, October 2020. URL <https://hal.science/hal-02961317>.
- Jonathan Tennant, Ritwik Agarwal, Ksenija Baždarić, David Brassard, Tom Crick, Daniel J. Dunleavy, Thomas Rhys Evans, Nicholas Gardner, Monica Gonzalez-Marquez, Daniel Graziotin, Bastian Greshake Tzovaras, Daniel Gunnarsson, Johanna Havemann, Mohammad Hosseini, Daniel S. Katz, Marcel Knöchelmann, Christopher R Madan, Paolo Manghi, Alberto Marocchino, Paola Masuzzo, Peter Murray-Rust, Sanjay Narayanaswamy, Gustav Nilsson, Josmel Pacheco-Mendoza, Bart Penders, Olivier Pourret, Michael Rera, John Samuel, Tobias Steiner, Jadranka Stojanovski, Alejandro Uribe-Tirado, Rutger Vos, Simon Worthington, and Tal Yarkoni. A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship. preprint, SocArXiv, March 2020. URL <https://osf.io/2kxq8>.
- Roy Tennant. MARC must die. *Library Journal*, 127(17):26–27, October 2002. URL <http://soiscompsfall2007.pbworks.com/f/marc%20must%20die.pdf>.
- Melissa Terras, Stephen Coleman, Steven Drost, Chris Elsdén, Ingi Helgason, Susan Lechelt, Nicola Osborne, Inge Panneels, Briana Pegado, Burkhard Schafer, Michael Smyth, Pip Thornton, and Chris Speed. The value of mass-digitised cultural heritage content in creative contexts. *Big Data & Society*, 8(1):20539517211006165, January 2021. ISSN 2053-9517. doi:[10.1177/20539517211006165](https://doi.org/10.1177/20539517211006165). URL <https://doi.org/10.1177/20539517211006165>.
- Nishad Thalpath, Mitsuharu Nagamori, Tetsuo Sakaguchi, and Shigeo Sugimoto. Wikidata Centric Vocabularies and URIs for Linking Data in Semantic Web Driven Digital Curation. In Emmanouel Garoufallou and María-Antonia Ovalle-Perandones, editors, *Metadata and Semantic Research*, Communications in Computer and Information Science, pages 336–344, Cham, 2021. Springer International Publishing. ISBN 978-3-030-71903-6. doi:[10.1007/978-3-030-71903-6_31](https://doi.org/10.1007/978-3-030-71903-6_31).
- Florian Thiery. Sphere 7 Data: LOUD and FAIR Data for the Research Community. Working Paper, Römisch-Germanisches Zentralmuseum, Mainz, Germany, April 2019. URL <https://doi.org/10.5281/zenodo.2643469>.
- Florian Thiery. Linked COVID-19 Data – Semantische Modellierung von Linked Geodata. *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 2020(4):198–204, 2020. ISSN 1618-8950. doi:[10.12902/zfv-0312-2020](https://doi.org/10.12902/zfv-0312-2020). URL <https://doi.org/10.12902/zfv-0312-2020>.
- Florian Thiery, Timo Homburg, Sophie Charlotte Schmidt, Martina Trognitz, and Monika Przybilla. SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data. In *FIG Working Week 2020 Proceedings*, Amsterdam, Netherlands, 2020a. Fédération Internationale des Géomètres. ISBN 978-87-92853-93-6. doi:[10.5281/zenodo.3751770](https://doi.org/10.5281/zenodo.3751770). URL <https://zenodo.org/records/3751770>.
- Florian Thiery, Sophie Charlotte Schmidt, and Timo Homburg. SPARQLing Publication of Irish – Ogham Stones as LOD. In Julian Bogdani, Riccardo Montalbano, and Paolo Rosati, editors, *Archeo. FOSS: 2020 Proceeding of the 14th international conference 15-17 october 2020*, pages 119–127, Oxford (GB), 2020b. Archaeopress. ISBN 978-1-80327-125-5.
- Martin Thomson and Cory Benfield. HTTP/2. Request for Comments RFC 9113, Internet Engineering Task Force, June 2022. URL <https://doi.org/10.17487/RFC9113>.
- Stefan Tilkov. A Brief Introduction to REST, December 2017. URL <https://www.infoq.com/articles/rest-introduction/>.

- Jouni Tuominen, Eero Hyvönen, and Petri Leskinen. Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research. In Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt, editors, *Proceedings of the Second Conference on Biographical Data in a Digital World 2017*, volume 2119 of *CEUR Workshop Proceedings*, pages 59–66, Linz, Austria, November 2017. CEUR. URL <https://ceur-ws.org/Vol-2119/#paper10>.
- Christopher Tweed and Margaret Sutherland. Built cultural heritage and sustainable urban development. *Landscape and Urban Planning*, 83(1):62–69, November 2007. ISSN 0169-2046. doi:10.1016/j.landurbplan.2007.05.008. URL <https://www.sciencedirect.com/science/article/pii/S0169204607001442>.
- Erzsébet Tóth-Czifra, Marta Błaszczczyńska, Francesco Gelati, Gelati Femmy, Mirjam Blümm, Erik Buelinckx, Vera Chiquet, Rita Gautschy, Peter Gietz, Péter Király, Maria Vivas-Romero, Walter Scholger, Bartłomiej Szleszyński, and Ulrike Wuttke. Research Data Management for Arts and Humanities: Integrating Voices of the Community. Report, DARIAH-EU, September 2023. URL <https://doi.org/10.5281/zenodo.8059626>.
- UNESCO. Charter on the Preservation of the Digital Heritage. Circular Letter CL/3865, United Nations Educational, Scientific and Cultural Organization, January 2009. URL <https://n2t.net/ark:/48223/pf0000179529>.
- UNESCO. Preliminary study of the technical, financial and legal aspects of the desirability of a UNESCO recommendation on Open Science. Programme and meeting document 40 C/63, United Nations Educational, Scientific and Cultural Organization, Paris, October 2019. URL <https://n2t.net/ark:/48223/pf0000370291>.
- UNESCO. Implementation of the UNESCO Recommendation on Open Science. Programme and meeting document SC-PCB-SPP/2021/OS/UROS, United Nations Educational, Scientific and Cultural Organization, Paris, November 2021. URL <https://n2t.net/ark:/48223/pf0000379949>.
- UNESCO. Basic texts of the 2003 Convention for the Safeguarding of the Intangible Cultural Heritage. Programme and meeting document CLT-2022/WS/3, United Nations Educational, Scientific and Cultural Organization, Paris, France, 2022. URL <https://n2t.net/ark:/48223/pf0000383762>.
- UNESCO. Culture for Development Indicators. *Methodology Manual*. United Nations Educational, Scientific and Cultural Organization, Paris, France, 2014. ISBN 978-92-3-001227-4. URL <https://n2t.net/ark:/48223/pf0000229608>.
- UNESCO Institute for Statistics. *UNESCO Framework for Cultural Statistics (FCS)*. United Nations Educational, Scientific and Cultural Organization, Montreal, Quebec, Canada, 2009. ISBN 978-92-9189-075-0. URL <https://doi.org/10.15220/978-92-9189-075-0-en>.
- Herbert Van de Sompel. FAIR Digital Objects and FAIR Signposting, May 2023. URL <https://doi.org/10.5281/zenodo.7977333>.
- Herbert Van de Sompel and Michael L. Nelson. Reminiscing About 15 Years of Interoperability Efforts. *D-Lib Magazine*, 21(11/12), November 2015. ISSN 1082-9873. doi:10.1045/november2015-vandesompel. URL <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html>.
- Sigrid Van der Auwera. UNESCO and the protection of cultural property during armed conflict. *International Journal of Cultural Policy*, 19(1):1–19, January 2013. ISSN 1028-6632. doi:10.1080/10286632.2011.625415. URL <https://www.tandfonline.com/doi/full/10.1080/10286632.2011.625415>.
- Sandra van Ginhoven and Claartje Rasterhoff. Art Markets and Digital Histories. *Arts*, 8(3):105, September 2019. ISSN 2076-0752. doi:10.3390/arts8030105. URL <https://www.mdpi.com/2076-0752/8/3/105>.
- Anne van Kesteren. Fetch Living Standard, November 2023. URL <https://fetch.spec.whatwg.org/>.
- Jan van Leeuwen. On Floridi’s Method of Levels of Abstraction. *Minds and Machines*, 24(1):5–17, February 2014. ISSN 1572-8641. doi:10.1007/s11023-013-9321-7. URL <https://link.springer.com/article/10.1007/s11023-013-9321-7>.
- L.W.C. van Lit. The Digital Materiality of Digitized Manuscripts. In *Among Digitized Manuscripts. Philology, Codicology, Paleography in a Digital World*, pages 51–72. Brill, Leiden, the Netherlands, 2020. ISBN 978-90-04-41521-8. URL <https://www.jstor.org/stable/10.1163/j.ctv2gjwzrd.6>.
- Lise Vandenhende and Geert Van Hoorick. The management of cultural heritage and nature : complementary or conflicting regulations? In *EELF Annual Conference, 5th, Abstracts*, 2017. URL <http://hdl.handle.net/1854/LU-8722614>.
- Marilena Vecco. A definition of cultural heritage: From the tangible to the intangible. *Journal of Cultural Heritage*, 11(3): 321–324, July 2010. ISSN 1296-2074. doi:10.1016/j.culher.2010.01.006. URL <https://www.sciencedirect.com/science/article/pii/S1296207410000361>.

