

Ethical and Legal Perspectives on Big Data in Research

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Christophe Olivier Schneble Butz

2024

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

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag

von ErstbetreuerInnen: Prof. Dr. Bernice Elger, Dr. David Shaw

Zweitbetreuer: Prof. Dr. Heiko Schuldt

Externe Expertin: Prof. PD Mira Burri

Basel, den 25.5.2021

Prof. Dr. Marcel Mayor

Table of Contents

ACKNOWLEDGEMENTS.....	5
LIST OF TABLES	6
SUMMARY.....	7
LIST OF ABBREVIATIONS.....	8
CHAPTER 1: INTRODUCTION	9
I. Dissertation Structure	10
II. Big Data the Buzz Word of The Decade	11
III. Research Ethics.....	14
IV. Legal Aspects.....	18
V. Building a Bridge	19
VI. References.....	20
CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGY	24
I. Research Objectives.....	25
II. Methodology	27
III. Limitations.....	27
IV. Individual Contributions.....	28

CHAPTER 3: ALL OUR DATA WILL BE HEALTH DATA ONE DAY: THE NEED FOR UNIVERSAL DATA PROTECTION AND COMPREHENSIVE CONSENT	29
CHAPTER 4: SOCIAL MEDIA TERMS AND CONDITIONS AND INFORMED CONSENT FROM CHILDREN: AN ETHICAL ANALYSIS	45
CHAPTER 5: INFORMED CONSENT IN THE AGE OF BIG DATA - "A CONTRADICTION IN TERMS"? A QUALITATIVE STUDY OF LAWYERS AND DATA PROTECTION OFFICERS	67
CHAPTER 6: DATA PROTECTION LAWS IN THE AGE OF BIG DATA: A QUALITATIVE STUDY WITH LAWYERS AND DATA PROTECTION OFFICERS IN THE UNITED STATES AND SWITZERLAND.....	88
CHAPTER 7: "EIGENTUM" AN PATIENTENDATEN – WER MEINT EIGENTLICH WAS?"	105
CHAPTER 8: SOCIAL-MEDIA UND SOCIAL MESSAGING IM SPITAL	123
CHAPTER 9: THE CAMBRIDGE ANALYTICA AFFAIR AND INTERNET- MEDIATED RESEARCH	137
CHAPTER 10: GOOGLE’S PROJECT NIGHTINGALE HIGHLIGHTS THE NECESSITY OF DATA SCIENCE ETHICS REVIEW	142
CHAPTER 11: DATA PROTECTION DURING THE CORONAVIRUS CRISIS 147	
CHAPTER 12: DISCUSSION.....	153
I. Major Findings.....	154

II. Consent in the Era of Big Data	155
III. The “Ethical Legal Tension”	158
IV. Solution to the Big Data Consent Dilemma	161
V. Towards a Procedural Approach	163
VI. Limitations.....	163
VII. Conclusion and Further Research	164
VIII. References	165
APPENDIX	169
I. Tables	170
II. Interview Guide Swiss and US Expert Interviews.....	173
III. Curriculum Vitae.....	175

“Treat data as you treat any other consumer good. Let, as a general matter people to make whatever choices they want to make, but don’t let them buy exploding toasters.” (USL-13)

Acknowledgements

I would like to thank Helen, my biggest supporter, for the many hours that you had my back and took care of the children, but also made me happy with your music.

To my supervisors and especially to Prof. Bernice Elger, for her support, encouragement, and the possibility to work in many different projects, that broaden my horizon. An enormous thanks goes to Dr. David Shaw for his support, guidance, and many discussions on the subject. I also thank him for allowing me to work on the autonomous self-driving car project. The numerous ideas that we could develop in several opinion pieces, that make ethics tangible. I also thank Prof. Heiko Schuldt for his guidance over the years and his valuable input on computer science matters. Finally, a great thanks goes to Prof. PD Mira Burri for evaluating my thesis.

To my friends and colleagues at the Institute. Andrea for his valuable input on the legal issues in my work. Virgilia for your friendship and numerous coffee-breaks and chats. Finally, I would thank Dr. Tenzin Wangmo, for her precious methodological input, and the train rides to Zurich and Basel, which would have been annoying without the possibility to chat with you.

Finally, a great thanks goes to Prof. Jens Gaab with whom I was allowed to build up the block internet research within the framework of his lecture and on whose full support I could always count on. A great thanks goes to my friend Godefroy, who believed in me and whose great advice and entrepreneurial spirit kept my curious.

List of Tables

Table 1.1 Ethical Principles / dimensions and mechanisms	page 15
Table 3.1. Overview of the most important federal laws on data protection in the United States.	page 34
Table 3.2. Consent based on different health data types related to their temporal origin.	page 41
Table 4.1. Overview of the different Platforms / Apps	page 51
Table 4.2. Scoring the most popular social media apps	page 52
Table 4.3. Constraints of the scoring system.	page 53
Table 11.1. Overview of selected contact tracing apps	page 149
Table App.1 Link to the Terms and Conditions	page 170
Table App.2 How are the Terms and Conditions presented?	page 171

Summary

This PhD thesis focuses on the legal and ethical aspects of Big Data in general and in research. Big data holds the promise of resolving many pressing questions in commercial and research settings. In fact, the use of Big Data to reveal individual or group patterns and the use of predictive analysis are already effective in several areas ranging from building smart hospitals and cities to fighting climate change and to predict individual health related topics. Big data thus touches the core of today's massively interlinked and connected society; this omnipresent use raises societal questions and also affects individual rights, raising many issues for regulators. To assess issues in regulation and identify ethical challenges with the use of Big Data, interviews were conducted in Switzerland and the USA. In Switzerland we interviewed Data-Protection-Officers of all cantons with Universities and University hospitals) and hospital lawyers. The US sample consisted of lawyers with a connection to Big Data or dealing with medical Big Data. The publications that arose from the project can be grouped into the following categories.

A first group of publications deals with the relationship between tech companies and their users and explores how to build a "trusted partnership". A second group has the focus on health data in general and the fact that the distinction between health and non-health data gets more and more blurry. A third group of publications based on the results of our interviews deals with legal and ethical issues on the use of big Data ranging from special hands-on topics of for example data sharing (Social Messaging) in the hospital to discussing the building blocks of data protection law and ownership.

List of Abbreviations

DSG:	Datenschutzgesetz
ECHR:	European Convention on Human Rights
ECtHR:	European Court of Human Rights
EKNZ:	Ethics Committee Northwest/Central Switzerland
ELSI:	Ethical, Legal and Social Implications
EPDG:	Swiss Law on Electronic Health Record (Bundesgesetz über das elektronische Patientendossier)
FADP:	Federal Act on Data Protection
EU:	European Union
FDA:	US Food & Drugs Administration
GDPR:	General Data Protection Regulation
HIPAA:	US Health Insurance Portability and Accountability Act
GDPR:	EU General Data Protection Regulation
IoT:	Internet of Things
IRB:	Institutional Review Board
OECD:	Organization for Economic Co-operation and Development
PIA:	Privacy Impact Assessment
T&C's	Terms and conditions
UDHR:	Universal Declaration of Human Rights
UDHGR:	Universal Declaration on the Human Genome and Human Rights
UK:	United Kingdom
US:	United States

Chapter 1: Introduction

I. Dissertation Structure

Big data holds the promise of resolving many pressing questions in commercial and research settings. In fact, the use of Big Data to reveal individual or group patterns and the use of predictive analysis are already effective in several areas ranging from building smart hospitals and cities to fighting climate change and to predict individual health related topics. Big data thus touches the core of today's massively interlinked and connected society; this omnipresent use raises societal questions and also affects individual rights, raising many issues. This dissertation analyses and gives solutions on how Big Data can be used ethically and how regulators should approach them. This chapter starts by setting out a general introduction to this thesis and the underlying ethical and legal issues of Big Data in research and the commercial sphere.

Chapter 2 gives an overview of the methods used for the research and critically examines them for their strengths and weaknesses. It also provides an overview of the contributions of the researchers and students involved in some of the results of this dissertation

Chapter 3 is composed of a paper which takes a theoretical approach to examine the entangling between legislation and the ever-growing use of data in today's interlinked world. And questions whether the distinction between sensitive and non-sensitive data is still valid before presenting possible solutions to approach the problematic.

Chapter 4 a scoping review of the most popular social-media platforms focuses on the most vulnerable group of users, children. It questions whether today's approach of notice and consent is ethically sound in view of the potential harm that can result of its use. It also links the ethical principles to real-world measures and show potential solution that could lead out of the use – harm dilemma.

Chapter 5 the result of our interview study in the US provides insight how consent mechanisms can still be upheld in times of Big Data and provides solution on how to tackle the "crisis of consent".

Chapter 6 delinates the results from our studies amongst US and Swiss Lawyers shows what current data protection legislation are lacking and question if Big Data exceptionalism could be a solution.

