

# Améliorer la valorisation des données du patrimoine culturel grâce au Linked Open Usable Data (LOUD)

Improving the valorisation of cultural heritage data through Linked Open Usable Data (LOUD)

Julien A. Raemy<sup>1</sup>

<sup>1</sup>Universität Basel, Digital Humanities Lab, Spalenberg 65, CH-4051 Basel, Suisse

## Abstract

The re-usability, and thus the valorisation, of Linked Open Data (LOD) in the Digital Humanities can be greatly improved by applying the Linked Open Usable Data (LOUD) design principles and the adhering standards which are the Presentation API 3.0 of the International Image Interoperability Framework (IIIF), the W3C Web Annotation Data Model (WADM) and Linked Art, an RDF application profile based on CIDOC-CRM that uses JSON-LD and Getty vocabularies. The various organisations that have taken advantage of these specifications, both technical and social solutions, have greatly contributed to making the most of cultural heritage data and have opened up new possibilities for end users, in particular by meeting the needs of both the scientific community and software developers. As an example, the research project titled "Participatory Knowledge Practices in Analogue and Digital Image Archives" (PIA) is highlighted. This project aims to leverage IIIF, WADM and Linked Art as part of its data model in the context of a Citizen Science initiative focusing on three photographic collections of the Swiss Society for Folklore Studies (SSFS). Thanks to this technological foundation putting LOUD into practice, PIA will enable a more participatory use of the archive, offering possibilities for different target audiences to contextualise, link, contrast and annotate images and their associated metadata.

## Keywords

CIDOC-CRM, Application Programming Interface (API), International Image Interoperability Framework (IIIF), JavaScript Object Notation for Linked Data (JSON-LD), Linked Art, Linked Open Data (LOD), Linked Open Usable Data (LOUD), cultural heritage, participatory knowledge practices in analogue and digital image archives, Web Annotation Data Model (WADM)

## 1. Introduction

L'application du web sémantique et plus précisément du *Linked Open Data* (LOD) ou « données ouvertes liées » dans le domaine du patrimoine culturel se manifeste principalement dans la publication de jeux de données, mises en évidence ici de trois manières, allant de la plus élémentaire à la forme de publication la plus complexe.

Premièrement, les notices descriptives peuvent être accompagnées d'une représentation conforme à la syntaxe *Resource Description Framework* (RDF) à télécharger, par exemple dans des formats de sérialisation RDF/XML ou Turtle. Une deuxième manière est de mettre en place un mécanisme de négociation de contenu permettant autant aux humains qu'aux machines d'accéder aux métadonnées. Enfin, mais de façon plus ponctuelle, lorsque des services SPARQL

---

. *Workshop on Digital Humanities and Semantic Web*

.  [julien.raemy@unibas.ch](mailto:julien.raemy@unibas.ch) (J. A. Raemy)

.  <https://github.com/julsraemy> (J. A. Raemy)

.  0000-0002-4711-5759 (J. A. Raemy)

sont déployés, ce qui permet de réaliser des requêtes fédérées en ligne ou via un terminal [1].

Ce travail de publication est autant réalisé par les institutions du patrimoine culturel, que sont les bibliothèques, archives et musées, que par les plateformes d'agrégation tel qu'Europeana qui structure les données selon le Europeana Data Model (EDM) lors de leur ingestion [2] voire les agrège directement en RDF lors de projets pilotes [3].

Un autre cas d'utilisation du LOD est celle liée à la curation de données et notamment aux efforts d'alignement de termes descripteurs et d'entités avec des notices d'autorité de bibliothèques telles que Rameau, le Gemeinsame Normdatei (GND) ou encore le Library of Congress Subject Headings (LCSH). Il y a également les services web collaboratifs comme Wikimédia et notamment Wikidata, leur projet de données structurées, qui prend de plus en plus d'importance. Ce dernier est devenu un nœud reliant de nombreux identifiants et une ressource promouvant autant l'utilisation de concepts liés [4] qu'un instrument facilitateur dans les efforts d'alignement, entre autres avec Wikibase, logiciel lui aussi créé sous l'égide de Wikimédia [5].

Les humanités numériques (HN) bénéficient de ces efforts, conduits par les institutions du patrimoine culturel (bibliothèques, archives, musées) et les agrégateurs, principalement pour les deux motifs suivants : améliorer la mise en contexte des données et faire la critique des méthodes et outils appliqués pour les produire. Les HN contribuent même souvent à ces processus, y compris au stade initial, en raison de leur proximité avec ce maillage de communautés. Si les finalités entre les domaines du patrimoine culturel numérique (*Digital Cultural Heritage*) et les HN varient dans une large mesure, ceux-ci partagent des concepts tels que « [...] *mettre à disposition des informations détaillées sur les objets comme fondement de la recherche et un lien étroit avec la création et la perception de la visualisation et de l'imagerie* »<sup>1</sup>. [6].

Hyvönen [7] parle également qu'une transition dans l'utilisation des portails de web sémantique est nécessaire, passant d'une logique de publication de données à celle de l'analyse et de la découverte de connaissances fortuites. Il identifie trois générations de portails :

1. Une première génération de systèmes comprenant le développement de portails pour l'harmonisation des données, l'agrégation, la recherche et la navigation ;
2. Une deuxième génération de systèmes fournissant aux utilisateurs un ensemble d'outils intégrés pour résoudre des problèmes de recherche de manière interactive ;
3. Une troisième génération de systèmes basée sur l'intelligence artificielle qui pourrait résoudre automatiquement des questions de recherche en fonction des contraintes fixées par les scientifiques.

Hyvönen [7] avance que si les HN ont contribué au déploiement de systèmes de la deuxième génération, une demande plus élevée de critique des sources ainsi que des compétences dans l'utilisation d'outils informatiques sophistiqués seront nécessaire pour atteindre la troisième étape afin de stimuler le champ d'action des HN.

Une des clés pour développer de tels systèmes, autant destinés aux humains qu'aux machines et qui puissent être réutilisables facilement, résiderait dans la capacité de faire « partie du web » et non seulement de créer des pièces « sur le web » [8].

---

1. Traduction de l'auteur

## 2. Un premier tournant : JSON-LD

Les projets autour du LOD se sont donc principalement concentrés sur la publication et la consommation de données et plutôt dirigés pour un public d'expert-e-s ayant des connaissances en RDF. De plus, les projets LOD ont rarement pris avantage de l'architecture du web, par exemple en construisant des interfaces de programmation applicatives (API - *application programming interface*) ou services web respectant le style d'architecture RESTful.

Un premier tournant a été effectué en 2014 avec la publication de *JavaScript Object Notation for Linked Data* (JSON-LD), une sérialisation RDF qui permet d'exprimer les caractéristiques de RDF de façon à ce qu'elles puissent être interprétées par les développeurs comme du JSON standard, une syntaxe très courante sur le web et qui a pour avantage pour les développeurs de ne pas avoir besoin de comprendre le « formalisme de RDF avant de pouvoir accéder aux données » [9].