- Guillermo Vega-Gorgojo. LOD4Culture: Easy exploration of cultural heritage linked open data. *Semantic Web*, Preprint (Preprint):1–30, January 2023. ISSN 1570-0844. doi:10.3233/SW-233358. URL <https://content.iospress.com/articles/semantic-web/sw233358>. Publisher: IOS Press.
- Tommaso Venturini and Bruno Latour. The Social Fabric. In *Actes Futur en Seine 2009*. Cap Digital, May 2009. URL <https://hal-sciencespo.archives-ouvertes.fr/hal-01293394>.
- Tommaso Venturini, Dominique Cardon, and Jean-Philippe Cointet. Méthodes digitales: Approches quali/quantitative des données numériques - Curation and Presentation of the Special Issue. *Réseaux : communication, technologie, société*, 188(6), 2015. doi:10.3917/res.188.0009. URL <https://hal.science/hal-01672302>.
- Tommaso Venturini, Liliana Bounegru, Jonathan Gray, and Richard Rogers. A reality check(list) for digital methods. *New Media & Society*, 20(11):4195–4217, November 2018. ISSN 1461-4448. doi:10.1177/1461444818769236. URL <https://doi.org/10.1177/1461444818769236>. Publisher: SAGE Publications.
- Vice President for Education. Guidelines «Citing AI Tools»: How to use tools based on artificial intelligence. Version 2.2, University of Basel, Basel, Switzerland, June 2024. URL http://web.archive.org/web/20240721105011/https://www.unibas.ch/dam/jcr:fb36d17a-71ba-4e1e-8676-2cf019254f93/Guidelines-citing-AI-Tools_v2.2.pdf.
- Ruben Vicente-Saez and Clara Martinez-Fuentes. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88:428–436, July 2018. ISSN 0148-2963. doi:10.1016/j.jbusres.2017.12.043. URL <https://www.sciencedirect.com/science/article/pii/S0148296317305441>.
- Bianca Vienni-Baptista, Isabel Fletcher, and Catherine Lyall, editors. *Foundations of Interdisciplinary and Transdisciplinary Research: A Reader*. Bristol University Press, July 2023. ISBN 978-1-5292-3501-2. URL <https://doi.org/10.56687/9781529235012>.
- Dominique Vinck. *Les métiers de l'ombre de la Fête des Vignerons*. Médias et Histoire. Editions Antipodes, Lausanne, Suisse, 2019. ISBN 978-2-88901-171-1. URL <https://doi.org/10.33056/antipodes.1711>.
- Valeria Vitale and Simon Rainer. Linked Data Networks: How, Why and When to Apply Network Analysis to LOD. In Tom Brughmans, Barbara J. Mills, Jessica Munson, Matthew A. Peeples, Tom Brughmans, Barbara J. Mills, Jessica Munson, and Matthew A. Peeples, editors, *The Oxford Handbook of Archaeological Network Research*, Oxford Handbooks, pages 378–389. Oxford University Press, Oxford, New York, November 2023. ISBN 978-0-19-885426-5.
- W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition), December 2012. URL <https://www.w3.org/TR/owl2-overview/>.
- G. Walsham. Actor-Network Theory and IS Research: Current Status and Future Prospects. In Allen S. Lee, Jonathan Liebenau, and Janice I. DeGross, editors, *Information Systems and Qualitative Research*, pages 466–480. Springer US, Boston, MA, 1997. ISBN 978-1-4757-5487-2 978-0-387-35309-8. URL https://doi.org/10.1007/978-0-387-35309-8_23.
- Gary Watson. Free Agency. *The Journal of Philosophy*, 72(8):205–220, 1975. ISSN 0022-362X. doi:10.2307/2024703. URL <https://www.jstor.org/stable/2024703>.
- Jim Webber. A programmatic introduction to Neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*, SPLASH '12, pages 217–218, New York, NY, USA, October 2012. Association for Computing Machinery. ISBN 978-1-4503-1563-0. doi:10.1145/2384716.2384777. URL <https://dl.acm.org/doi/10.1145/2384716.2384777>.
- Jie Wei, Zhuoming Xu, and Wenzhe Xia. DQAF: Towards DQV-Based Dataset Quality Annotation Using the Web Annotation Data Model. In *2016 13th Web Information Systems and Applications Conference (WISA)*, pages 24–27, Wuhan, China, September 2016. IEEE. ISBN 978-1-5090-5437-4. doi:10.1109/WISA.2016.15. URL <https://ieeexplore.ieee.org/abstract/document/7799649>.
- Stuart L. Weibel and Traugott Koch. The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions. *D-Lib Magazine*, 6(12), December 2000. ISSN 1082-9873. doi:10.1045/december2000-weibel. URL <https://mirror.dlib.org/dlib/december00/weibel/12weibel.html>.
- David M. Weigl, Werner Goebl, David J. Baker, Tim Crawford, Federico Zubani, Aggelos Gkiokas, Nicolas F. Gutierrez, Alastair Porter, and Patricia Santos. Notes on the Music: A social data infrastructure for music annotation. In *Proceedings of the 8th International Conference on Digital Libraries for Musicology, DLfM '21*, pages 23–31, New York, NY, USA, July 2021. Association for Computing Machinery. ISBN 978-1-4503-8429-2. doi:10.1145/3469013.3469017. URL <https://dl.acm.org/doi/10.1145/3469013.3469017>.
- Marie O. Weil. Community Building: Building Community Practice. *Social Work*, 41(5):481–499, September 1996. ISSN 0037-8046. doi:10.1093/sw/41.5.481. URL <https://doi.org/10.1093/sw/41.5.481>.

- Dianne Weinthal and Dawn Childress. IIF for Open Access, October 2019. URL <https://escholarship.org/uc/item/260616w7>.
- Richard Weiss. Atlas der schweizerischen Volkskunde : die bisherigen Erfahrungen der Exploratoren. *Schweizerisches Archiv für Volkskunde/ Archives suisses des traditions populaires.*, 38(1):105–118, 1940. doi:10.5169/SEALS-113634. URL <https://www.e-periodica.ch/digbib/view?pid=sav-001:1940:38::303>.
- Etienne Wenger. Communities of practice: a brief introduction. *National Science Foundation*, pages 1–7, October 2011. URL <http://hdl.handle.net/1794/11736>.
- Anna Wessman, Suzie Thomas, Ville Rohiola, Jutta Kuitunen, Esko Ikkala, Jouni Tuominen, Mikko Koho, and Eero Hyvönen. A Citizen Science Approach to Archaeology : Finnish Archaeological Finds Recording Linked Open Database (SuALT). In *Proceedings of the DHN 2019 Conference*, Copenhagen, Denmark, May 2019. CEUR. URL <https://hdl.handle.net/10138/303221>. Accepted: 2019-06-17T12:05:01Z Publisher: CEUR-WS.org.
- Jon White. 10 Lessons Learned From Creating My First Mirador Plugin, October 2023. URL <https://blog.cogapp.com/10-lessons-learned-from-creating-my-first-mirador-plugin-d13b011a35eb>.
- Jens Wietschorke. Historische Kulturanalyse. In Christine Bischoff, Karoline Oehme-Jüngling, and Walter Leimgruber, editors, *Methoden der Kulturanthropologie*, number 3948 in UTB Kulturwissenschaft, pages 160–176. Haupt Verlag, Bern, 1. auflage edition, 2014. ISBN 978-3-8252-3948-0.
- David Wigg-Wolf, Kerstin P. Hofmann, Karsten Tolle, Katja Rösler, Markus Möller, Chrisowalandis Deligio, Julia Tietz, and Caroline von Nicolai. Frankfurt am Main, Deutschland. ClaReNet. Klassifikation und Repräsentation keltischer Münzprägungen im Netz. Das Projekt von 2021 bis 2024. *e-Forschungsberichte des DAI*, pages 1–21 (§), November 2022. ISSN 2198-7734. doi:10.34780/9rgb-or3d. URL <https://publications.dainst.org/journals/efb/article/view/3975>.
- Indrajit Wijegunaratne and George Fernandez. The Three-Tier Application Architecture. In Indrajit Wijegunaratne and George Fernandez, editors, *Distributed Applications Engineering: Building New Applications and Managing Legacy Applications with Distributed Technologies*, Practitioner Series, pages 41–78. Springer, London, 1998. ISBN 978-1-4471-1550-2. URL <https://doi.org/10.1007/978-1-4471-1550-2.3>.
- Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3:160018, March 2016. ISSN 2052-4463. doi:10.1038/sdata.2016.18. URL <https://www.nature.com/articles/sdata201618>.
- Florian Windhager, Paolo Federico, Gunther Schreder, Katrin Glinka, Marian Dork, Silvia Miksch, and Eva Mayr. Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2311–2330, June 2019. ISSN 1077-2626, 1941-0506, 2160-9306. doi:10.1109/TVCG.2018.2830759. URL <https://ieeexplore.ieee.org/document/8352050/>.
- Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, EASE '14, pages 1–10, New York, NY, USA, May 2014. Association for Computing Machinery. ISBN 978-1-4503-2476-2. doi:10.1145/2601248.2601268. URL <https://doi.org/10.1145/2601248.2601268>.
- David Wood, Marsha Zaidman, Luke Ruth, and Michael Hausenblas. *Linked data: structured data on the Web*. Manning, Shelter Island, NY, 2014. ISBN 978-1-61729-039-8.
- Marcia Lei Zeng. Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(2-3):160–182, 2008. ISSN 0943-7444. doi:10.5771/0943-7444-2008-2-3-160. URL <https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-160/>.
- Marcia Lei Zeng. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El Profesional de la Información*, 28(1), January 2019. ISSN 1699-2407, 1386-6710. doi:10.3145/epi.2019.ene.03. URL <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2019.ene.03>.
- Marcia Lei Zeng and Jian Qin. *Metadata*. ALA Neal-Schuman, Chicago, third edition edition, 2022. ISBN 978-0-8389-3795-2 978-0-8389-3796-9.