Chapter 7 presents the results of an interview study conducted in Switzerland asks the questions who owns the data and links the findings to the current legal system and provides clarifications whether such a concept could be useful.

Chapter 8 presents an aspect of the results of our interview study amongst data-protection officers and hospital lawyers in Switzerland. Namely the use of social-media and social-messaging in the hospital environment. It analysis the legal background and offers guidelines for a safe use of those tools and technologies.

Chapters 9-11 Are short commentary articles published in various high-impact journals on the governance of Big Data mainly in the field of Big Data research.

II. Big Data the Buzz Word of The Decade

Big Data holds the promise to solve many of today's problems. Starting from climate change (Faghmous & Kumar, 2014) and ranging to the design of smart hospitals (Mertz, 2014). It is present across many disciplines such as social sciences (Kaplan, 2015), to medicine (Hulsen et al., 2019; NIH, 2015) and psychology (Adjerid & Kelley, 2018), natural sciences, engineering and even law. Whilst some of the predictions of a paradigm shift towards a complete hypothesis free¹ and correlation based sciences (Anderson 2008) have been proven wrong, Big Data and its underlying methods have become mainstream in many disciplines.

The emergence of Big Data goes hand in hand with the fact that a lot of data is produced, collected, and curated today. On the one hand, this has to do with the development of higher computing power² and greater storage capacity; on the other hand, omnipresent devices such as smartphones and the Internet of Things allow data to be collected on every single individual. This inevitably leads to many advantages and benefits, but also poses risks to the individual, whether this concerns prediction that reveal details that one would rather keep private or that data is being used to discriminate individuals based on different components (Favaretto et al., 2019). It's thus key to develop a framework on the ethical

¹ In its article Anderson proclaimed the end of hypothesis driven research (Anderson 2008). Although his claim has not be subject to scientific research it has mainly influenced the public discussion and especially in the commercial field fuelled high expectations on the use of Big Data refer back to his claim.

² According to Moore's law the number of transistors in a dense integrated circuit double every two years. With the upcoming development of quantum computers the development could be altered by a magnitude in computer power in the years to come (Ménard et al., 2020)

use of Big Data to have benefit from this technological advancement and in to respect individual rights.

a) The Definition of Big Data

Big data is not a new term and data-linkage is one of the main characteristics has been commonly known since the mid 1990's. The increasing computing power and the evolution of data storages made those technologies accessible to a wider public in the early 2000's. This was when the term Big Data became prominent to a wider public starting with the 3 V's *Volume* (huge amounts), *velocity* (high-speed processing) and *variety* (heterogeneous data) (Laney, 2001; Mayer-Schonberger & Cuiker, 2013; McKinsey Global Institute, 2011). In the following years the definition was slightly modified by adding additional properties to the definition such as *veracity* (IBM, 2012). Common to those definitions however are that they focus on large amount of data stemming from different sources as the definition from various legislative bodies and founding agencies such as the European Commission, the Organization for Economic Co-operation and Development (OECD) or the US-National Science foundation (NSF) shows³. This definition is not free from criticism. Some researchers have though argued that focusing only on attributes causes problems and is too restrictive. Research that does not meet all the criteria could fall through the net even though it clearly qualifies as big Data research. An investigation within the framework of our research project has also clearly shown that the definition in the various disciplines is different, but also personally influenced (Favaretto, de Clercq, et al., 2020). This is problematic, as it means that the Institutional Review Boards (IRB) may lack oversight over such research. Other scholar have argued the of the value resulting of Big data has often been neglected being especially in the commercial field on of the key drivers in promoting Big data (Cate & Mayer-Schonberger, 2013; De Mauro et al., 2016). Others have argued that a definition should focus on the procedural and methodological aspects defining Big data as the information asset characterized by such high volume,

³ The European Commission defines Big Data as: "large amounts of different types of data produced from various types of sources, such as people, machines or sensors. This data includes climate information, satellite imagery, digital pictures and videos, transition records or GPS signals. Big Data may involve personal data: that is, any information relating to an individual, and can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address" (Commission européenne, 2016) The NSF defines Big Data as following: "large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future" (National Science Foundation (NSF), 2014)

velocity, and variety to requiring specific technology and analytical methods for its transformation into values (Stucke & Grunes, 2016).

b) The Problem of Massive, Interlinked Data

Data has become ubiquitous in our every-day life. Many of us use a Smartphone, some other even drive semi-automated vehicles, thus many of us are being constantly monitored by this technology. Common to all those technologies and uses is that they heavily depend on massive amounts of data to either make decisions, or in some other cases data is collected without any purpose for later use. But in other use cases it is not that obvious that data is collected, a person posting comments on Instagram is probably not conscious that his data could be used to predict his mental state (Eichstaedt et al., 2018; Reece & Danforth, 2017). A person ordering a pizza is probably unaware that during fulfilling the order-process big tech players are tracking their behaviour and use that data as asset. Although in many cases, the insights gained may not touch sensitive personal data like as for example, in a much-discussed incident in the United States, where Target, a large chain of discount department stores, used its analytics team to see if it could discover customers' pregnancies through their purchasing patterns so they could market various products (Kashmir, H. 2012).

But in fact every one of us has by now a large digital footprint as Jain and colleague describe in their concept of the digital phenotype (Jain et al., 2015). The digital phenotype, an enlarged notion of the ex-tended phenotype, encompasses the ubiquity of digital technologies and linkage of their data to virtually any other data, possibly resulting in potentially health-relevant data (Mayer et al., 2016). This notion is underpinned by not only the large numbers of studies conducted by universities that take advantage of the ubiquity of massive amounts of publicly available data, but also research involving the use of apps to predict individual behaviour (Harari et al., 2016). The recent scandal surrounding Facebook and Cambridge Analytica (Schneble et al., 2018) shows that these aspects of security and privacy are often not taken into account. Cambridge Analytica, a British consulting firm, was able to collect data from as many as 87 million Facebook users without their consent. Though the information was anonymized and aggregated, the fact that app users were able to consent to the use of their friends' data is very unusual, both in terms of research ethics and social media terms and conditions. With the ongoing digitalisation the

collection of data will not become smaller and the possibility of misuse will only increase. It is thus key that frameworks are being developed that foster a responsible use of data.

III. Research Ethics

Way back biomedical research used to have a bad reputation in terms of involving participants (McCallum et al., 2006) . People were included in medical studies without their consent (Vollmann & Winau, 1996) and their data were shared without their knowledge. Some studies continued even though a treatment had been available discovered. This resulted not only in a great much suffering, but also in lives being deliberately put at risk. By introducing guidelines such as the Nuremberg Code (“The Nuremberg Code,” 1996) or the Helsinki Declaration (JAVA, 2013), as well as medical guidelines such as Good Clinical Practices (GCP) (Brown, 2013) or the Belmont Report (Research, 1978), conditions were created to prevent such suffering in the future. At the heart of all these guidelines are the 4 principles respect for autonomy, beneficence, non-maleficence and justice (Beauchamp & Childress, 2013) of biomedical ethics. Seen in a wider context also non-discrimination, and the right to health. Associated to those ethical principles are Mechanisms to safeguard them in the daily practice. Those principle grew out of biomedical ethics but are now also at the heart of research ethics. Table 1.1 gives an overview of the ethical principles and the mechanism.

Starting from the medical context, research ethics has also established itself in other disciplines. For example, ethical guidelines have been established in psychological research (APA Committee on Human Research, 2013; Swiss Psychological Society, 2003) for several decades and many guidelines on national and international levels have been implemented, but other research fields are also increasingly dealing with the questions of research ethics in the context of their research⁴.

⁴ Research ethics goes beyond the question of scientific integrity, which has always been part of good research.

Table 1.1 Ethical Principles / dimensions and mechanisms

Ethical Principles	Dimensions	Mechanism
Respect for autonomy	Participant has the right to decide to be involved in any kind of research Participant has the right to be informed about goals of a study such as side effects etc. Patient has the right to drop out of a study at any time	Shared decision making Informed Consent Right to withdraw
Justice	Burden and benefits should be equally distributed amongst all groups in a study	Define clear inclusion and exclusion criteria
Beneficence	The study must adhere to the principle of doing good for the	Uphold existing laws
Non-maleficence	Requires that a procedure does not harm the patient involved or others in society	Uphold to existing laws

a) **Human Subject Research and Research Ethics**

Whether or not research is subject to ethics review either by IRB's or Research Ethics Committees (REC's) has been mainly the question if it concerns human subjects and causes harm to the person. In Switzerland research is subject to ethical review upon fulfilment of Article 2⁵ of the Federal Act on Research involving Human Beings (HRA). Research outside this scope is thus not subject to ethical review unless the University has set-up a mandatory IRB. In the US the Code of Federal Regulations 45 CFR 46⁶ is the main instrument. But has also defined broad exemptions that are subject to limited IRB review (National Institutes of Health, 2021). These laws mostly focus on the biomedical context. However, Big data and the associated methodologies have been extensively used also in other disciplines where the concept of human subject has not been known yet (Meatcalf & Crawford, 2016). As Big Data might subsequently cause harm to subjects its paramount to either adapt the scope of those laws or implement other safeguards.