Au sein des HN, le développement des API au sein de la communauté IIF, spécifications toutes sérialisées en JSON-LD, est un exemple de réussite à grande échelle. Les principes de conception promulgués par IIF ont d'ailleurs servi de base à ceux du *Linked Open Usable Data* (LOUD).

## 3. Linked Open Usable Data (LOUD)

LOUD ou « données ouvertes liées et utilisables », est un terme et une démarche proposés par Robert Sanderson, impliqué dans la conception des normes et standards présentés ci-après.

Une des premières intentions du LOUD est de permettre autant à la communauté scientifique qu'aux développeurs de logiciels d'accéder aux données. Il convient de trouver un équilibre qui prenne en compte les besoins en matière d'exhaustivité et de précision des données, qui dépend de la construction ontologique, et les préoccupations pragmatiques que sont l'évolutivité et la facilitation d'utilisation.

A l'instar du programme de déploiement en cinq étoiles des données ouvertes liées de Tim-Berners Lee<sup>2</sup>, cinq principes de conception (*design principles*) encadrent le LOUD<sup>3</sup> :

### A. La bonne abstraction en fonction du public (*The right Abstraction for the audience*)

Il faut privilégier les cas d'utilisation au lieu de la rigueur ontologique afin de déterminer le niveau d'interopérabilité.

### B. Peu d'obstacles à l'entrée (*Few Barriers to entry*)

Les données, et le modèle sous-jacent, doivent être faciles à utiliser et à exploiter. La mise en place de tels systèmes incitera davantage de personnes à y recourir activement.

### C. Compréhensible par introspection (*Comprehensible by introspection*)

Les données doivent être compréhensibles dans une large mesure simplement en les consultant, sans requérir de l'aide extérieure. Cela peut être achevé en sérialisant les informations en JSON-LD, un format d'encodage de donnée structurées simple à lire et répandu sur le Web.

---

2. Open Data 5 étoiles : <https://5stardata.info/>

3. LOUD : <https://linked.art/loud/>

#### D. **Documentation comportant des exemples concrets** (*Documentation with working examples*)

Une documentation des plus exhaustives doit être réalisée afin de clarifier la mise en œuvre des cas d'utilisation.

#### E. **Peu d'exceptions, mais de nombreux modèles cohérents** (*Few Exceptions, instead many consistent patterns*)

Un modèle doit pouvoir contenir le moins d'exceptions possibles pour éviter d'ajouter des règles demandant la création de champs personnalisés au cas par cas.

## 4. Les standards LOUD

Selon Sanderson [8], il y aurait trois systèmes qui suivent les principes de conception du LOUD : IIF et plus particulièrement la troisième version de l'API Présentation, le *Web Annotation Data Model* (WADM) du *World Wide Web Consortium* (W3C) et *Linked Art*. Ces trois spécifications sont complémentaires et peuvent être utilisés séparément ou conjointement.

Si l'on considère la relation entre les principes FAIR (*Findable, Accessible, Interoperable, Reusable*) [10] et LOUD, il apparaît que les premiers sont liés à l'environnement dans lequel se trouvent les données et que le second concerne le contenu en soi. En décomposant l'acronyme LOUD, les termes *Linked, Usable* et (*machine-readable*) *Data* peuvent être considérés comme des caractéristiques des données (et de leur utilisation une fois transférées dans un environnement). *Open* peut constituer en quelque sorte le pendant du principe de réutilisabilité énoncé dans FAIR [11].

### 4.1. International Image Interoperability Framework (IIIF)

Le « cadre international d'interopérabilité des images » (IIIF) est une initiative communautaire, réunissant d'importants acteurs des domaines académique et culturel, qui a défini plusieurs API partagées permettant de normaliser la manière dont les ressources iconographiques et audiovisuelles sont décrites et diffusées sur le web. L'implémentation des API de IIIF permet aux institutions de mieux valoriser leurs collections numérisées ou nées numériques en offrant, par exemple, des possibilités de zoom profond, de comparaison, de recherche plein texte d'objets occlus ou encore d'annotation [12]. Depuis mars 2022, il existe désormais six API stables qui sont conçues et approuvées par le consortium et la communauté IIIF<sup>4</sup>. Les deux spécifications principales sont les API Image et API Présentation. La première étant un « *service web pour manipuler une image à travers une URL* » et la seconde « *spécifie les informations nécessaires à la présentation d'un objet numérique* » [13].

IIIF a établi des principes de conception<sup>5</sup> pour guider la manière dont les spécifications sont créées, en veillant notamment à ce que le périmètre de ces travaux soit constitué autour de cas d'utilisation communs, en évitant de dépendre de technologies qui constituent une barrière à l'entrée et en étant conforme aux normes d'architecture web du W3C et de l'*Internet Engineering*

---

4. API Specifications - International Image Interoperability Framework : <https://iiif.io/api/>

5. IIIF Design Principles : [https://iiif.io/api/annex/notes/design\\_principles/](https://iiif.io/api/annex/notes/design_principles/)

*Task Force* (IETF). Par exemple, lors des sorties des troisièmes versions des deux API principales en 2020, IIIF a suivi les évolutions du W3C et de JSON-LD en intégrant JSON-LD 1.1, sortie la même année, et de s'appuyer sur le WADM au lieu de *Open Annotation*.

Le modèle conceptuel de IIIF et de l'API Présentation se base sur le *Shared Canvas Data Model* décrivant la représentation numérique d'un objet physique au sein d'un canevas auquel est attribué des dimensions pouvant accueillir le contenu, autrement dit une ou plusieurs images ou annotations associées [14]. Ce modèle a été réalisé lors des réunions du *Digital Manuscript Technical Working Group*, groupe actif entre 2010 et 2013, qui fut une initiative préliminaire à IIIF réunissant principalement des médiévistes à l'initiative de Université de Stanford [13]<sup>6</sup>.

Qui plus est, si IIIF ne se base pas *stricto sensu* sur RDF, il respecte sa syntaxe dans un sens conceptuel puisqu'il est en quelque sorte « *un support visuel pour le LOD* »<sup>7</sup> [17]. Si des assertions RDF peuvent bien être créés à partir de ressources IIIF, ce n'est en tout cas pas un but recherché de la communauté. IIIF permet de faire pointer vers d'autres données structurées ou représentations différentes d'un objet via la propriété *seeAlso*, mais il n'existe pas encore de consensus ou de bonnes pratiques sur la manière dont cela devrait être effectué à des buts d'agrégation même si cela pourrait changer avec, premièrement l'API Change Discovery sortie en 2021 et la création du groupe communautaire Discovery for Humans [18, 19]. Néanmoins, des expérimentations ont déjà été réalisés, notamment par Europeana [3, 20].