- Xiaozhu Zou, Siyi Xiong, Zhi Li, and Ping Jiang. Constructing Metadata Schema of Scientific and Technical Report Based on FRBR. *Computer and Information Science*, 11(2):34–39, March 2018. ISSN 1913-8989. doi:10.5539/cis.v11n2p34. URL <https://ccsenet.org/journal/index.php/cis/article/view/73872>.
- Katerina Zourou and Mariana Ziku. Citizen Enhanced Open Science in Cultural Heritage - Review and analysis of practices in Higher Education. Study 2020-1-BE02-KA203-07427, Erasmus+: EU Programme for Education, Training, Youth and Sport, July 2022. URL <https://www.citizenheritage.eu>.
- Joris van Zundert. On Not Writing a Review about Mirador: Mirador, IIF, and the Epistemological Gains of Distributed Digital Scholarly Resources. *Digital Medievalist*, 11(1):1–48, August 2018. ISSN 1715-0736. doi:10.16995/dm.78. URL <http://journal.digitalmedievalist.org//articles/10.16995/dm.78/>.
- Maja Žumer. Functional requirements for bibliographic records: FRBR: The end of the road or a new beginning. *Bulletin of the American Society for Information Science and Technology*, 33(6):27–29, 2007. ISSN 1550-8366. doi:10.1002/bult.2007.1720330608. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bult.2007.1720330608>.

A

Appendix A: AI Tools

Following the ‘Citing AI Tools’ guidelines of the University of Basel [see [Vice President for Education 2024](#)], I have summarised all the AI-based applications that I have used in Table A.1.

Table A.1: AI Tools used for the thesis

Tools	Usage	Affected parts of the work
ChatGPT 4o	Help with structuring the sections and paragraphs Help with formatting texts in L ^A T _E X Help with summarising datasets and articles Help with generating or debugging Python scripts	Entire thesis Entire thesis Literature Review Empirical chapters (6, 7, 8)
DeepL Translator	Translation of text passages	Literature Review and Discussion
DeepL Write	Reformulation of text passages in British English	Entire thesis
Perplexity AI	Help with finding further relevant articles	Literature Review
Whisper	Speech recognition	Transcripts of the LUX interviews

B

Appendix B: Supplementary Figures

This appendix contains a selection of figures that have been enlarged to improve the clarity and ease of reading of their detailed visual information. These figures have been rotated and resized to fit the page optimally, ensuring that complex relationships and intricate details are more accessible for closer examination.

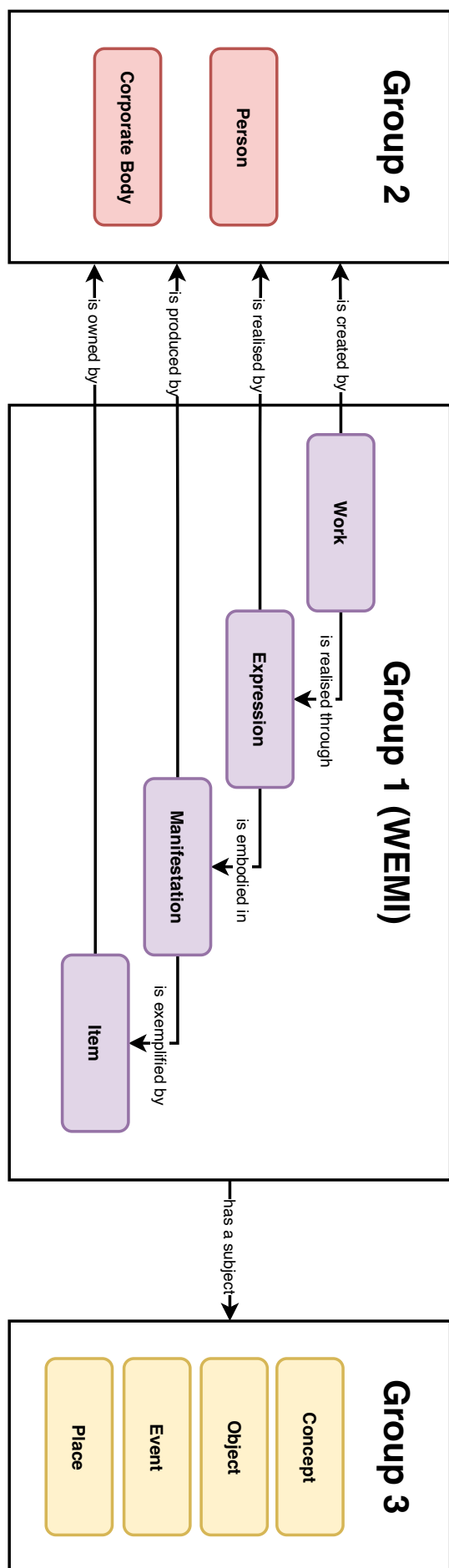


Figure B.1: Enlarged View of The FRBR Conceptual Framework. Adapted from Zou et al. [2018, p. 36].