⁵ The scope of the HRA and thus ethical review of such research is mandatory to research concerning human diseases and concerning the structure and function of the human body, which involves a) persons, b) deceased persons, c) embryos and foetuses, d.) biological material, e) health-related personal data

⁶ 45 CFR 46 defines the following criteria when research is subject to IRB-Review: Research, involving a living individual about whom data or biospecimens are obtained/used/studied/analyzed through interaction/intervention, or identifiable, private information is used/studied/analyzed/generated

b) Research Ethics in Age of Big Data

As pointed out earlier in this chapter ethics regulation has its background in biomedical research and the translation to Big data research needs some recalibration of those guidelines. One of the aspects that needs recalibration is the aforementioned concept of human subject research as sole inclusion category to IRB-Review. Big Data research projects – especially if Big data is defined as set of methodologies - finds applications amongst various disciplines many of them which were exempted from ethical review so far. For different reason, first many of them are not considering themselves conducting human subject research (Metcalf & Crawford, 2016) or because they are exempt from current legislation or rarely are direct intervention in life or body (Department of Health and Human Services, 2009). This turn towards the “virtual data subject” results in different notions and mainly intangible harm. For example, disrespecting privacy will not harm individuals on the physical level; rather, violates their rights of informational self-determination. Nevertheless, harm might arise from privacy violations or data breaches or predictive algorithms may have severe effects on individuals. Therefore, it is crucial to adapt notions of autonomy, justice, beneficence and non-maleficence to the needs of Big Data.

Consent:

Consent is one of the most important mechanisms in research ethics. Traditionally getting informed consent involves the researcher proactively explaining the research setting and informing the participant about its fundamental rights as well as potential harm that might result of this research. This is typically done through face-to-face contact either in person or by video-conferencing. Big Data challenges this concept for several reasons: the distance between researchers and data subjects, the multitude of sources and the high number of participants involved might mean that participants are exposed to risks in big data without being informed about the research⁷. Thus, the lack of face-to-face contact is a barrier to obtaining participants’ consent. Furthermore, it is difficult to check whether the participants truly understood the study description. This problem has also been known in Internet mediated research where participants have been presented with an online version of the form and gave consent by clicking the “Agree” button. The issue of consent is an

⁷ Especially harvesting social media data where a lot of data is collected without being truly informed poses a problem and has been subject to intense scholar discourse (Zimmer, 2018). It is especially unclear whether social media data should be regarded as public data or not. The University of Oxford for example has developed a framework that takes into account the if the user is a public person.

unsolved problem, especially also in the commercial environment, where the notice and consent procedure has become widely accepted and Terms and Conditions and Data Policies are the sole mechanism in informing the user. It is thus questionable whether users are truly informed regarding length and technical languages of those documents (Cate & Mayer-Schonberger, 2013). This calls for new ways of implementing this process which will intensively be discussed in Chapter 3 and 5.

Privacy and Confidentiality:

With the almost infinite amount of data about individual, the concept of privacy, and especially the understanding of what is private data has become challenged (Zimmer, 2018). Although there might be consensus that informational self-determination is still important the variety of data that is currently being used might make it difficult to approach this issue with a one fits all approach. For example, social media data is largely perceived as being public data. The caveat lies in the issue that users of such services have rarely all read the terms and conditions and thus it might be questionable if they are fully aware of their data being exposed and used for research or to build predictive models that might pose harm to them. This might be problematic as vulnerable population (for example children) might be exposed to such issues (Schneble et al., 2020a). When research is collecting personal identifiable Information researchers are prompted to develop protocols with sufficient safeguards and maintain confidentiality of research data. Strategies and many legal regulations have therefore proposed to minimize the amount of data to be collected or to rely on anonymization both concepts currently being challenged by Big Data.

A possible way to escape this discussion has been the proposition of Nissenbaum and colleagues which propose a contextual approach to privacy (Nissenbaum, 2017). She claims that privacy is context dependent. Data that might be seen sensitive in one domain might be perceived different in another domain. Another approach is differential privacy that takes advantage of statistical obfuscation to hide the single individual in a data-set (Kairouz et al., 2017).

Anonymity

The issue of anonymity is strongly interconnected with confidentiality and the ambiguity of private vs public data. When data is perceived as public, anonymity might be less important because the data subject should know about the potential issues when sharing the

data. Current guidelines and regulations have taken this into account by either exempting such research from IRB-Review or being not covered by current Dataprotection Laws⁸ (DPO-Laws).

Big Data is also challenging the concept of anonymisation, as it is now easier than ever to re-identify anonymised data. For example, Torra and colleagues (Torra & Navarro-Arribas, 2016) have demonstrated how easy it is to infer a person's identity from just a few criteria. Zimmer and colleagues pointed out how easy it is to re-identify an anonymised social media data set (Zimmer, 2008).

Harm

The use of Big Data also questions the concept of harm. It has been often argued that harm shifted from tangible harm to intangible harm as most of the big data research does not affect the body's integrity. However the reduction to this notion falls too short as Zimmer and colleagues pointed out (Zimmer, 2018). Instead the use of data as for example the use of social media data out of its own primary context can be seen as an invasion to the persons dignity. Carpenter and Dittrich have therefore introduced the notion of human-harming research. They argue that "distance" between themselves and their subjects might have an effect on the perception of the human subject. In order to overcome this issue, they propose a risk-based assessment of research. Although harm might not be bodily its notion will be subject to more in-depth research in the future a notion that is also underpinned by the various legal definition of harm (McMahon et al., 2020).

IV. Legal Aspects

Ensuring the protection of privacy and compliance has become a central issue for researchers using big data. However data protection law is often perceived as very complex and difficult to interpret putting additional burden on researcher (Martani et al., 2020). Particularly in the area of sensitive data, data protection is becoming increasingly important in the research environment and has been part of ongoing discussion. Data-driven initiatives, such as the SPHN (Meier-Abt et al., 2018) or the human genome project (Collins et al., 2003), the human brain project (Amunts et al., 2016) as well as the increased inclusion of clinical data in research have contributed significantly to this development. Although

⁸ Many Dataprotection laws as for example as the Swiss Federal Data Protection Legislation apply only to personal identifiable data. Whether or not anonymised Data is subject to the scope is debated highly amongst law scholars (Martani et al., 2020).

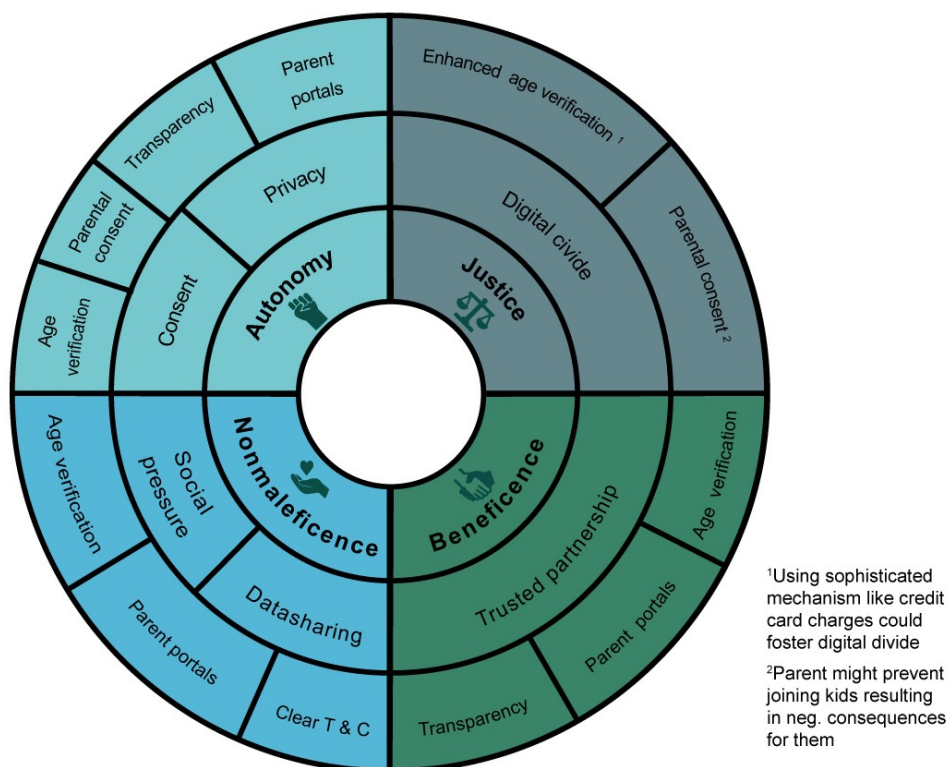
Switzerland's data-protection setting foresees an umbrella kind of overarching data protection regulation. Many areas especially in the research field are however also covered by sectorial regulations making the delineation of which legislations applies difficult (Martani et al., 2020). Research in psychology for example falls under the HRA if it concerns human diseases or the structure and function of the human body (Art. 2 HRA) any other research in psychology not being subject to the scope of Art. 2 HRA is subsequently governed by the Federal Data Protection Regulation.