## 4.2. Web Annotation Data Model

Le WADM a été créé parallèlement à un vocabulaire et un protocole par un groupe dédié du W3C qui a retravaillé la spécification *Open Annotation*, développée en 2013, « (...) *définissant un cadre interopérable pour la création d'associations entre des ressources connexes, des annotations, à l'aide d'une méthodologie conforme à l'architecture du Web* »<sup>8</sup> [21].

Le WADM est, comme tous les standards LOUD, sérialisé en JSON-LD. Son principe de fonctionnement, comme illustré sur la Figure 1, repose sur la division d'une annotation en deux parties distinctes que sont le corps (*body*), qui correspond à la ressource sur laquelle on cherche à annoter quelque chose, et la cible (*target*), qui représente l'objet que l'on annote.

Le contexte de création des annotations est défini par le biais d'une dizaine de « motivations » qui seront interprétées selon le client. Par exemple, le visualiseur d'images *Mirador 3*, qui est compatible aux API de IIIF et au WADM, utilisera et affichera par défaut les annotations ayant *commenting* comme motivation, un de cas d'utilisation les plus fréquents.

## 4.3. Linked Art

*Linked Art* est un profil RDF de l'ontologie CIDOC-CRM sérialisé en JSON-LD qui se veut pragmatique et moins complexe que CIDOC-CRM en mettant à disposition une API et qui

---

6. La plupart des ressources compatibles à IIIF restent à ce jour des manuscrits et livres anciens. Premièrement, car il s'agit en quelque sorte du cœur de la communauté IIIF qui a réellement débuté en 2014 lors de la publication de la deuxième version de *e-codices*, bibliothèque virtuelle des manuscrits en Suisse, avec le déploiement des deux API principales [15]. Dans un deuxième temps, énormément d'argent et d'effort ont été investi par les institutions dans la numérisation de ces objets, tombés dans le domaine public depuis longtemps et qu'il est plus aisé de disséminer [16].

7. Traduction de l'auteur

8. Traduction de l'auteur

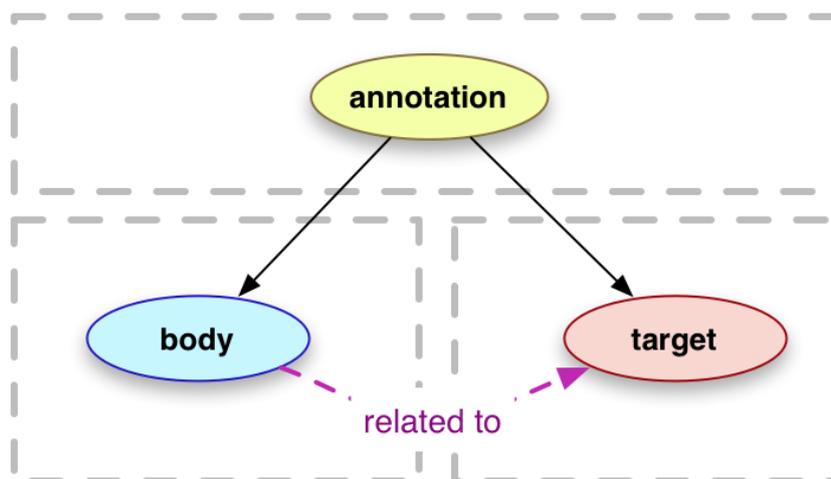


FIGURE 1 : Principes de fonctionnement du WADM

fait appel aux vocabulaires du Getty (tels que AAT et ULAN) pour préciser les entités des ressources du patrimoine culturel décrites. Linked Art utilise également d'autres ontologies RDF couramment utilisées comme RDFS ou Dublin Core pour compléter son modèle, par exemple lorsqu'il faut désambiguïser des noms de propriétés proches ou similaires déjà utilisés par CIDOC-CRM.

Linked Art décrit les assertions dans un paradigme axé sur les événements, plutôt que sur les objets (comme c'est le cas avec la plupart des normes de bibliothèques ou encore avec Dublin Core ou Schema.org) et est documenté, tout comme IIF, dans une démarche ascendante où les cas d'utilisation communs des parties prenantes influencent grandement le modèle. Techniquement, voici comment Linked Art fonctionne [9] :

« *Linked Art utilise le concept contexte de JSON-LD pour faire correspondre les classes sélectionnées et leurs relations respectives à partir de l'ontologie RDF publiée de CIDOC-CRM vers des propriétés JSON en appliquant un ensemble de règles établies.* »<sup>9</sup>.

Parmi ces règles, ou patterns, les suivantes méritent d'être mentionnées :

- **Spécificité des classes et classifications** (*Types and Classifications*) : CIDOC-CRM est un environnement qui doit être enrichi de vocabulaires et d'ontologies additionnels pour être fonctionnel. Le mécanisme fourni pour ce faire est la propriété `classified_as`, qui fait référence à un terme d'un vocabulaire contrôlé. Ceci est en contraste avec la propriété `type`, qui est utilisée pour les classes définies par CIDOC-CRM (`crm:P2_has_type`). Ainsi, la responsabilité du maintien de la structure des classes est déplacée de l'ontologie vers le vocabulaire [9].
- **Noms et identifiants d'une ressource** (*Names and Identifiers for a Resource*) : La désignation des objets se fait via la propriété `_label` (`rdfs:label`) pour la documentation interne. En revanche, pour leur donner un nom ou un label si aucun identifiant n'a été créé via un système, il faut utiliser la propriété `_identified_by` en indiquant que la ressource

9. Traduction de l'auteur

a le type Name. Ce pattern a pour objectif de rendre l'instance du nom comme unique.

- **Partitionnement d'entités (Parts)** : Linked Art a choisi le partitionnement d'entités via la propriété `part` pour tous les cas où l'on souhaite regrouper des propriétés spécifiques d'une entité. Ceci dans le but d'obtenir des descriptions de plus en plus granulaires ou spécifiques. L'appartenance à un set est traité différemment dans le cas où cet ensemble pourrait ne plus comporter de membres et toujours en être un (on peut imaginer une institution où tous les employés décident de se retirer). Dans ce cas là, les propriétés `member` ou `member_of` sont utilisés.

Par ailleurs, Un avantage non négligeable du modèle est la possibilité d'intégrer des objets et services numériques<sup>10</sup>, de sorte qu'il est aisé de pointer vers des ressources IIF en donnant des informations quant à leur niveau de conformité par rapport à une API au sein d'un modèle de données [22].