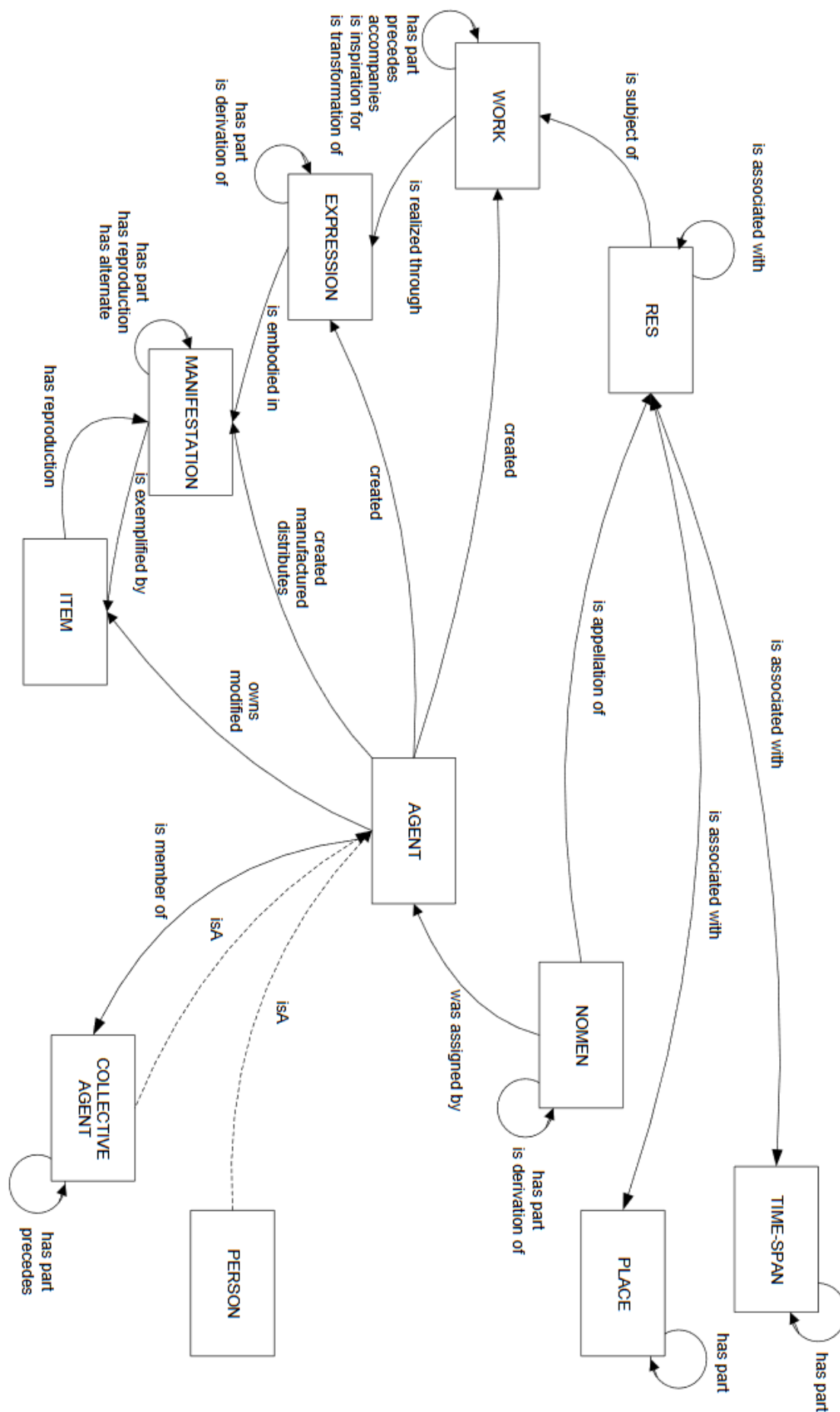


Figure B.2: Enlarged View of the Overview of Relationships in LRM, adapted from Riva et al. [2017, p. 86].

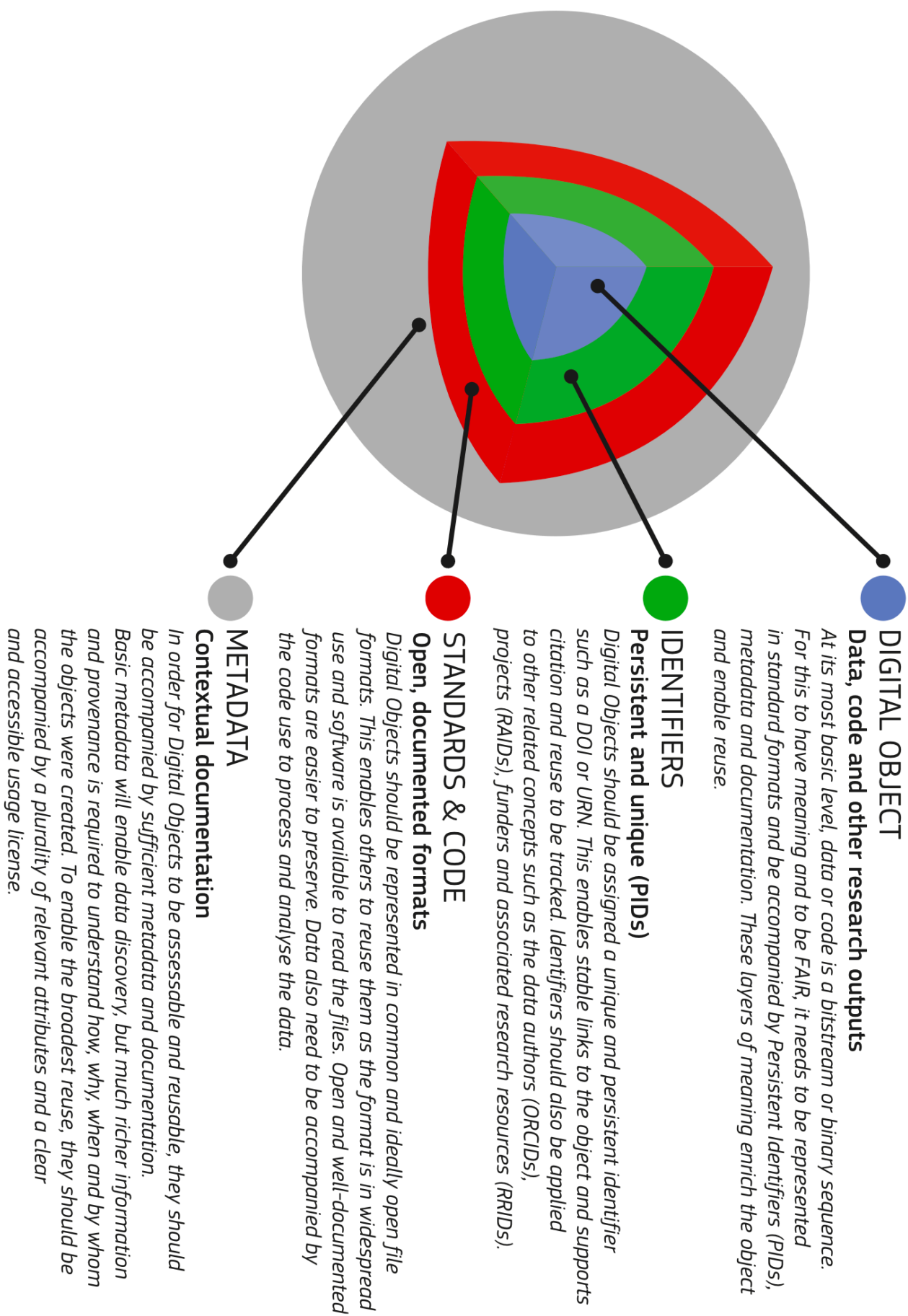


Figure B.3: Enlarged View of The FAIR Digital Object (FDO) Model [European Commission. Directorate General for Research and Innovation. 2018].

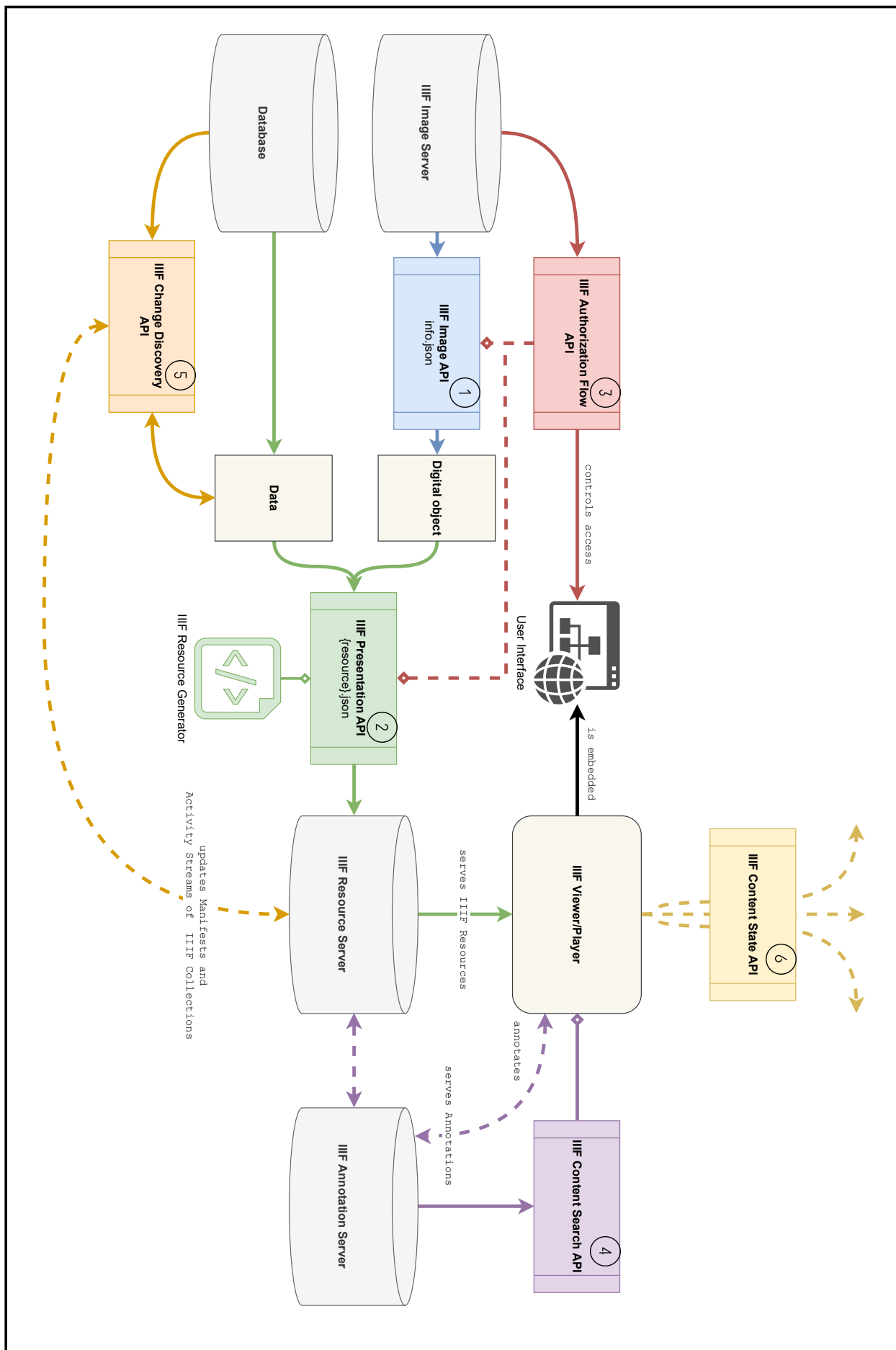


Figure B.4: Enlarged View of an Example of a IIF Architecture Implementing the Six APIs.