In addition to the actual genuine questions of pure data protection, there are also questions of ownership of data. Although there is no actual ownership of data from a legal point of view, this question also arises with regards to the commercial value of data (Amstutz, 2018). Research however has shown that there is still ambiguity on this issue (Schneble, Widmer, et al. 2021). Especially researchers (as the data producers) might have a different view on this topic either as they perceive some entitlement towards data that they manage and clashes with the view of the patients that feel that their data should be subject to their control.

V. Building a Bridge

As discussed before Big Data needs a recalibration of well-established concepts in research ethics and data protection law. Research on the ethics of big data have also shown that researchers might also not be aware that they are actually conducting human subject research and thus neglect the ethical issue (Favaretto, De Clercq, Briel, et al., 2020). There is the need to develop frameworks that look at the cross functional pipeline of such Big Data research projects. Figure 1.1 illustrates the various levels which are touched by Big Data. On the highest level of abstraction the basic ethical principles group the different aspects on the second level they are matched with the different mechanism to safeguard them and finally in the third level concrete implementations of those mechanisms need to be in place. Current guidelines stick mostly to one of those levels either they provide to general questions (Markham & Buchanan, 2012) or they are tight to a specific discipline and thus are too narrow as to serve as general model (British Psychological Society et al., 2013).

Figure 1.1) Different levels of Abstraction in Big Data The example of Social Media (Schneble 2021)



This figure maps the four biomedical ethical principles to issues arising from the use of social media and links them to possible fields of actions.

VI. References

- Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*. <https://doi.org/10.1037/amp0000190>
- Amstutz, M. (2018). Dateneigentum. *Archiv Für Die Civilistische Praxis*, 218(2–4).
- Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., & Lippert, T. (2016). The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain. In *Neuron* (Vol. 92, Issue 3). <https://doi.org/10.1016/j.neuron.2016.10.046>
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. <https://www.wired.com/2008/06/pb-theory/>
- APA Committee on Human Research. (2013). *Guidelines for Ethical Conduct of Behavioral Projects Involving Human Participants by High School Students*.
- Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics*. Oxford University Press.

-
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67–73. <https://doi.org/10.1093/idpl/ipt005>
- Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. In *Science* (Vol. 300, Issue 5617). <https://doi.org/10.1126/science.1084564>
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. In *Library Review* (Vol. 65, Issue 3). <https://doi.org/10.1108/LR-06-2015-0061>
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44), 11203–11208. <https://doi.org/10.1073/PNAS.1802331115>
- Faghmous, J. H., & Kumar, V. (2014). A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data*, 2(3), 155–163. <https://doi.org/10.1089/big.2014.0026>
- Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0177-4>
- Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science. *Perspectives on Psychological Science*, 11(6), 838–854. <https://doi.org/10.1177/17456916166650285>
- Hulsen, T., Januar, S. S., Moody, A. R., Karnes, J. H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D. A., & McKinney, E. F. (2019). From big data to precision medicine. In *Frontiers in Medicine* (Vol. 6, Issue MAR). <https://doi.org/10.3389/fmed.2019.00034>
- IBM. (2012). IBM Big Data Success Stories. IBM Big Data Success Stories.
- Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature Biotechnology*, 33(5), 462–463. <https://doi.org/10.1038/nbt.3223>
- JAVA. (2013). Declaration of Helsinki World Medical Association Declaration of Helsinki. *Bulletin of the World Health Organization*, 79(4).
- Kairouz, P., Oh, S., & Viswanath, P. (2017). The Composition Theorem for Differential Privacy. *IEEE Transactions on Information Theory*, 63(6). <https://doi.org/10.1109/TIT.2017.2685505>
- Kaplan, F. (2015). A Map for Big Data Research in Digital Humanities. *Frontiers in Digital Humanities*, 2. <https://doi.org/10.3389/fdigh.2015.00001>
- Kashmir, A. (2012). How target figured out a teen girl was pregnant before her father did. *Forbes*. <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and

-
- Variety. Application Delivery Strategies, 949(February 2001).
- Martani, A., Egli, P., Widmer, M., & Elger, B. S. (2020). Data protection and biomedical research in Switzerland: setting the record straight. *Swiss Medical Weekly*, 150. <https://doi.org/10.4414/smw.2020.20332>
- Mayer-Schonberger, V., & Cuiquer, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (John Murra).
- Mayer, M. A., Fernández-Luque, L., & Leis, A. (2016). Big Data For Health Through Social Media. *Participatory Health Through Social Media*, 67–82. <https://doi.org/10.1016/B978-0-12-809269-9.00005-0>
- McCallum, J. M., Arekere, D. M., Green, B. L., Katz, R. V., & Rivers, B. M. (2006). Awareness and knowledge of the U.S. Public Health Service syphilis study at Tuskegee: Implications for biomedical research. In *Journal of Health Care for the Poor and Underserved* (Vol. 17, Issue 4). <https://doi.org/10.1353/hpu.2006.0130>
- McKinsey Global Institute. (2011). *Big Data: The Next Frontier for Innovation, Competition and Productivity*”.
- McMahon, A., Buyx, A., & Prainsack, B. (2020). Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond. *Medical Law Review*, 28(1). <https://doi.org/10.1093/medlaw/fwz016>
- Meier-Abt, P., Lawrence, A. K., Selter, L., Vayena, E., & Schwede, T. (2018). The Swiss Approach to Precision Medicine. *Swiss Medical Weekly*. <https://smw.ch/en/op-eds/post/the-swiss-approach-to-precision-medicine/>.
- Ménard, A., Ostojic, I., Patel, M., & Volz, D. (2020). A game plan for quantum computing. *McKinsey Quarterly*, February 1.
- Mertz, L. (2014). Saving lives and money with smarter hospitals: Streaming analytics, other new tech help to balance costs and benefits. *IEEE Pulse*, 5(6), 33–36. <https://doi.org/10.1109/MPUL.2014.2355306>
- Metcalfe, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data and Society*, 3(1). <https://doi.org/10.1177/2053951716650211>
- National Institutes of Health. (2021). Definition of Human Subjects Research. <https://grants.nih.gov/policy/humansubjects/research.htm>
- NIH. (2015). Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director – Building a Research Foundation for 21st Century Medicine. <http://acd.od.nih.gov/reports/DRAFT-PMI-WG-Report-9-11-2015-508.pdf>.
- Nissenbaum, H. (2017). Deregulating Collection: Must Privacy Give Way to Use Regulation? *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- Research, T. N. C. for the P. of H. S. of B. and B. (1978). Belmont Report.

-
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579.
<https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020a). All our data will be health data one day: the need for universal data protection and comprehensive consent (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/16879>
- Swiss Psychological Society. (2003). Ethische Richtlinien.
https://www.swisspsychologicalsociety.ch/fileadmin/user_upload/PDF-Dateien/Ethic_Guidelines/DE/d3-ethische_richtlinien18.dt_.pdf
- The Nuremberg Code. (1996). *JAMA: The Journal of the American Medical Association*, 276(20). <https://doi.org/10.1001/jama.1996.03540200077043>
- Torra, V., & Navarro-Arribas, G. (2016). Big Data Privacy and Anonymization. In A. Lehmann, D. Whitehouse, S. Fischer-Hübner, L. Fritsch, & C. Raab (Eds.), *Privacy and Identity Management. Facing up to Next Steps: 11th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2.2 International Summer School, Karlstad, Sweden, August 21-26, 2016, Revised Selected Papers* (pp. 15–26). Springer International Publishing. https://doi.org/10.1007/978-3-319-55783-0_2
- Vollmann, J., & Winau, R. (1996). Informed consent in human experimentation before the Nuremberg code. In *British Medical Journal* (Vol. 313, Issue 7070).
<https://doi.org/10.1136/bmj.313.7070.1445>
- Zimmer, M. (2008). On the “Anonymity” of the Facebook Dataset.
<https://michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook-dataset/>
- Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media and Society*, 4(2).
<https://doi.org/10.1177/2056305118768300>

Chapter 2: Research Objectives and Methodology

I. Research Objectives

The research presented in this thesis has been part of a larger research project with the title *Regulating Big Data Research: A new frontier*. The project itself is part of the larger National Research Program (NRP75 Big Data). The aim of the study was to analyse the ethical and legal challenges related to the research and commercial use of Big Data. The research project was subdivided into two parts. The first part was an interview study by academic scholars in the fields of psychology and sociology who performed research using Big Data methods. The second part (this thesis) was an interview study with data protection officers and data lawyers from Switzerland and the United States.