Linked Art est une initiative encore récente et en développement. Néanmoins, le modèle a notamment été adopté par Pharos, le consortium international des archives photographiques, pour mettre à disposition ses descriptions de collections [23] et est en train d'être implémenté par la Galerie nationale d'art des États-Unis (*National Gallery of Art*) ou encore l'université de Yale pour uniformiser leurs pratiques en matière de publication de données.

## 5. Pratiques de connaissance participatives dans les archives d'images analogiques et numériques (PIA)

Dans ce chapitre sont présentés le projet de recherche PIA, la mise en œuvre globale des principes de conception et des standards LOUD suivis de trois cas d'utilisation décrits de manière plus détaillée, ainsi que les perspectives d'avenir et l'impact que les standards LOUD peuvent apporter lors de leur déploiement dans le cadre d'un projet de science citoyenne comme celui-ci.

### 5.1. Bref descriptif du projet PIA

En février 2021, le séminaire d'anthropologie culturelle et d'ethnologie européenne (*Seminar für Kulturwissenschaft und Europäische Ethnologie*) et le laboratoire d'humanités numériques (DHLab) de l'Université de Bâle ainsi que le *Institute of Design Research* de la Haute école des arts de Berne ont démarré un projet de recherche interdisciplinaire intitulé « Pratiques de connaissance participatives dans les archives d'images analogiques et numériques » (PIA - *Participatory Knowledge Practices in Analogue and Digital Image Archives*)<sup>11</sup>.

Le projet PIA, qui se déroule de février 2021 à janvier 2025, prévoit la mise en place de processus, d'outils et d'interfaces permettant de générer et rendre visible la connaissance de manière participative en permettant notamment un accès et une exploration intuitifs, basée sur l'exemple de trois collections des archives photographiques<sup>12</sup> de la Société suisse des traditions populaires (SSTP) - *Schweizerische Gesellschaft für Volkskunde* (SGV) - qui rassemblent des

---

10. Linked Art - Digital Integration : <https://linked.art/model/digital/>

11. PIA Project : <https://about.participatory-archives.ch/>

12. Fotoarchiv der Schweizerischen Gesellschaft für Volkskunde : <https://archiv.sgv-sstp.ch/>

photographies sur des sujets divers tels que la vie quotidienne, la tradition et l'identité, les formes de travail et d'habitation [22].

Parallèlement au déploiement de l'infrastructure et au design de ces interfaces, le projet étudie les phases des archives analogiques et numériques dans une perspective d'anthropologie de la connaissance, de technique et de communication.

Les nouvelles collectes de données font également partie d'une utilisation participative des archives : les utilisatrices et utilisateurs mettent à disposition leurs propres images, ce qui permet d'élargir les collections, elles et ils écrivent ou racontent des faits intéressants à propos des objets ou font part de leurs propres impressions, expériences et analyses. Toutes ces données peuvent à leur tour faire partie des archives et contribuer ainsi, par leur recontextualisation, à la production de connaissances dans différents domaines. Ainsi, les thèmes d'une collection peuvent également être actualisés et étudiés sous de nouveaux angles.

L'objectif du projet est d'analyser et de décrire systématiquement les pratiques archivistiques historiques et actuelles ; de faire des archives un lieu vivant de création, préservation et transmission des connaissances. PIA veut s'inscrire dans la mouvance de l'ouverture des données et afin d'y parvenir, sept priorités ont été identifiées au sein d'une vision commune prenant en compte les aspects interdisciplinaires du projet :

1. Accessibilité ;
2. Hétérogénéité ;
3. Matérialité ;
4. Interopérabilité ;
5. Interconnexions ;
6. Intelligence artificielle ;
7. Gestion des biais.

## 5.2. Application des principes de conception et des standards LOUD

Si instinctivement, les principes de conception encadrant le LOUD s'alignent et peuvent plus facilement contribuer à la réalisation des priorités 1, 4 et 5, toutes les priorités du projet PIA peuvent en fait être réalisés ou optimisés par le biais de Linked Art, IIF ou le WADM.

Plus concrètement, PIA utilisera les trois standards LOUD mentionnés ci-avant. Dans un premier temps, les deux API principales de IIF vont être déployées, ce qui permettra, entre autres, aux utilisatrices et utilisateurs d'annoter les ressources conformément au WADM. Vu qu'il s'agit d'une initiative promouvant la science citoyenne ou participative<sup>13</sup>, il a par exemple été convenu d'autoriser le téléversement d'images privés sur l'interface utilisateur pour contextualiser, comparer, compléter ou encore déconstruire le corpus de la SSTP, obtenu via l'API DaSCH Service Platform (DSP), service du DaSCH qui est actif dans la préservation à long-terme des données de sciences humaines en Suisse.

Les métadonnées, en plus d'une sérialisation en Schema.org, seront également exposées via le modèle Linked Art [25]. Les différentes informations seront reliées dans l'API centrale de PIA, par le biais de la propriété `seeAlso` au sein des ressources IIF pointant vers une représentation de l'objet en Linked Art, en ayant des listes d'annotations WADM rattachées aux ressources IIF et enfin en indiquant au sein du modèle Linked Art lorsqu'un service web IIF est disponible.

---

13. Au sein du titre de travail de la future thèse de doctorat de l'auteur, le terme *Citizen Humanities* - « science citoyenne dans les sciences humaines » - est employé pour le démarquer des initiatives de science citoyenne émanant des sciences naturelles [24]

Cet « écosystème LOUD », schématisé dans la Figure 2, permettra de répondre aux besoins de différents publics, que ce soit des humains ou des machines.

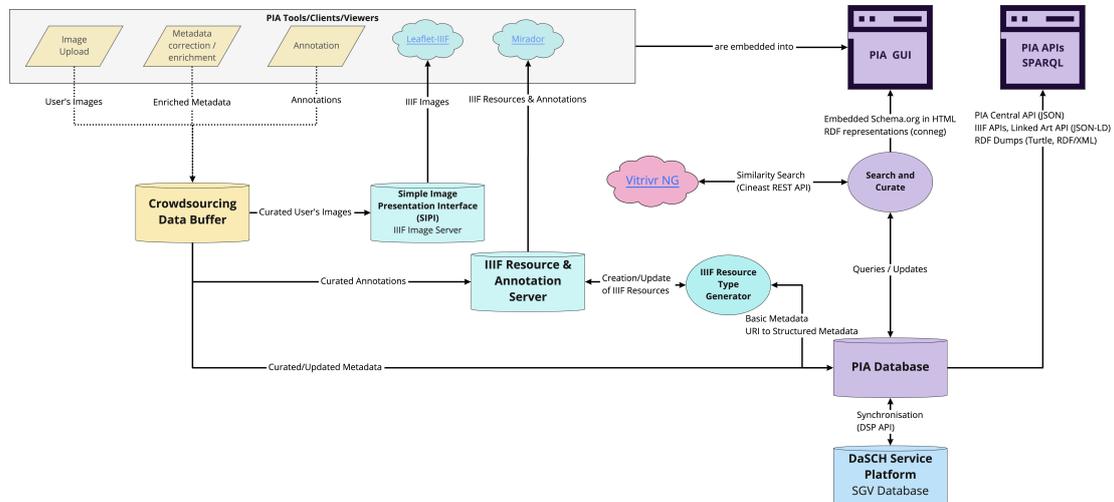


FIGURE 2 : Panorama simplifié de l'architecture de PIA

Il convient donc de déployer des efforts particuliers dans la modélisation des différents cas d'utilisation, mis en évidence dans les sous-sections successives par le biais d'une photographie issue de la collection Ernst Brunner (cf. Figure 3), et de pouvoir les formaliser tout en gérant au mieux les flux entre les différents services et logiciels de l'infrastructure.