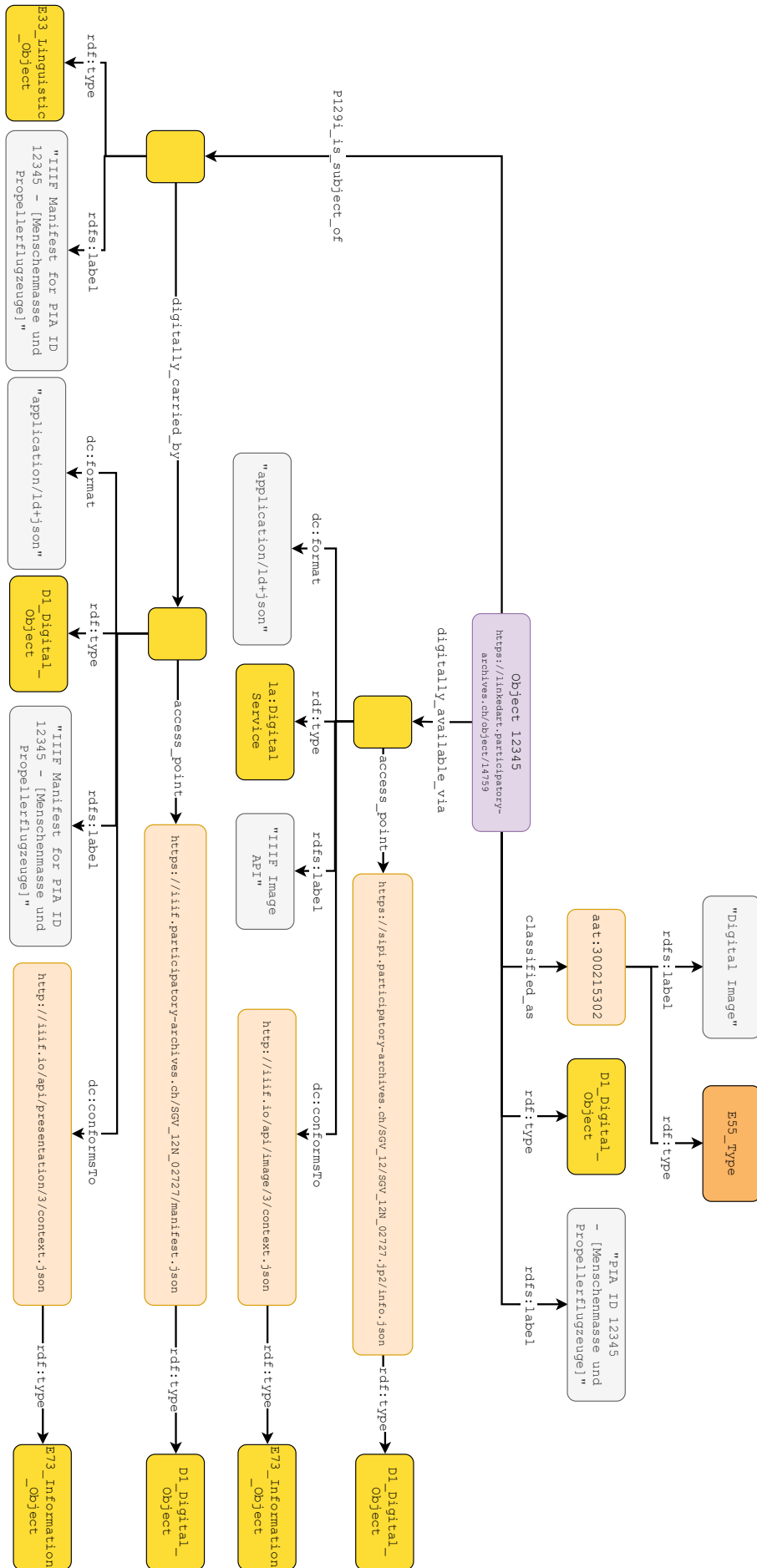


Figure B.5: Enlarged View of Linked Art Digital Integration: Possible Description of Available IIIF Resources.

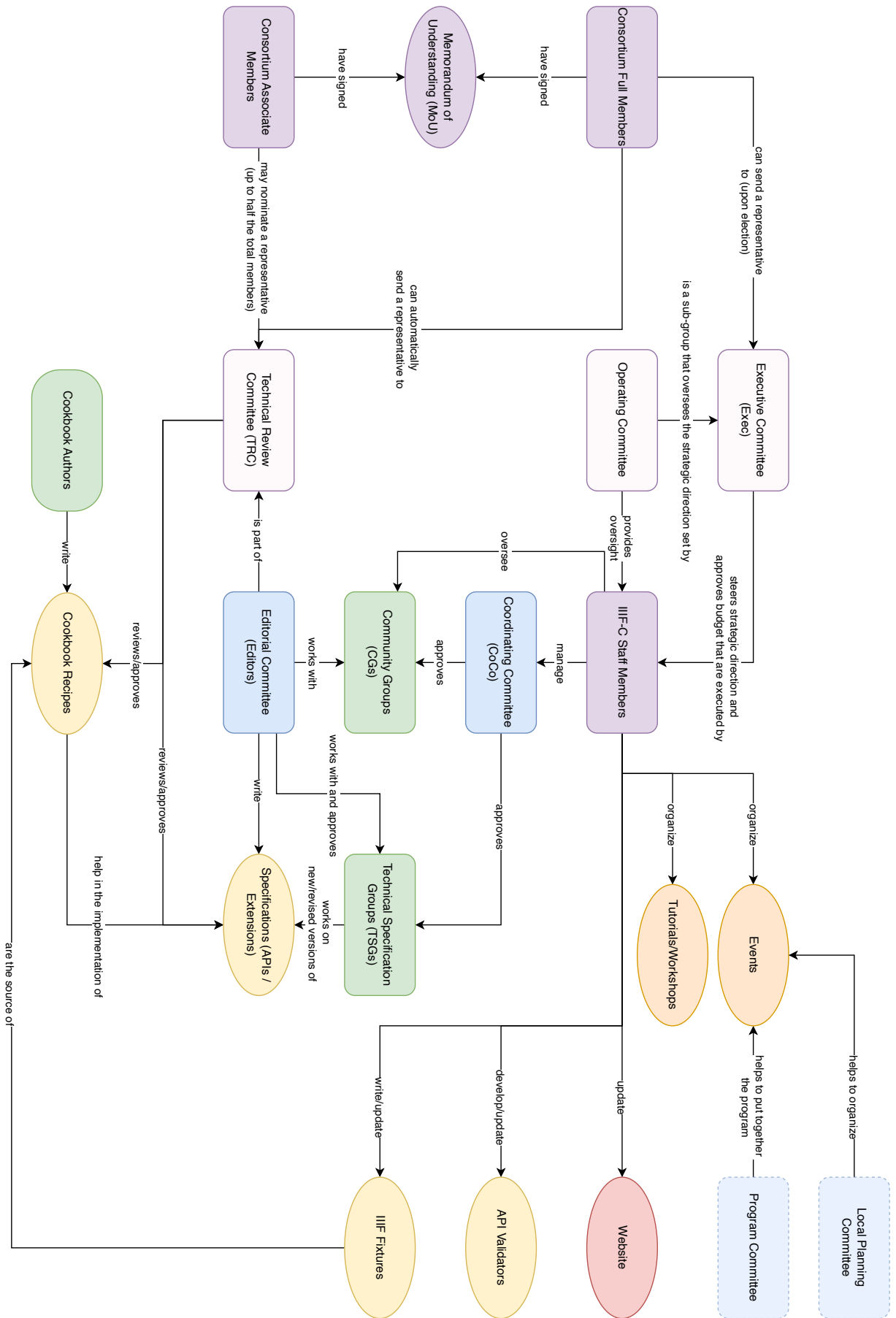


Figure B.6: Enlarged View of the Overview of the IIIF Community.