This PhD thesis focuses on the legal and ethical aspects of Big Data in general and in research. Big data holds the promise of resolving many pressing questions in commercial and research settings. In fact, the use of Big Data to reveal individual or group patterns and the use of predictive analysis are already effective in several areas, ranging from building smart hospitals and cities to fighting climate change and predicting individual health-related topics. Thus, Big Data touches the core of today's massively interlinked and connected society; this omnipresent use raises societal questions and affects individual rights, raising many issues for regulators.

More specifically, the thesis investigates the following questions.

a) Is the distinction between sensitive and nonsensitive data still appropriate, and how should the temporality of data be handled? (Chapter3)

The tremendous growth in data that are collected, and their interlinkages are enabling more predictions of individuals' behaviour, health status and diseases. Legislation in many countries treats health-related data as a special sensitive type of data. However, today's massive linkage of data could transform "non-health" data into sensitive health data. In *chapter 3*, we argue that the notion of health data should be broadened. It should also take into account past and future health data and indirect, inferred and invisible health data. We also lay out the ethical and legal implications of our model.

b) How can users be informed seriously, especially children? (Chapter 4)

Terms and conditions define the relationship between social media companies and users. However, these legal agreements are long and written in complex language. Children interacting from a young age with social media gives companies large amounts of data,

resulting in longitudinal data sets that most researchers can only dream of. Children's use of social media is highly relevant to their mental and physical health for two reasons: their health can be adversely affected by social media, and their data can be used to conduct health research. Chapter 4 offers an ethical analysis of how the most common social media apps/services inform users and obtain their consent regarding privacy and other issues. The chapter also discusses how lessons from research ethics can lead to trusted partnerships between users and social media companies, with a focus on children who represent a vulnerable group amongst users of those platforms.

c) What are the current challenges of consent, and is there an alternative to consent? (Chapter 5)

Consent has been the main pillar of research ethics and is closely linked to respect for autonomy, as discussed in Section III of *chapter 1*. Traditional methods of obtaining consent have been questioned in the realm of Big Data. The aim of *chapter 5* was to investigate this issue present future-proof models.

d) Regulate the use rather than collection? (Chapter 6)

A vivid discussion has occurred over whether the use (Big Data exceptionalism) or the collection of data should be regulated. Whereas opponents of Big Data exceptionalism argue that the collection, not the use, of data should be regulated, others still believe that regulating the collection is the best approach. The aim of *chapter 6* was to investigate those positions based on the experiences of our interviews and to line out the implications.

e) What does data-ownership mean? (Chapter 7)

Data are becoming increasingly important in the healthcare sector. In the context of the Big Data Research Ethics research project of the NRP 75 Big Data, cantonal data protection officers and hospital lawyers/data protection officers were interviewed. One topic that came up in the interviews was a discussion of nontechnical "data ownership", especially in the context of personal patient data. It became apparent that this term is blurred and unclear and that various misunderstandings exist in this regard. Therefore, this article analyses the question of what stakeholders understand by "data ownership" and how these issues are reflected by the law and aims to clarify certain misunderstandings that often exist in this context for non-lawyers.

f) What are possible approaches to governing the use of Big Data? Some examples form practice. (Chapter 7-11)

Governance on the use of Big Data has become increasingly important. Clear guidelines prevent harm from resulting from the use of Big Data, such as identifying private information that a participant or user wants to keep secret. On the flipside, clear governance provides tools to oversee data science projects. Therefore, the aim of *Chapters 7-11* was to implement and showcase such guidelines.

II. Methodology

To meet the objectives outlined in Section I, this thesis followed a mixed-methods design. First, *Chapter 4* used thematic analysis to perform an ethical analysis of the social media terms and conditions.

Second, an interview study was conducted to grasp the perspectives of lawyers in Big Data research (*Chapters 5, 6, 7, and 8*). The interviews were conducted between January 2018 and August 2019. They comprise 35 semistructured interviews with (hospital-) lawyers/DPO Switzerland (n=20) and lawyers and law scholars in the United States (n=15) who were selected systematically and through snowballing.

The specific methods used for the different articles are described in detail in the methods sections of the following chapters and are not presented here.

III. Limitations

This section addresses the limitations of the interview study. Additional limitations are discussed in Section VI of *chapter 12*. For this study, interviews were conducted amongst Swiss cantonal data protection officers and hospital lawyers and the comparative group in the United States consisting of lawyers and ethical/legal scholars. Whereas the Swiss sample covers a majority of the DPO in cantons with research facilities, universities and university hospitals, the US sample's findings are less encompassing for several reasons. The US field of privacy and data protection law is heterogeneous. First, because it is sectorial, overseeing all different sectors in such a study is difficult. Thus, the study addressed general aspects and focused on the issue introduced by big tech companies and the distinct aspects of health data. Second, no direct privacy law exists at the federal level, and many states have, thus, enacted special laws.

IV. Individual Contributions

This study was conceived and supervised by Professor Dr med. Bernice Simone Elger, Head of the Institute for Biomedical Ethics (IBMB) at the University of Basel. This study was part of a larger project within the scope of NRP 75 on Big Data founded by the Swiss National Foundation (SNF, Grant No: 407540_167211). Dr T. Wangmo and F. Zimmerman and myself prepared the submission to obtain ethical approval from the competent Cantonal Ethics Committee (Ethics Committee northwest/central Switzerland (EKNZ)) and drafted the interview guide. I developed the protocol for participant recruitment. I also recruited the participants and transcribed the interviews. Qualitative data analysis was carried out by Dr D. Shaw and myself. I devised the conceptualization, development and writing of each paper included in this dissertation with the revision, supervision and approval of Dr. D. Shaw, Prof. Dr. med. B. Elger and the other co-authors. *Chapter 7* and *9* were co-authored with two lawyers – Andrea Martani for *chapter 8* and Dr Michael Widmer for *chapter 7* – who co-authored on the legal parts of the paper.

Chapter 3: All Our Data will be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent

Full Reference: Schneble CO, Elger BS, Shaw DM. All Our Data Will Be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent. Journal of Medical Internet Research, 2020;22(5): e16879, DOI: 10.2196/16879, PMID: 32463372, PMCID: 7290498

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, Bernice Simone Elger^{1,2}, MD, PhD, David Martin Shaw¹, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²University Center for Legal Medicine, University of Geneva, Switzerland

I. Abstract

Tremendous growth in the types of data that are collected and their interlinkage are enabling more predictions of individuals' behavior, health status, and diseases. Legislation in many countries treats health-related data as a special sensitive kind of data. Today's massive linkage of data, however, could transform "nonhealth" data into sensitive health data. In this paper, we argue that the notion of health data should be broadened and should also take into account past and future health data and indirect, inferred, and invisible health data. We also lay out the ethical and legal implications of our model.

II. Introduction

a) Background

Data intensive software, such as social media, wellness, and mobile health (mHealth) apps, have become ubiquitous in everyday life and are frequently used in a variety of situations. Years ago, social media networks were mostly accessed from traditional computers, but the rising use of smartphones and apps to access those networks has opened a Pandora's Box regarding data collection, including geolocation, motion data, health-related data, and behavioral data (Ienca et al., 2018). The collection of additional behavioral data about users was initially very limited, and only a fraction of basic data was collected (eg, IP address, operating system, and browser version). In contrast, current apps on smartphones have begun continuous monitoring of users by harvesting geolocation and motion data, and thus, they have the ability to infer users' physical and mental health states, for example, to detect signs of depression (Reece & Danforth, 2017; Saeb et al., 2015) and predict their next likely location (Do & Gatica-Perez, 2014; Cao & Lin, 2017). Moreover, app companies have collected a tremendous amount of data on individuals' public and private activities in the digital world, which are being reused not only for the sake of their primary platforms, but also in other lucrative business sectors, such as robotics, life sciences, car manufacturing, and health data provision. Previously, users were able to simply opt out of these services, but this is becoming increasingly challenging nowadays given the monopoly market structure instilled by the companies that drive digital transformation. Indeed, customers are increasingly forced to use these services because they either do not have any other equivalent alternatives in terms of services provided or they are influenced by their peers or parents (in case of children) to use the services. Sometimes these companies nudge users with marketing strategies, such as substantial advantages and discounts offered only

on these platforms. This proves to be problematic since to use these now important services, at least to some extent, users have to consent to some mandatory data sharing (Kaplan, 2016) and consequently expose their privacy.