FIGURE 3 : SGV\_12N\_02727 (Alte Bildnummer : BC 27). [Menschenmasse und Propellerflugzeuge]. Dübendorf, undatiert. Negativ schwarz/weiß 6x6cm. SGV\_12 (Ernst Brunner). Schweizerische Gesellschaft für Volkskunde (SGV). <https://archiv.sgv-sstp.ch/resource/424962>.

### 5.2.1. Cas d'utilisation 1 : modélisation d'une photographie numérisée et gestion de ses identifiants

La modélisation en Linked Art des différentes entités (objet, personne, événement, etc.) se basera sur la sélection de ressources qui puissent représenter un ensemble cohérent. Pour la ressource qui nous intéresse, une numérisation à partir d'un négatif noir et blanc, celle-ci est très représentative de la collection Ernst Brunner comportant plus de 47'000 objets ayant cette typologie.

Tout d'abord, il a fallu choisir comment un tel objet allait être représenté au centre de la modélisation, de manière conceptuelle, physique ou numérique. Autrement dit est-ce qu'il s'agit plutôt de la photographie telle qu'elle nous est présentée sur le site de la SSTP, du négatif conservé au sein des archives ou encore du fichier issu de la numérisation, réalisée il y a quelques années par le DHLab, à partir du négatif. C'est la dernière option qui a été retenue, car la première était difficilement envisageable vu qu'il s'agit plutôt d'une reconstruction intellectuelle de la photographie et qui serait plus complexe à modéliser en Linked Art et la deuxième nous paraissait superflue à sérialiser car l'utilisation majeure de cette ressource est désormais numérique.

```
1 {
2   "@context": "https://linked.art/ns/v1/linked-art.json",
3   "id": "https://linkedart.participatory-archives.ch/object/14759",
4   "type": "DigitalObject",
5   "_label": "PIA ID 14759 - [Menschenmasse und Propellerflugzeuge]",
6   "classified_as": [
7     {
8       "id": "http://vocab.getty.edu/aat/300215302",
9       "type": "Type",
10      "_label": "Digital Image"
11    }
12  ],
```

Extrait 1 – Modélisation Linked Art de la photographie comme DigitalObject

En d'autres termes, plutôt que d'utiliser la classe CIDOC-CRM HumanMadeObject, la modélisation débute avec une instanciation de la classe CRMdig DigitalObject en suivant notamment les consignes promulguées pour la modélisation de l'entité Object le site web de Linked Art, par exemple en mettant en avant le vocabulaire AAT (cf. Extrait 1).

S'il a été décidé de modéliser à partir de la photographie numérisée, son contexte de création est évidemment évoqué (cf. Extrait 2). Au sein du fichier JSON-LD complet - encore en phase de développement - des métadonnées concernant le lieu où la photographie a été prise ainsi que sur ses dimensions physiques viennent enrichir la description du négatif.

```
1   "created_by": {
2     "type": "Creation",
3     "_label": "Digitisation of Photograph",
4     "used_specific_object": [
5       {
6         "type": "HumanMadeObject",
7         "_label": "Negative of [Menschenmasse und Propellerflugzeuge]",
8         "classified_as": [
```

```

9      {
10     "id": "http://vocab.getty.edu/aat/300128343",
11     "type": "Type",
12     "_label": "Black and White Negative",
13     "classified_as": [
14       {
15         "id": "http://vocab.getty.edu/aat/300435443",
16         "type": "Type",
17         "_label": "Type of Work"
18       }
19     ]
20   },
21   ...
22 }
23 ]
24 ],
25 },

```

### Extrait 2 – Contexte de création et lien vers le format original

Cette ressource comporte plusieurs identifiants :

- BC 27 : identifiant attribué par le photographe Ernst Brunner
- SGV\_12N\_02727 : identifiant attribué par les archives de la SSTP
- 14759 : identifiant créé pour les besoins du projet et retenu comme point d'entrée principal.

En plus de ceux-ci et pour suivre l'un des patterns de Linked Art, s'ajoute l'identification par le nom, ici du titre qui a été attribué par la SSTP. Vu que chaque identifiant a été créé de manière distincte (logique propre, *numerus currens* de manière manuelle et identifiant assigné automatiquement par un système informatique), chaque nœud pointe vers un terme AAT différent et a une labélisation propre (cf. Extrait 3).

En outre, Il faut encore noter que cette ressource a aussi un numéro dédié sur le site web des archives de la SSTP (424962) et que prochainement, elle obtiendra un nouvel identifiant lors de la migration de la base de données qui devrait être effectué par le DaSCH dans le courant de l'année 2022. Néanmoins, ces deux hyperliens sont ou seront plutôt traités comme des représentations complémentaires, le premier est d'ailleurs aussi modélisé dans le fichier JSON complet via la propriété `subject_of`.

```

1  "identified_by": [
2    {
3      "type": "Name",
4      "content": "[Menschenmasse und Propellerflugzeuge]",
5      "classified_as": [
6        {
7          "id": "http://vocab.getty.edu/aat/300404670",
8          "type": "Type",
9          "_label": "Owner-Assigned Title"
10       }
11     ],
12     "language": [
13       {
14         "id": "http://vocab.getty.edu/aat/300388344",
15         "type": "Language",
16         "_label": "German"
17       }
18     ]
19   },
20   {
21     "type": "Identifizier",
22     "content": "BC 27",
23     "classified_as": [
24       {
25         "id": "http://vocab.getty.edu/aat/300417447",
26         "type": "Type",
27         "_label": "Creator-Assigned Number"
28       }
29     ]
30   },
31   {
32     "type": "Identifizier",
33     "content": "SGV_12N_02727",
34     "classified_as": [
35       {
36         "id": "http://vocab.getty.edu/aat/300312355",
37         "type": "Type",
38         "_label": "SGV Signature"
39       }
40     ]
41   },
42   {
43     "type": "Identifizier",
44     "content": "14759",
45     "classified_as": [
46       {
47         "id": "http://vocab.getty.edu/aat/300404621",
48         "type": "Type",
49         "_label": "PIA ID"
50       }
51     ]
52   }
53 ],

```

### 5.2.2. Cas d’utilisation 2 : dissémination des ressources numériques et intégration au modèle de données

Afin de valoriser les ressources numériques et de les disséminer de manière standardisée, les deux API principales de IIIF ont été déployées au sein de l’infrastructure PIA.