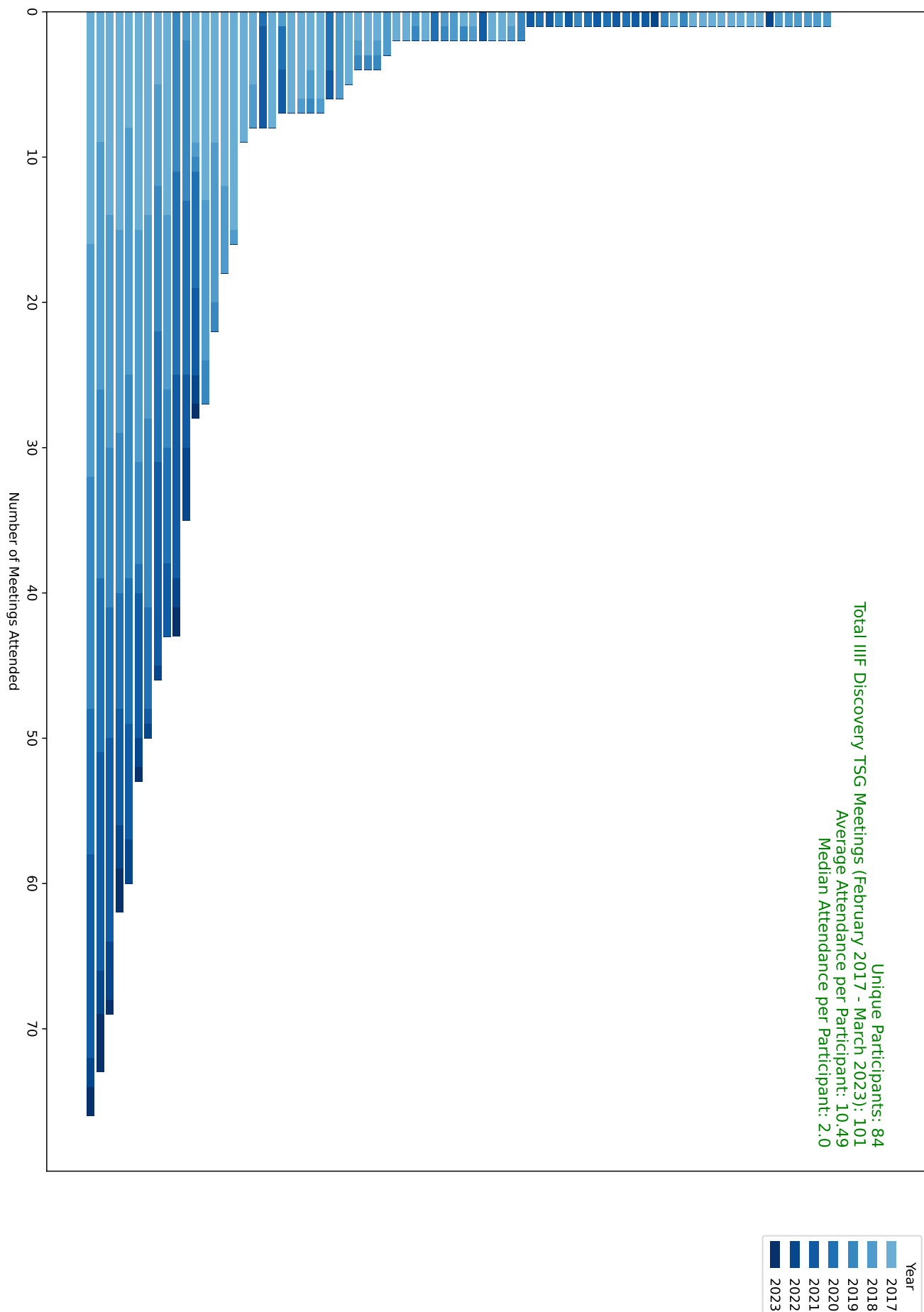


Figure B.7: Enlarged View of the Stacked Bar Chart Showing IIIIF Discovery TSG Meeting Attendance.

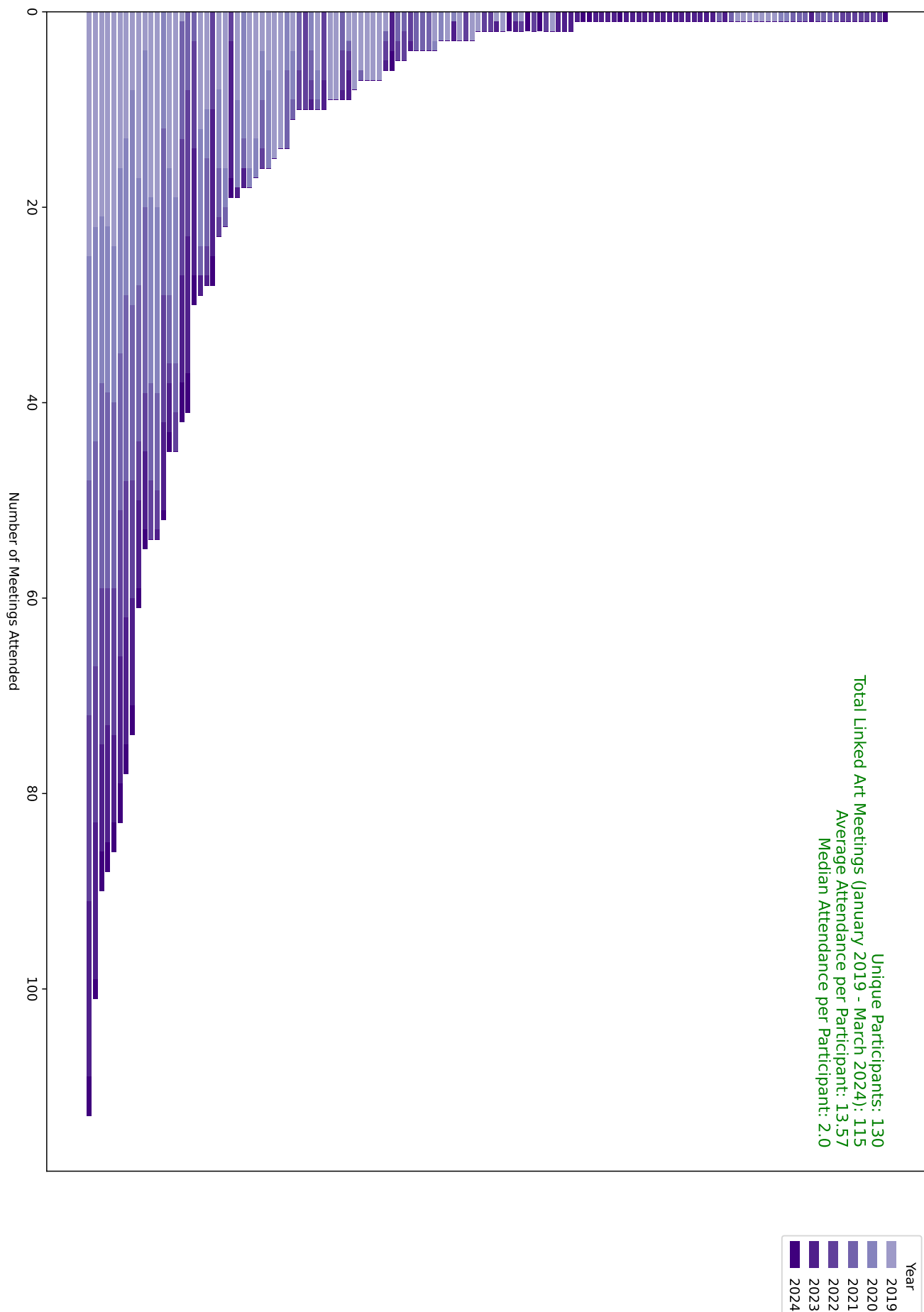


Figure B.8: Enlarged View of the Stacked Bar Chart Showing Linked Art Meeting Attendance.