b) Ambiguous Terms and Conditions

In view of the aforementioned facts, it is vital that the relationship between users and companies is transparent and regulated. This relationship is currently mostly defined in the terms and conditions (T&Cs), terms of service, and data privacy notices, which are unfortunately lacking in several aspects in term of enabling potential users to make an informed decision when signing up for a service. For instance, users are not warned about possible harms that might result from their activities on the platforms (eg, linkage of several anonymous data sources could lead to reidentification of otherwise anonymous datasets and lack of awareness of secondary use could undermine user privacy and confidentiality). Additionally, the information provided in the T&Cs of different platforms is not reader-friendly and not succinctly summarized to nudge users to read them thoroughly. They are also not harmonized in the sense that each platform has its own implementation or they are simply not prominent enough during the process of signing up for the service (Cate & Mayer-Schonberger, 2013; UK Children’s Commissioner, 2018). Another weakness of T&Cs is that they do not make it clear that social media and the linkage of several independent unique databases can yield health data. Health data represent a special data category (General Data Protection Regulation, 2016), requiring special security and privacy policies for governance. Indeed, many international legislations define health data as special data needing more protection than “usual” nonhealth-related data. The flipside of current legislation is that nonhealth-related data is subject to less strict governance. In the era where a digital phenotype (S. H. Jain et al., 2015) is emerging, data linkage can be very predictive and can be used to, for example, derive personal traits (Kosinski et al., 2013), predict psychosomatic diseases (Reece & Danforth, 2017), and obtain other types of behavioral information.

c) Aim of the Paper

This paper presents a new approach for considering data, with the four categories of direct, indirect, inferred, and invisible health data, and suggests different types of possible consent frameworks that are up to this challenge, especially as data might not be conceived as distinct health data when produced. We first describe the already recognized categories of

direct and indirect health data and then present the other two categories, describe the legal framework in the European Union and United States, and explore the different potential consent mechanisms and their suitability for these four categories of data.

III. Ubiquity of Health Data

a) Evolution of Health Data

Until the end of the last decade, health data were easily defined, and they included medical records, diagnostic images, laboratory testing data, and data produced by biomedical or clinical means. However, as rightly pointed out by Vayena et al (Vayena & Blasimme, 2017), the notion of what is considered health data has considerably evolved. So-called biomedical Big Data nowadays ranges from data produced by health services, public health activities, and biomedical research to data registering exposure to environmental factors, such as sunlight and pollution, or data revealing lifestyle, socioeconomic conditions, and behavioral patterns, such as those from wellness and fitness apps, social media, and wearable devices. There is thus a paradigm shift from the notion of individual data producers and distinct categories to a more complex notion of a data ecosystem (Vayena & Blasimme, 2018).

b) Implications of Massively Interlinked Data

The massive amount of data produced and interlinked has an effect on the characterization of individuals today, including their behavioral profile. Jain et al (S. H. Jain et al., 2015) developed the concept of the digital phenotype. The digital phenotype, an enlarged notion of the extended phenotype, encompasses the ubiquity of digital technologies and linkage of their data to virtually any other data, possibly resulting in potentially health-relevant data (Mayer et al., 2016). This notion is underpinned by not only the large numbers of studies conducted by universities that take advantage of the ubiquity of massive amounts of publicly available data, but also research involving the use of apps to predict individual behavior (Harari et al., 2016). Further complicating these issues is the possibility that other types of data could be health data one day (P. Jain et al., 2016). For example, some social media data concern exercise, which is highly relevant to health. Most people would agree that exercise data are health-related data and represent an example of indirect health data (S. H. Jain et al., 2015). However, it is less obvious that other data, such as address and shopping data, location data, smart home data (eg, Amazon Alex and Siri data), smart car

data, and articles shared on social media, can be combined with other datasets to infer health data [1,11], and this needs to be made more transparent in the future. It is already possible to use data to infer the degree of exposure to pollution and its likely health effects over long periods, and driving behavior data, such as acceleration patterns, can indicate the risks people take, which again could be used to infer their health conditions when combined with other elements.

c) Digitalization of Past Paper-Based Data

At another level, digitalization of past paper-based data could also yield direct or indirect health data (either through digitalization of paper medical records or fitness diaries), and future technologies, such as machine learning, may be able to identify health-relevant uses for data that have not yet even been conceived.

Thus, the days of health data as a distinct category are numbered, and soon, we will have not only direct health data, but also what we term indirect health data, inferred health data, and currently invisible health data (where the relevance to health might be perceptible only by machine intelligence in the future). This new understanding indicates the need for new governance, compliance, and regulatory mechanisms to handle data, protect individuals' privacy, and uphold the security of such new sensitive data. We briefly examine the current legislation that is relevant to these different categories of health data before continuing our ethical analysis.

d) Legal Situation in the United States and European Union

In contrast to the European Union, the United States is not subject to a single overarching data protection law. Data protection issues are implemented at the federal and state level. Table 3.1 summarizes some of the major federal laws that deal with data protection issues.

Table 3.1. Overview of the most important federal laws on data protection in the United States.

Law	Scope	Main Points
Gramm Leach Billey Act	Governs protection of the personal information in the hands of banks, insurances companies, and other companies in the financial service industry.	Addresses “nonpublic information” (NPI) that institutions collect from their customers in connection with the provision of services. Imposes requirements for securing NPI, restricting disclosures, and using NPI. Obligation to notify customers when NPI is improperly exposed.
Fair Credit Reporting Act	Federal law regulating the collection of consumer credit information and access to credit card reports.	Governs how credit bureaus can collect and share information about individual consumers. Businesses check credit reports for many purposes, such as deciding whether to provide a loan or sell insurance to a consumer. This act also gives consumers certain rights, including free access to their own credit reports.
Health Information Portability Act	Protects information held by covered entities concerning health status, provision of health care, or payment for health care.	Breach notification. Data handling by covered entities and definition of safeguards.
Telephone Consumer Protection Act	The order is relevant to any company that uses automated technology to phone or send text messages to consumers.	Regulates telemarketing and forces companies to respect do not call registries.
Family Educational Rights and Privacy Act	Federal law that protects the privacy of student education records.	Offers students the right to correct information about themselves.

In the United States, some privacy frameworks provide a different definition of personal data and the notion of sensitive personal data varies among several federal state laws, with the Californian legislation being the most comprehensive, and among different economic sectors. Health data regulation is mostly included in the Health Insurance Portability and Accountability Act (HIPAA), which only covers entities that are directly related to health care operations, such as health care providers, health plans, and health care clearing houses, and the statement is as follows: “Except as otherwise provided, the standards,

requirements, and implementation specifications adopted under this subchapter apply to the following entities: (1) a health plan; (2) a health care clearinghouse; and (3) a health care provider who transmits any health information in electronic form in connection with a transaction covered by this subchapter” (45 CFR § 160.103; 45th Code of Federal Regulations, Office of the Federal Register, United States of America). When data leave these entities, the imposed safeguards by HIPAA do not apply anymore, resulting in less strict regulations. A further guidance issued by the US Food and Drug Administration (FDA) concerns mHealth apps (Vayena & Blasimme, 2017). It addresses, for example, apps that pose a high risk to the public. Apps only fall within its scope if they transform a mobile phone or any other electronic device into a medical device. As the FDA acknowledges in its guidance document, it does not address a substantial number of health data collectors, such as wellbeing apps; websites, especially patient centered portals like PatientsLikeMe (Vayena & Blasimme, 2018); and social networks, and thus, it excludes most indirect, inferred, and invisible health data, which subsequently are subject to the US Federal Trade Commission guidance, resulting in lower safeguards of potentially highly personal data (Federal Trade Commission (FTC), 2016).

In contrast, the European Union General Data Protection Regulation (GDPR) offers a comprehensive framework for any kind of personal data and adds different notions to, for example, health or research data. The GDPR treats health data as a special category of data, which is sensitive by its nature, along with several other types, and the statement is as follows: “Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.” (Article 9, section 1; General Data Protection Regulation, The European Union). Two key points are illustrated by this quotation. First, it is clear that social media data could reveal any of these different sensitive types of data. Second, the term “health data” is not actually mentioned; the phrase used is “data concerning health.” This opens the door to indirect and inferred health data falling within the scope of the GDPR.

Processing of health data is prohibited unless exceptions apply, and one of them is the provision of the individual’s explicit consent. The collection of consent from the data subject remains one of the most common exceptions that organizations processing health data

will be able to rely on when it has been explicitly provided and the purpose for processing the data has been explicitly defined.

Comparing the two legislations (EU and US), it can be said that most US regulations on health data remain at the national or state level, such as the HIPAA in the US, which is tightly attached to parts of the health system, such as hospitals, health insurance companies, and pharmacies. Other important data stemming from social networks and other providers are not covered by many of these regulations because they are out of their scope. Furthermore, there are no specific regulations other than broad principles, such as the fair information principles, issued by the US trade commission. In contrast, the GDPR offers at least some boundaries, but there are a number of serious questions on how to interpret the approach of the GDPR to the regulation of social media data. Our broadening of the notion of health data implies both that higher safeguards are applied to nonspecial categories of data and that the lawful grounds for processing such kinds of data need to take care of this situation, and imposing implicit consent could be one approach. As today's consent mechanisms are unable to handle this extra burden, there is an urgent need to foster the development of an alternative consent mechanism. Some examples are delineated later in this article.