Premièrement, chaque image est servie via notre instance de serveur Simple Image Presentation Interface (SIPI), développé et maintenu par le DHLab. Autant la collection photographique de la SSTP que les futurs fichiers téléversés par les utilisatrices et utilisateurs sont et seront converties dans le format JPEG2000 pour générer les différentes tuiles permettant un zoom profond. En parallèle, SIPI s’assure de générer également les informations nécessaires (*Image Information Request URI Syntax*) pour une compatibilité à l’API Image 3.0 [26].

Ensuite, sur la base d’un modèle de Manifest, ressource IIIF principale au sein de l’API Présentation [27], nous avons généré à l’aide d’un script, pour tous les objets disponibles dans notre base de données, des fichiers JSON-LD en liant les données ainsi que les métadonnées descriptives et légales. Pour notre ressource, il s’agit de <https://iiif.participatory-archives.ch/14759/manifest.json> - fichier qui va continuer d’être enrichi, notamment en pointant vers plus de métadonnées structurées<sup>14</sup>.

Vu que notre API Linked Art est en phase de développement, il n’y a pas encore d’intégration au sein des Manifestes IIIF, par le biais de la propriété `seeAlso`, pointant vers une description Linked Art des objets. Dans l’autre direction, la Figure 4 illustre une intégration des deux API de IIIF au sein du modèle de données en suivant les règles de bonnes pratiques édictées par la communauté Linked Art, par exemple en indiquant la version des API, leur format (JSON-LD) ainsi qu’en allouant une classe à chaque nœud<sup>15</sup>.

### 5.2.3. Cas d’utilisation 3 : enrichissement des métadonnées et annotation des contenus

Un des buts du projet PIA est de permettre l’enrichissement et la correction des métadonnées descriptives par le biais d’un approvisionnement par la foule (*crowdsourcing*). Ceci pourrait être réalisé par l’intégration ou l’utilisation d’outils *Open Source* conformes à IIIF tels que *Madoc* ou *Zooniverse* et avec lesquels il est possible de modifier les métadonnées en important des ressources IIIF ou de les annoter directement [28, 29].

Au sein de la ressource 14759, la date est par exemple inconnu même s’il s’agit assez certainement d’une photographie prise à la fin des années 1930 en suivant la logique de numérotation du photographe. De même, sur l’un des avions, on peut apercevoir une inscription portant le nom « József Kiss ». Il reste par contre à savoir s’il s’agit du héros militaire hongrois de la première

---

14. A noter que nous nous efforçons de faire en sorte que les liens des ressources IIIF restent stables et sommes en train de développer une politique d’attribution d’identifiants pérennes à l’aide du schème *Archival Resource Key* (ARK). Cf. <https://github.com/Participatory-Image-Archives/ark>

15. Un graphe complet au format SVG a été généré en parallèle de la sérialisation en JSON-LD.

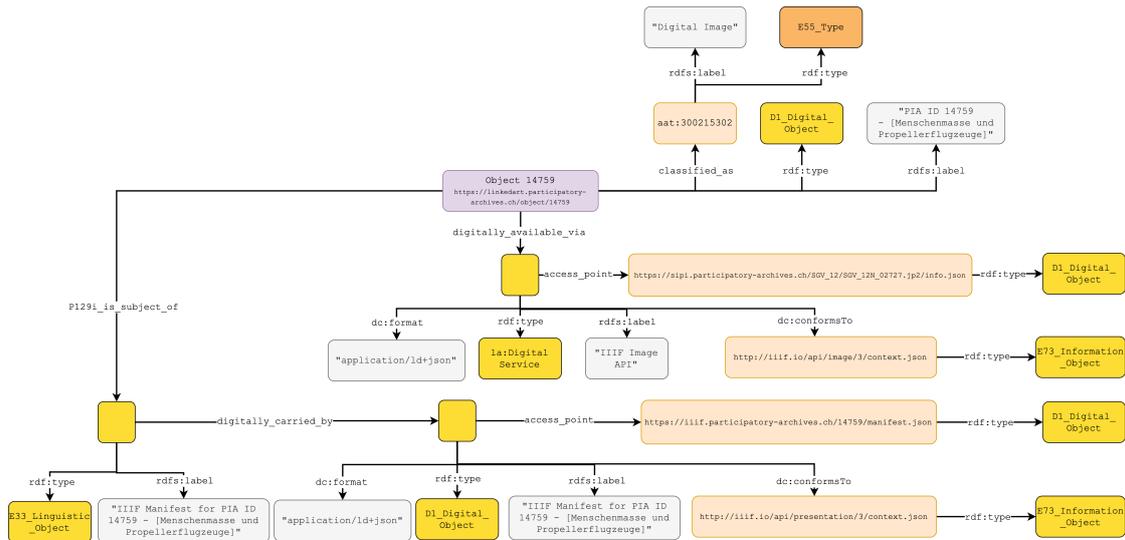


FIGURE 4 : Intégration de ressources/services IIFF au sein d'une modélisation Linked Art

guerre mondiale (<https://www.wikidata.org/entity/Q385425>) ou de l'un de ses homonymes, notamment le poète de la fin du XIXe siècle (<https://www.wikidata.org/entity/Q920515>).

Grâce à l'outil Annotate, permettant de créer des annotations conformes au WADM sur la base de ressources IIFF<sup>16</sup>, il a été assez facile de sélectionner une partie d'image et de l'annoter (cf. Figure 5). Cette annotation a généré le fichier JSON-LD suivant : <https://julsraemy.github.io/annotate/annotations/14759-p1-list.json>.