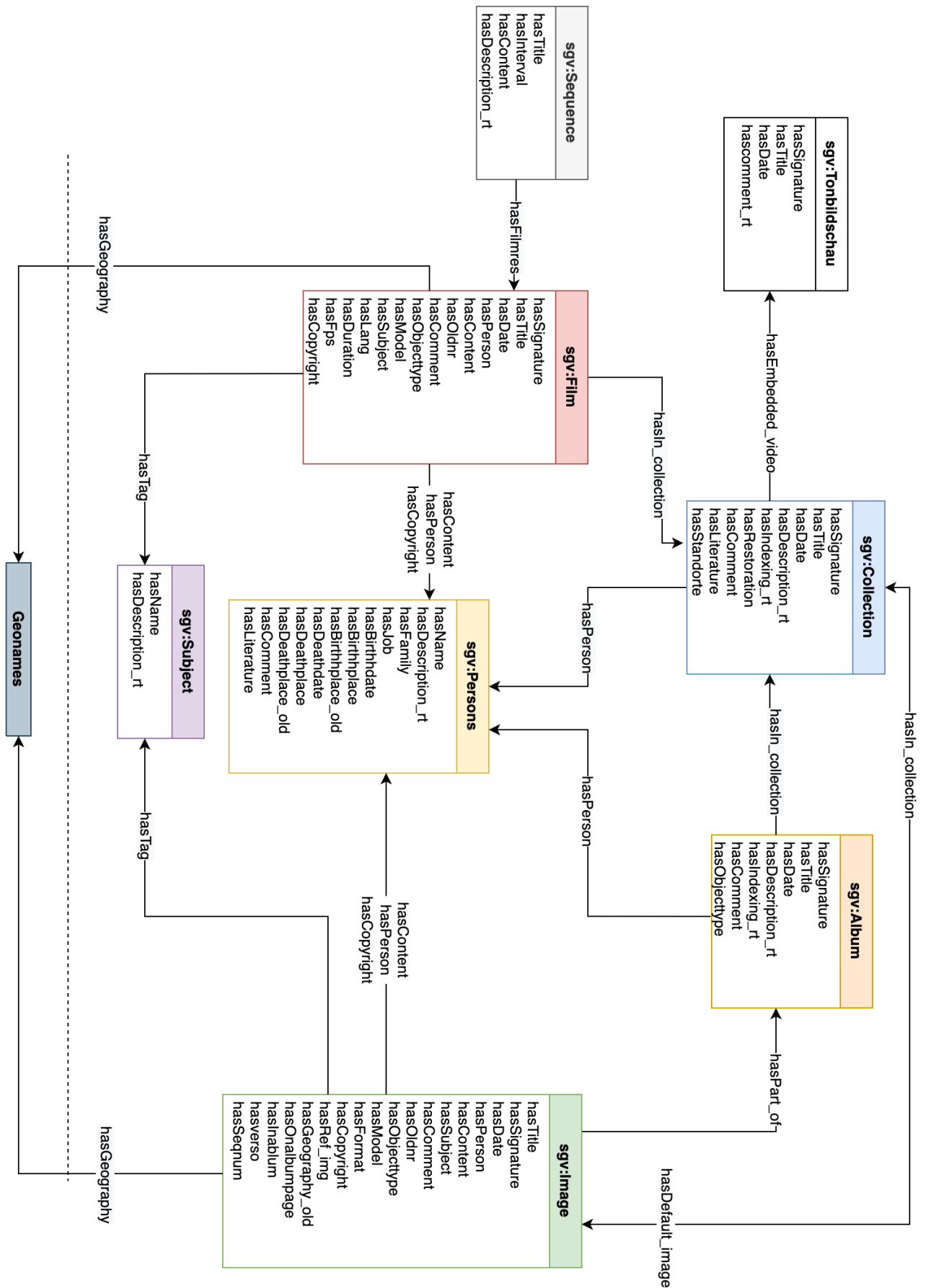


Figure B.11: Enlarged View of the Overview of the Legacy SSFS Data Model.

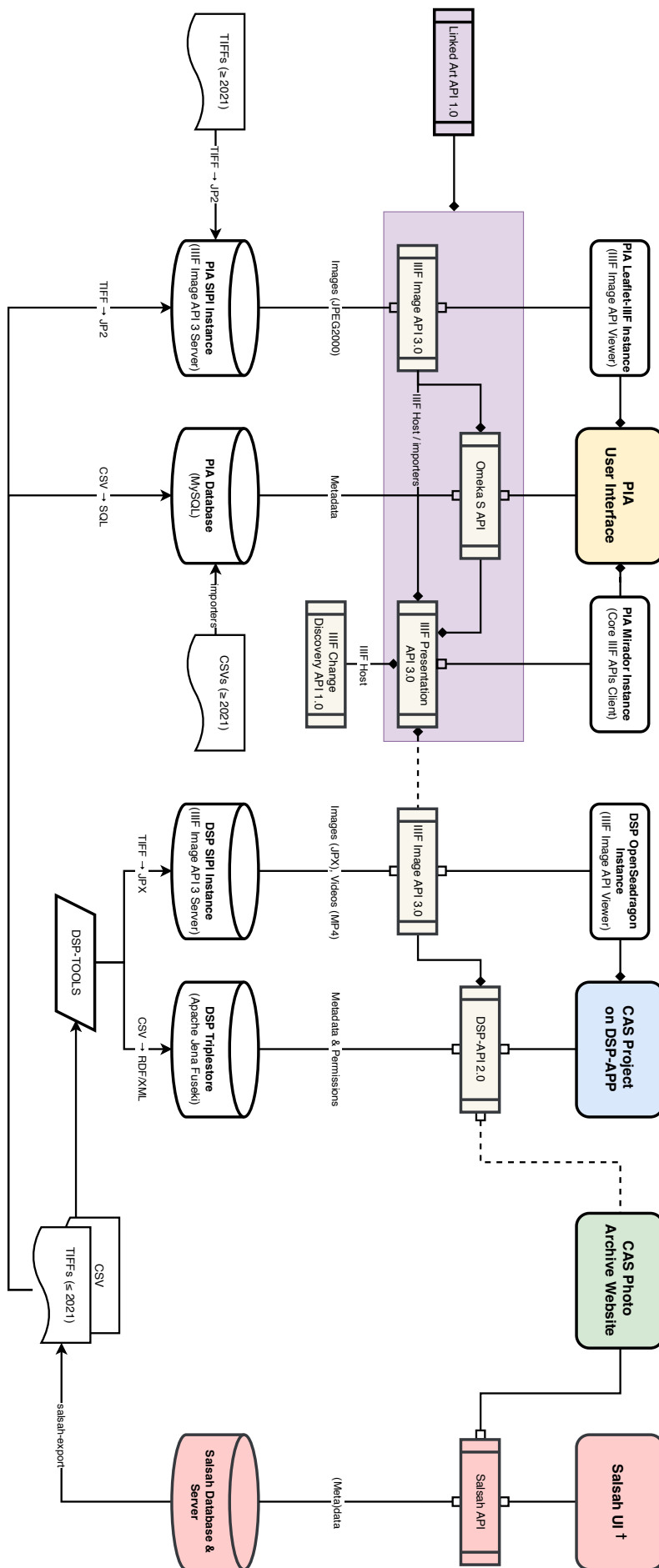


Figure B.12: Enlarged View of the Synoptic View of the PIA Infrastructure: Showcasing its Connection to DSP and the CAS Photo Archive Website.