The issue of different approaches to data protection has been the subject of disagreement in the literature. Cate et al, who are opponents of a more liberal approach, have argued for self-regulation of data protection principles (Cate et al., 2013), shifting the burden to users. In view of the recent scandals involving major data intensive companies, this view seems rather inappropriate and neglects the ethical responsibilities of companies toward their customers (Schneble et al., 2018; Kelley et al., 2009).

In contrast, authors like McDermont et al advocate data protection as a human right, with its underpinning ethical principles of privacy, transparency, autonomy, and nondiscrimination. These principles are particularly important in light of the increasing use of large amounts of data and algorithmic prediction (McDermott, 2017). This is also highlighted by Wachter et al, who called for a new right of reasonable inferences to close the accountability gap currently posed by “high-risk inferences,” especially regarding predictions drawn from Big Data analytics with low verifiability and thus possible damaging effects on individuals (Wachter & Mittelstadt, 2019).

IV. Extended Notions of Health Data and Consent

Until now, we have assumed that data used in research are primarily generated prospectively (ie, right now or recently) and have excluded data from the more distant past. However, research projects, such as the Time Machine project (EPFL, 2019), aim to generate digital copies of vast amounts of past paper-based data. Indeed, digitization of past data is progressing rapidly, for example, in the digital humanities and in the digitalization of large amounts of business and health data (Jones & Nevell, 2016). A new notion is thus introduced to the concept of consent. As today's consent is based on the assumption that the future use of data must be regulated, the linking of old or past data in connection with digitalization would also require consent when an identified or identifiable person is concerned. Consent will therefore have to deal with the past, present, and future use of both past and prospective data.

Consent to use pre-existing data is dealt with in a variety of ways. In medicine, research participants are sometimes invited to give "broad consent" to future reprocessing of data, which is subject to review by a research ethics committee. Such consent is necessarily broad because those giving it will often have no idea of the specific research projects that might use their data in the future. Other models require recontacting participants in order to obtain specific consent for each future project. More radical is the concept of "data donation," where people grant access to their data under very few limited conditions, if any (Shaw, 2019). However, all of these models are derived from health care and medical research. In the wider context of social media, financial, and location data, consent is based on the initial agreement to the T&Cs as described above and is mostly of a commercial background. Despite the fact that the data could be used in a myriad of ways, even if not ultimately health related, this consent is often entirely uninformed. This is even the case for data that are currently being generated. Awareness and transparency are even lower in terms of possible uses of past paper-based data that are digitalized or future data that might seem irrelevant now but could yield highly relevant health data when combined with other datasets. Current consent for data sharing is to a large extent blind owing to its broad nature. Consent systems actually need to look far back and far forward, as well as in close detail at the present. In essence, consent must be capable of time travel, just as data are capable of time travel.

V. Toward Comprehensive Consent in a Hyperconnected World

Several scholars have pointed out that traditional models of obtaining consent have reached their limits in today's highly data-driven and data-intensive research, and this articulates the need for new forms for obtaining participant consent (Mostert et al., 2016; Ioannidis, 2013). Concretely, the traditional form of obtaining consent by individually informing participants about their rights and protections is practically impossible in such environments (Christen et al., 2016; Kaye et al., 2015; Jacobs & Popma, 2019) owing to the sheer scale and challenges associated with such an endeavor. Several scholars have proposed possible ways to tackle this issue, ranging from information technology-based systems like dynamic consent (Williams et al., 2015), which offer a better way to inform and maintain a relationship between researchers and participants, to stewardship-based solutions, such as those where a community-based approach assures data governance (Blasimme et al., 2018), and radical solutions like data donation (Shaw, 2019). However, most of today's research projects simply use a digitized version of traditional consent procedures.

As we have already pointed out, past data will become increasingly health relevant. This applies particularly to public and commercial research where it is expected that increasingly more research will be data driven and data will stem from many disparate data sources, including commercial sources. There is thus an urgent need for a better approach of informing customers in a truly informative way. However, as stated above (P. Jain et al., 2016), the current approaches are ineffective as T&Cs are too long and written in a too complicated way, undermining what is at the heart of genuinely informed consent, namely the prevention and hold up of basic ethical principles (Federal Trade Commission (FTC), 2016; Cate et al., 2013). A possible way to tackle this is to introduce harmonized T&Cs, which would need stronger government interventions. Further possibilities are to move toward comic-based consent as developed by Brunschwig for contract law (Brunschwig, 2001) and to implement "nutrition label-like" consent (Kelley et al., 2009) for T&Cs.

The four distinct categories of direct, indirect, inferred, and invisible health data mentioned earlier in this article may each require a differentiated consent solution, and the past, present, and future aspects of consent for the use of Big Data complicate the situation further.

Direct health data are most easily governed, although they have their own set of challenges. A specific consent system may seem simple but can impose substantial limitations on

researchers and prove very burdensome for patients. Broad consent poses its own set of problems, particularly in terms of future data linkage. Even with specific consent to use an individual's data in a particular project, researchers might also want broad consent to access a participant's entire medical history and link it to medical records to facilitate follow-up, meaning that consent is provided for health data generated in not only the project itself (the present), but also the years or decades before it (the past) and the many years to come (the future). The growing discussion around data donation illustrates the ethical and legal challenges related to providing consent to ongoing use beyond death for an entire lifetime of genetic and nongenetic health data, some of which may have implications for relatives. In fact, most of the current data protection frameworks, such as GDPR, neglect data donation, and use of data after death is not within their scope (Recital 27 EU GDPR). At least until the point of death, a dynamic consent system seems to be a promising means of controlling different users' access to past and present medical data and for controlling data linkage with other studies. Another issue complicating the use of direct data concerns mHealth data (ie, data that are gathered when using mHealth solutions). Such apps are mostly overseen by national authorities, and, for instance in the United States, they need approval from the FDA. Data gathered by those apps need higher protection by law than US regulations and GDPR provide at present. In particular, the issue of consent is rather unregulated, and what could be seen as possible consent (the acceptance of the T&Cs) does not meet the high standards as imposed by traditional consent in research.

Indirect health data, such as exercise, social media, and movement data concerning an individual, are not currently regulated within the health data model in the United States. In the European Union, the GDPR covers such data, but they are not regarded as distinct health data and thus benefit from less protection, as is the case for inferred data. As stated above, consent is usually given via the T&Cs of relevant apps. This must change in the future. Either T&Cs must become much more user friendly and accessible or an entirely different consent model and system more akin to direct health data governance will have to be adopted.

Inferred data are particularly problematic in terms of consent, as they are the result of the combination of two or more datasets of one individual. Consent may have been given to the processing of each of these data points (or sets) but not to their combined processing, which can yield more revealing data not anticipated at the time consent was provided. One way to approach this problem is to make it clear at the point of consent to use direct and

indirect health data that their combination with other datasets is a real possibility despite goodwill and efforts on the part of researchers, companies, and other users to prevent it. With further technical developments, it might even be possible to send an alert to a dynamic consent portal for each new instance of combining data points of the individual, enabling tighter and more finely grained control of inferred data. This would ensure that consent is provided when new inferences are made. It is important to bear in mind in the discussion of inferred health data that some data, particularly genetic data, affect not only the data subject but also family members, which is also true for social media data when parents share information of their close relatives and children. Some individuals might be very happy to share all types of data, but this can have ramifications for close relatives, particularly identical twins, in the case of genetic data. In such cases, some form of collective consent may be required.

By its very nature, consent to use invisible health data cannot currently be provided, as we are still blind to the very nature of such data and their potential relevance to health. However, consent can be “future proofed” to a limited extent with careful legislation and regulation. By adopting a similar level of oversight for all types of data concerning a person (the GDPR is a step in this direction), safeguards will be in place once it emerges that seemingly entirely innocuous data can be used by artificial intelligence technologies to yield health findings. Once this transpires, alerts to dynamic consent systems will be a sensible precaution.

As the future use of data cannot be foreseen at present, alert mechanisms play a particularly important role, especially for inferred health data and invisible health data. Given our thesis that any data could turn out to be relevant to health, alerts might well be essential to ensure that people are kept informed when their data are put to a novel use with health implications. Whether citizens would or could have any right to stop the processing of such “new” health data is a difficult question that is outside the scope of this paper, but informing them seems to be a basic ethical requirement. If, as we suggest, all data will become health relevant, it might be impossible or, at least, very impractical for people to stop the processing of all data relevant to their health. Laws, such as the GDPR, will have to keep pace with developments in the conception of health data, as imposing current GDPR standards on all data that might be relevant for health in the future could have great implications for the processing of data, particularly in research. Table 3.2 summarizes the

different categories of health-related data and how consent to use past, present, and future data could be approached.

Table 3.2. Consent based on different health data types related to their temporal origin.