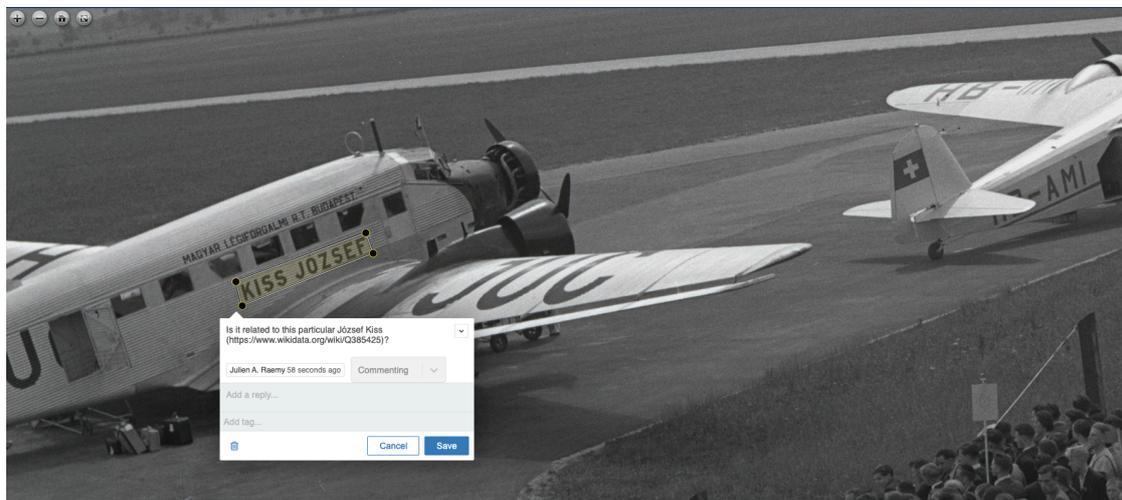


FIGURE 5 : Annotation du Manifeste IIFF 14759 au sein d'Annotate

16. Annotate permet une sélection fine des motivations ainsi que l'ajout de mots-clés associés. Il est donc théoriquement possible de recourir à un vocabulaire contrôlé.

Ce processus d'annotation a surtout été une démonstration de faisabilité et à quel point ces standards permettent de plus facilement échanger des données sur le Web. Pour le projet de recherche PIA, il s'agira, entre autres, de trouver des solutions techniques et organisationnelles pour modérer les flux d'information.

### 5.3. Perspectives

Pour chacun des cas d'utilisation présentés dans cet article, l'un des trois standards LOUD a été davantage sollicité (dans l'ordre : Linked Art, IIF, puis le WADM). Par rapport à une implémentation complète du modèle de données et des composantes LOUD, en plus de la modération des flux mentionnés ci-avant, il y a encore des incertitudes sur la création de modèles et de règles sous-jacentes pouvant accommoder les différentes typologies d'entités. Il faudra donc repenser les processus au sein de notre prototype et voir comment modéliser les cas d'utilisation de la manière la plus pragmatique possible, ce qui est déjà le cas avec la modélisation de la ressource 14759 en Linked Art où des choix ont été opérés pour garder la numérisation et le négatif au sein d'un même fichier pour les garder connectés sans être tout à fait correct d'un point de vue ontologique.

Ce qui est déjà certain c'est que cet écosystème LOUD est déjà un véritable actant - à prendre au sens latourien du terme - dynamique de l'infrastructure et influence grandement la façon dont les informations peuvent être sauvegardées et disséminées mais également à quel point cela va impacter les utilisatrices et utilisateurs des interfaces [30].

Dernièrement, il restera bien évidemment à étudier l'évolutivité d'un tel système et voir à quel point ces standards ouverts et maintenus par des communautés ou bâties autour d'une pratique communautaire peut servir d'exemple dans les HN et ainsi devenir ce que Hyvönen [7] mentionne comme cette troisième génération de portails pouvant résoudre automatiquement des questions de recherche en fonction des contraintes fixées par les scientifiques tout en gardant à l'esprit les besoins et apports d'un plus large public.

## Conclusion

Le point fort du LOUD ne réside pas seulement dans les solutions techniques qu'il peut offrir aux institutions, que ce soit l'amélioration de l'interopérabilité ou ces capacités d'annotation, mais bien dans sa capacité de résoudre des questions fondamentales communautaires et de briser les silos. C'est donc d'une approche qui est autant technique que sociale où le seuil de participation est de soumettre ses cas d'utilisation aux communautés respectives. Il s'agit ici bien des communautés IIF et celle de Linked Art. Quant au WADM, publié sous l'égide du W3C et qui devrait rester stable pour un certain nombre d'années, il est en effet plus compliqué de le modifier ou de pouvoir directement communiquer avec cette instance, mais au cas où il y aurait une modification de celui-ci, il est assez évident que IIF adaptera à nouveau ses spécifications pour y être conforme<sup>17</sup>.

---

17. A savoir qu'une extension des spécifications est toujours permise tant que celle-ci ne devient pas contradictoire avec le standard.

L'adoption de ce socle technologique, autant par les institutions du patrimoine culturel que par les HN, permettra de parler un langage commun, de pouvoir échanger et visualiser les données de manière à tout ce qui est décrit fera parti du web et ne sera ainsi pas seulement de simples pièces échangées grâce au web. Le projet PIA s'efforcera d'ici fin janvier 2025 d'apporter des éléments de réponse et si la participation se fera principalement par l'interface utilisateur, les standards LOUD permettront de plus facilement extraire et réutiliser des métadonnées et données en ne passant que par les API.

## Remerciements

Je tiens à remercier mon directeur de thèse et mes encadrants (PD Dr. Peter Fornaro, Prof. Dr. Walter Leimgruber et Dr. Robert Sanderson), les collègues du laboratoire d'humanités numériques (DHLab) de l'Université de Bâle et toutes les personnes qui participent au projet de recherche PIA, notamment les cinq autres doctorant-e-s (Murielle Cornut, Max Frischknecht, Birgit Huber, Fabienne Lüthi et Florian Spiess) ainsi que notre développeur (Adrian Demleitner).

PIA bénéficie du soutien du fonds national suisse (FNS) dans le cadre du programme interdisciplinaire Sinergia (<https://data.snf.ch/grants/grant/193788>).

## Références

- [1] I. Papadakis, K. Kyprianos, M. Stefanidakis, Linked Data URIs and Libraries : The Story So Far, *D-Lib Magazine* 21 (2015). URL : <http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>. doi :10.1045/may2015-papadakis.
- [2] N. Freire, G. Robson, J. B. Howard, H. Manguinhas, A. Isaac, Cultural heritage metadata aggregation using web technologies : IIF, Sitemaps and Schema.org, *International Journal on Digital Libraries* (2018). URL : <https://doi.org/10.1007/s00799-018-0259-5>. doi :10.1007/s00799-018-0259-5.
- [3] N. Freire, E. Meijers, S. d. Valk, R. Voorburg, A. Isaac, R. Cornelissen, Aggregation of linked data : A case study in the cultural heritage domain, in : 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 522–527. URL : <https://doi.org/10.1109/BigData.2018.8622348>. doi :10.1109/BigData.2018.8622348.
- [4] N. Thalhath, M. Nagamori, T. Sakaguchi, S. Sugimoto, Wikidata Centric Vocabularies and URIs for Linking Data in Semantic Web Driven Digital Curation, in : E. Garoufallou, M.-A. Ovalle-Perandones (Eds.), *Metadata and Semantic Research, Communications in Computer and Information Science*, Springer International Publishing, 2021, pp. 336–344. doi :10.1007/978-3-030-71903-6\_31.
- [5] A. Chardonens, La gestion des données d'autorité archivistiques dans le cadre du Web de données, 2020. URL : <https://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/315804>, publisher : Université libre de Bruxelles.
- [6] S. Münster, F. I. Apollonio, P. Bell, P. Kuroczynski, I. Di Lenardo, F. Rinaudo, R. Tamborino, Digital Cultural Heritage meets Digital Humanities, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W15*