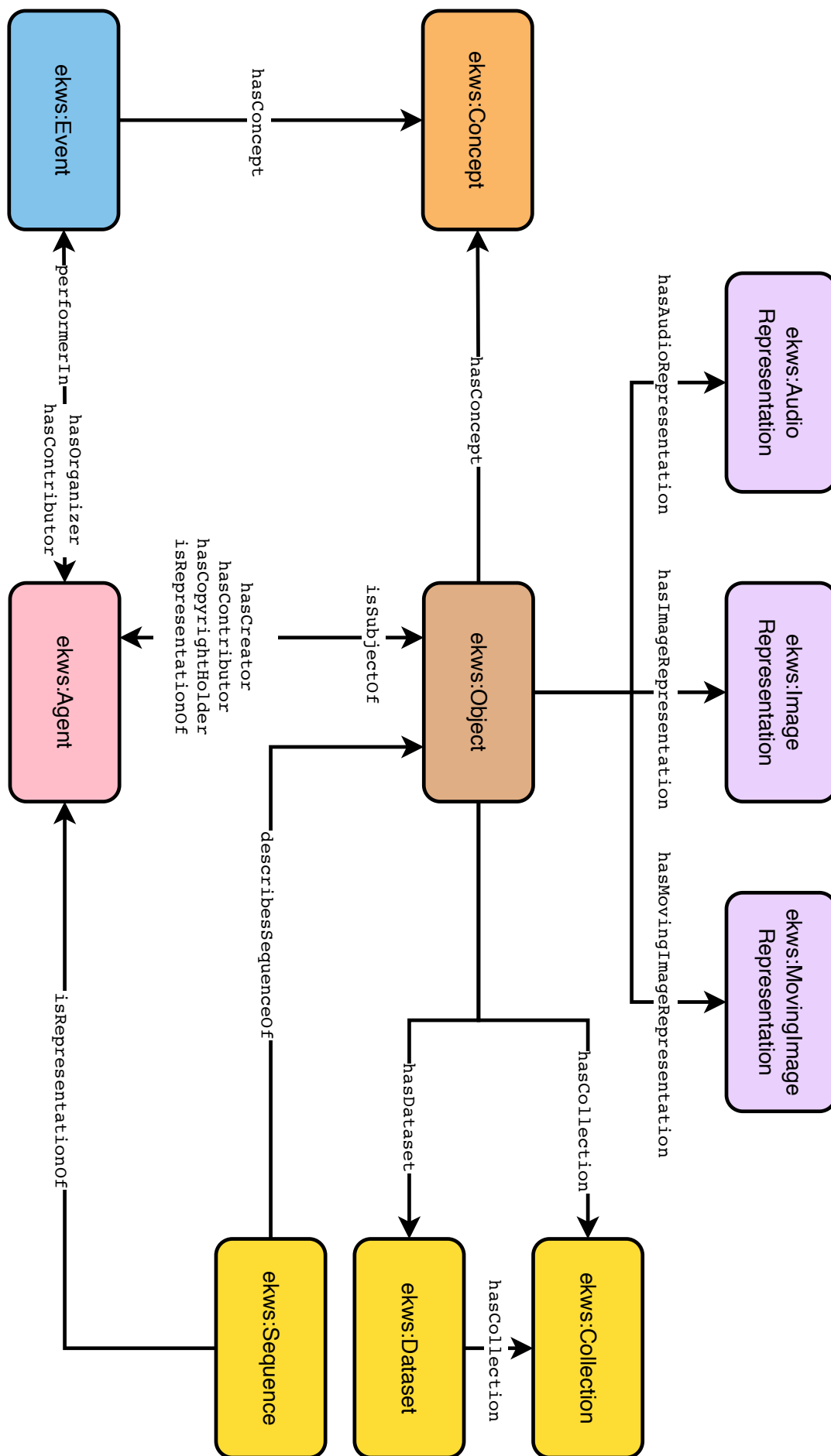


Figure B.13: Enlarged View of the Overview of the Updated CAS Data Model.

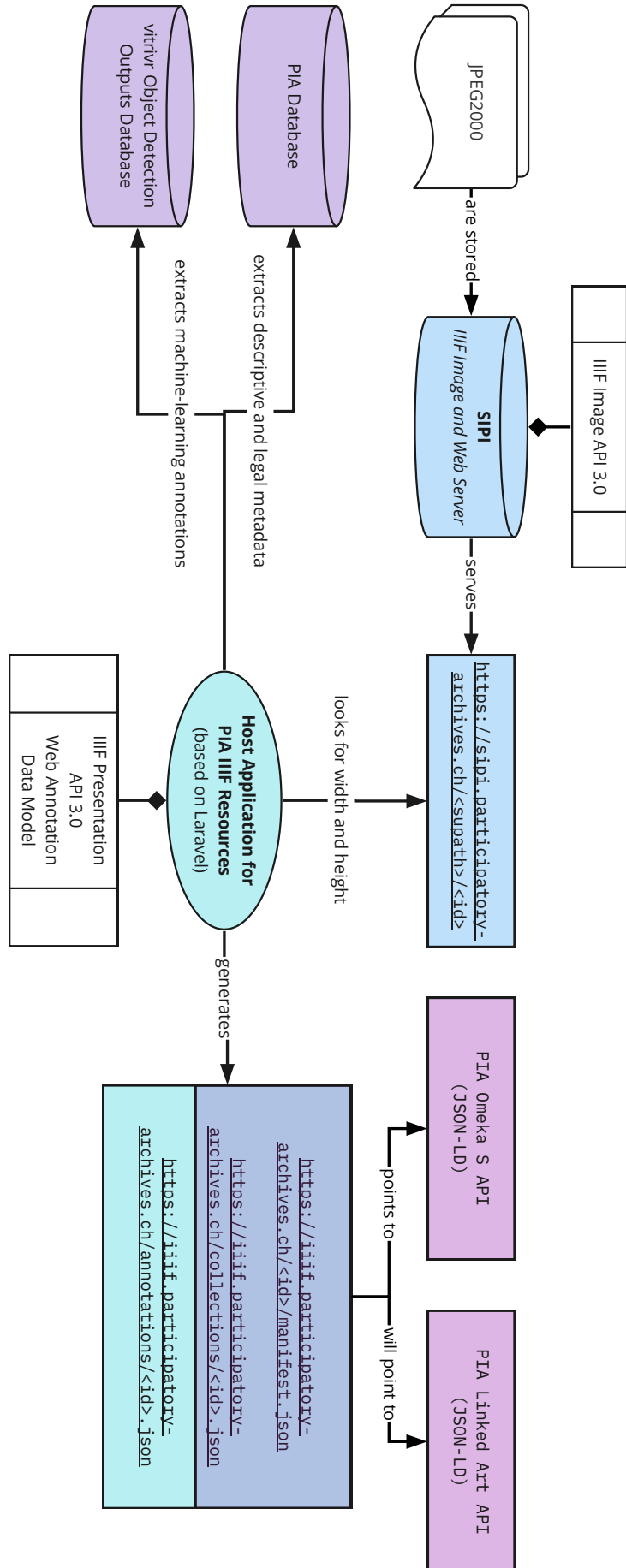


Figure B.14: Enlarged View of the Overview of the Legacy Automatic Annotation Workflow [Cornut et al. 2023, p. 12].

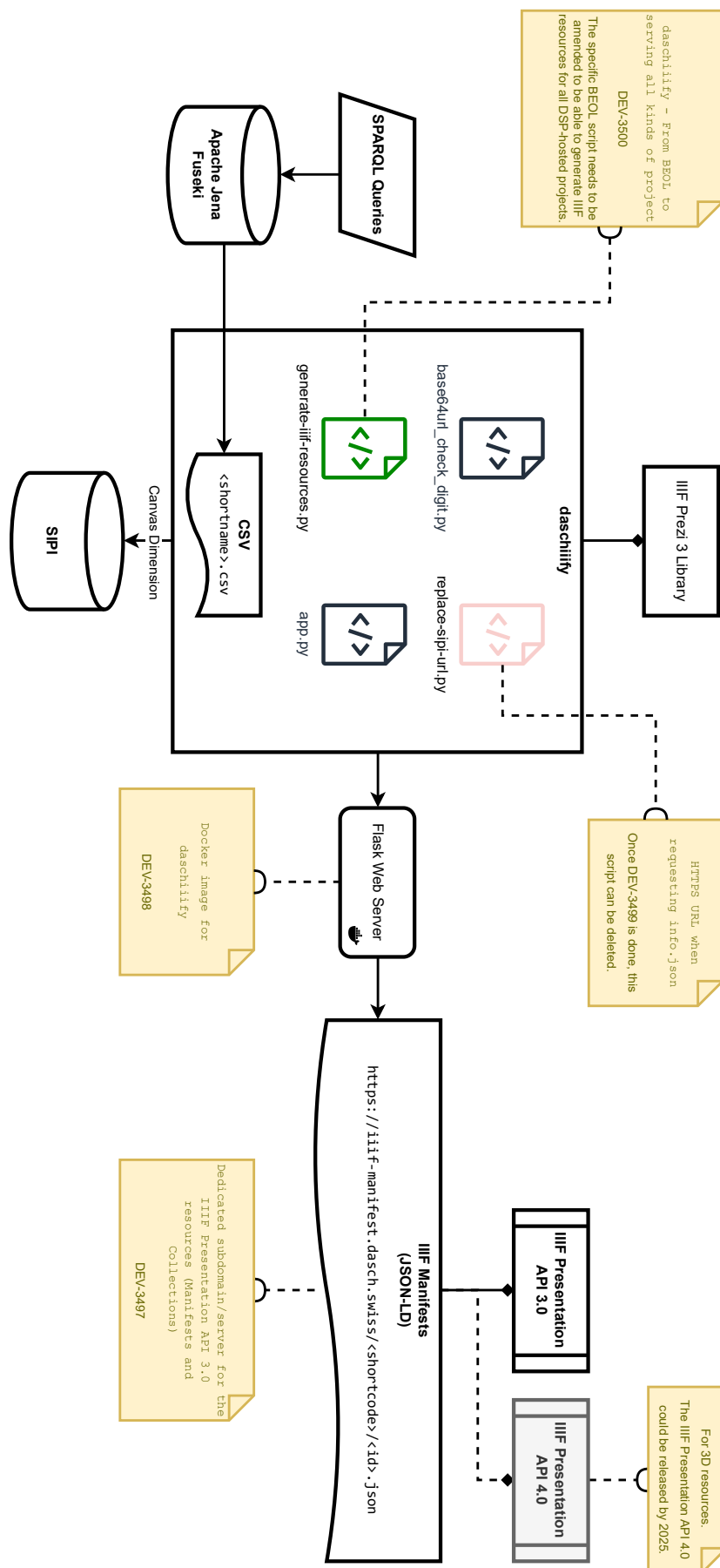


Figure B.15: Enlarged View of the Overview of the IIIF Presentation API Workflow Proof of Concept at DaSCH.