- (2019) 813–820. URL : <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W15/813/2019/>. doi :10.5194/isprs-archives-XLII-2-W15-813-2019.
- [7] E. Hyvönen, Using the Semantic Web in digital humanities : Shift from data publishing to data-analysis and serendipitous knowledge discovery, *Semantic Web* 11 (2020) 187–193. URL : <https://content.iospress.com/articles/semantic-web/sw190386>. doi :10.3233/SW-190386, publisher : IOS Press.
- [8] R. Sanderson, The Importance of being LOUD, 2020. URL : <https://www.slideshare.net/azaroth42/the-importance-of-being-loud>, LODLAM 2020.
- [9] D. Newbury, LOUD : Linked Open Usable Data and linked.art, in : 2018 CIDOC Conference, 2018. URL : [https://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2021/03/CIDOC2018\\_paper\\_153.pdf](https://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2021/03/CIDOC2018_paper_153.pdf).
- [10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3 (2016) 160018. URL : <https://www.nature.com/articles/sdata201618>. doi :10.1038/sdata.2016.18.
- [11] R. Sanderson, Cultural Heritage Research Data Ecosystem, 2020. URL : <https://www.slideshare.net/azaroth42/sanderson-cni-2020-keynote-cultural-heritage-research-data-ecosystem>, CNI Spring 2020 Virtual Meeting.
- [12] S. Snyderman, R. Sanderson, T. Cramer, The International Image Interoperability Framework (IIIF) : A community & technology approach for web-based images, in : Archiving Conference, volume 2015, IS&T, 2015, pp. 16–21. URL : <https://purl.stanford.edu/df650pk4327>.
- [13] R. Robineau, Introduction aux protocoles IIIF, 2019. URL : <https://doi.org/10.5281/zenodo.3760306>.
- [14] R. Sanderson, B. Albritton, R. Schwemmer, H. 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, Association for Computing Machinery, 2011, pp. 175–184. URL : <https://doi.org/10.1145/1998076.1998111>. doi :10.1145/1998076.1998111.
- [15] M. B. Reusser, e-codices : 15 Jahre – eine Erfolgsgeschichte, *ICOMOS – Hefte des Deutschen Nationalkomitees* 77 (2021) 64–76 Seiten. URL : <https://journals.ub.uni-heidelberg.de/index.php/icomoshefte/article/view/85149>. doi :10.11588/IH.2020.1.85149.
- [16] J. v. Zundert, On Not Writing a Review about Mirador : Mirador, IIIF, and the Epistemological Gains of Distributed Digital Scholarly Resources, *Digital Medievalist* 11 (2018) 5. URL : <http://journal.digitalmedievalist.org//articles/10.16995/dm.78/>. doi :10.16995/dm.78.
- [17] S. Cossu, IIIF at the getty : Vision & tactics, 2020. URL : [https://www.cni.org/wp-content/uploads/2020/03/scossu\\_iiif\\_getty\\_vision\\_and\\_tactics\\_cni\\_spring\\_2020.pdf](https://www.cni.org/wp-content/uploads/2020/03/scossu_iiif_getty_vision_and_tactics_cni_spring_2020.pdf), CNI Spring 2020.

- [18] R. Sanderson, Discovery of IIIF resources, 2018. URL : <https://www.slideshare.net/azaroth42/iiif-discovery-walkthrough>, 2018 IIIF Conference.
- [19] M. Appleby, D. Childress, T. Crane, J. Mixer, R. Sanderson, S. Warner, M. Whitaker, IIIF Content State API 1.0, 2022. URL : <https://iiif.io/api/content-state/1.0/>.
- [20] N. Freire, E. Meijers, S. de Valk, J. A. Raemy, A. Isaac, Metadata aggregation via linked data : Results of the europeana common culture project, in : E. Garoufallou, M.-A. Ovalle-Perandones (Eds.), *Metadata and Semantic Research, Communications in Computer and Information Science*, Springer International Publishing, 2021, pp. 383–394. doi :10.1007/978-3-030-71903-6\_35.
- [21] R. Sanderson, P. Ciccarese, H. Van de Sompel, Designing the W3C open annotation data model, in : *Proceedings of the 5th Annual ACM Web Science Conference, WebSci '13*, Association for Computing Machinery, 2013, pp. 366–375. URL : <https://doi.org/10.1145/2464464.2464474>. doi :10.1145/2464464.2464474.
- [22] J. A. Raemy, Applying Effective Data Modelling Approaches for the Creation of a Participatory Archive Platform, 2021. URL : <https://infoscience.epfl.ch/record/291219>.
- [23] E. Delmas-Glass, R. Sanderson, Fostering a community of PHAROS scholars through the adoption of open standards, *Art Libraries Journal* 45 (2020) 19–23. URL : <https://doi.org/10.1017/alj.2019.32>. doi :10.1017/alj.2019.32.
- [24] B. Heinisch, K. Oswald, M. Weißpflug, S. Shuttleworth, G. Belknap, Citizen Humanities, in : K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, K. Wagenknecht (Eds.), *The Science of Citizen Science*, Springer International Publishing, 2021, pp. 97–118. URL : [https://doi.org/10.1007/978-3-030-58278-4\\_6](https://doi.org/10.1007/978-3-030-58278-4_6). doi :10.1007/978-3-030-58278-4\_6.
- [25] A. Demleitner, J. A. Raemy, PIA data model, 2021. URL : <https://github.com/Participatory-Image-Archives/pia-data-model>. doi :10.5281/zenodo.5142605.
- [26] M. Appleby, T. Crane, R. Sanderson, J. Stroop, S. Warner, IIIF image API 3.0, 2020. URL : <https://iiif.io/api/image/3.0/>.
- [27] M. Appleby, T. Crane, R. Sanderson, J. Stroop, S. Warner, IIIF presentation API 3.0, 2020. URL : <https://iiif.io/api/presentation/3.0/>.
- [28] S. Fraser, An introduction to the Madoc Platform, 2019. URL : <https://medium.com/digirati-ch/an-introduction-to-the-madoc-platform-af516a67e0>.
- [29] S. Blickhan, Fun with IIIF, 2022. URL : <https://blog.zooniverse.org/2022/04/20/fun-with-iiif/>.
- [30] B. Latour, *Reassembling the social : an introduction to actor-network-theory*, Clarendon lectures in management studies, Oxford University Press, 2005.