Data Type	Past Data	Present Data	Future Data
Direct health data	Specific/broad consent	Direct/specific/broad consent	Direct consent
Indirect health data	Direct consent/terms and conditions	Direct consent/terms and conditions	Direct consent
Inferred health data	Not applicable	Alerts to dynamic consent systems	Alerts to dynamic consent systems
Invisible health data	Nota applicable	Not applicable	Alerts to dynamic consent systems

VI. Conclusion

With each passing day, billions of gigabytes of direct, indirect, and inferred health data are being recorded, with massive implications for privacy and harm prevention if adequate consent mechanisms for their use are not in place. The possibility of invisible health data complicates the situation further. If all our data will be health data one day, we need to start treating consent to data use with the respect that it deserves. Currently, most data collectors are gaining access to vast amounts of behavioral and health data effectively for free, without having to comply with any safeguards. Broadening the notion of health data, as we have suggested, would cause companies to give more thought to ethical acquisition and processing of data. However, broadening the notion of health data could have an adverse effect on research if it results in excessively burdensome regulations.

Our argument that all data are health data is primarily ethical, but it could have important legal ramifications. In jurisdictions where health data can only be processed with consent, widening the scope of health data in this way would vastly increase the burden on, for example, private companies who process indirect and inferred health data. This might be difficult but ethically appropriate, and the development of more modern and dynamic consent mechanisms could facilitate this shift. Alternatively, legislation could limit the legal scope of health data to direct health data, leaving soft laws and guidelines to regulate other categories of health data.

VII. References

- Blasimme, A., Vayena, E., & Hafen, E. (2018). Democratizing Health Research Through Data Cooperatives. *Philosophy & Technology*, 31(3), 473–479.
<https://doi.org/10.1007/s13347-018-0320-8>
- Brunschwig, C. R. (2001). *Visualisierung von Rechtsnormen : legal design (Vol. 45)*. Schulthess.
- Cao, H., & Lin, M. (2017). Mining smartphone data for app usage prediction and recommendations: A survey. In *Pervasive and Mobile Computing*.
<https://doi.org/10.1016/j.pmcj.2017.01.007>
- Cate, F. H., Cullen, P., & Mayer-Schonberger, V. (2013). *Data Protection Principles for the 21st Century*. Maurer Faculty.
- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67–73.
<https://doi.org/10.1093/idpl/ipt005>
- Christen, M., Domingo-Ferrer, J., Draganski, B., Spranger, T., & Walter, H. (2016). On the Compatibility of Big Data Driven Research and Informed Consent: The Example of the Human Brain Project (pp. 199–218). Springer, Cham.
https://doi.org/10.1007/978-3-319-33525-4_9
- Do, T. M. T., & Gatica-Perez, D. (2014). Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*.
<https://doi.org/10.1016/j.pmcj.2013.03.006>
- EPFL. (2019). *Time Machine: Big Data of the Past for the Future of Europe*.
<https://www.timemachine.eu/>
- Federal Trade Commission (FTC). (2016). *Mobile Health App Developers: FTC Best Practices | Federal Trade Commission*. <https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-app-developers-ftc-best-practices>
- Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science. *Perspectives on Psychological Science*, 11(6), 838–854.
<https://doi.org/10.1177/17456916166650285>
- Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PLOS ONE*, 13(10), e0204937. <https://doi.org/10.1371/journal.pone.0204937>
- Ioannidis, J. P. A. (2013). Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research. *American Journal of Bioethics*.
<https://doi.org/10.1080/15265161.2013.768864>
- Jacobs, B., & Popma, J. (2019). Medical research, Big Data and the need for privacy by design. *Big Data & Society*, 6(1), 205395171882435.
<https://doi.org/10.1177/2053951718824352>
- Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 25.
<https://doi.org/10.1186/s40537-016-0059-y>
- Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital

- phenotype. *Nature Biotechnology*, 33(5), 462–463. <https://doi.org/10.1038/nbt.3223>
- Jones, L., & Nevell, R. (2016). Plagued by doubt and viral misinformation: the need for evidence-based use of historical disease images. *The Lancet. Infectious Diseases*, 16(10), e235–e240. [https://doi.org/10.1016/S1473-3099\(16\)30119-0](https://doi.org/10.1016/S1473-3099(16)30119-0)
- Kaplan, B. (2016). How Should Health Data Be Used? Privacy, Secondary Use, and Big Data Sales. *Cambridge Quarterly of Healthcare Ethics*. <https://doi.org/10.1017/S0963180115000614>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics : EJHG*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Kelley, P. G., Bresee, J., Cranor, L. F., & Reeder, R. W. (2009). A “Nutrition Label” for privacy. *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09*, 1. <https://doi.org/10.1145/1572532.1572538>
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>
- Mayer, M. A., Fernández-Luque, L., & Leis, A. (2016). Big Data For Health Through Social Media. *Participatory Health Through Social Media*, 67–82. <https://doi.org/10.1016/B978-0-12-809269-9.00005-0>
- McDermott, Y. (2017). Conceptualising the right to data protection in an era of Big Data. In *Big Data and Society*. <https://doi.org/10.1177/2053951716686994>
- Mostert, M., Bredenoord, A. L., Biesart, M. C. I. H., & van Delden, J. J. M. (2016). Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics*, 24(7), 956–960. <https://doi.org/10.1038/ejhg.2015.239>
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research*, 17(7), e175. <https://doi.org/10.2196/jmir.4273>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). Google’s Project Nightingale highlights the necessity of data science ethics review. *EMBO Molecular Medicine*. <https://doi.org/10.15252/emmm.202012053>
- Shaw, D. M. (2019). Defining Data Donation After Death: Metadata, Families, Directives, Guardians and the Route to Big Consent (pp. 151–159). https://doi.org/10.1007/978-3-030-04363-6_10
- General Data Protection Regulation, Official Journal of the European Union (2016).

https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf

- UK Children's Commissioner. (2018). Simplified Social Media Terms and Conditions for Facebook, Instagram, Snapchat, YouTube and WhatsApp. <https://www.childrenscommissioner.gov.uk/publication/simplified-social-media-terms-and-conditions-for-facebook-instagram-snapchat-youtube-and-whatsapp/>
- Vayena, E., & Blasimme, A. (2017). Biomedical Big Data: New Models of Control Over Access, Use and Governance. *Journal of Bioethical Inquiry*, 14(4), 501–513. <https://doi.org/10.1007/s11673-017-9809-6>
- Vayena, E., & Blasimme, A. (2018). Health Research with Big Data: Time for Systemic Oversight. *The Journal of Law, Medicine & Ethics*, 46(1), 119–129. <https://doi.org/10.1177/1073110518766026>
- Wachter, S., & Mittelstadt, B. D. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review*, 2, 130. <https://ssrn.com/abstract=3248829>
- Williams, H., Spencer, K., Sanders, C., Lund, D., Whitley, E. A., Kaye, J., & Dixon, W. G. (2015). Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR Medical Informatics*, 3(1), e3. <https://doi.org/10.2196/medinform.3525>

Chapter 4: Social media terms and conditions and informed consent from children: An ethical analysis

Full Reference: Schneble CO, Favaretto M, Elger BS, Shaw DM. Social media terms and conditions and informed consent from children: An ethical analysis. JMIR Pediatr Parent., 2021, Vol4, No2 DOI: 10.2196/22281

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, Maddalena Favaretto¹, MSc, Bernice Simone Elger^{1,2}, MD, PhD, David Martin Shaw¹, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

I. Abstract

Background: Terms and conditions define the relationship between social media companies and users. However, these legal agreements are long and written in complex language. It remains questionable whether users understand the terms and conditions and are aware of the consequences of joining such a network. With children interacting from a young age with social media, companies are thus acquiring large amounts of data, resulting in longitudinal data sets that most researchers can only dream of. Children's use of social media is highly relevant to their mental and physical health for two reasons: their health can be adversely affected by social media, and their data can be used to conduct health research.

Objective: This paper offers an ethical analysis of how the most common social media apps/services inform users and obtain their consent regarding privacy and other issues, and discusses how lessons from research ethics can lead to trusted partnerships between users and social media companies, with a focus on children, who represent a vulnerable group amongst users of those platforms.

Methods: A thematic analysis of the terms and conditions of the 20 most popular social media platforms and the two predominant mobile phone ecosystems (Android and iOS) was conducted. The results of this analysis served as the basis for the scoring of these platforms.

Results: The analysis showed that most of the platforms comply with the age requirements issued by legislators. However, the consent process during signup is not taken seriously. Terms and conditions are often too long and difficult to understand, especially for younger users. The same applies to age verification, which is not realized proactively but instead relies on other users reporting underaged users.

Conclusion: Our analysis reveals that social media networks are still lacking in many respects in regard to adequate protection of users who are children. Consent procedures are flawed because they are too complex, and in some cases, children can create social media accounts without sufficient age verification or parental oversight. Adopting measures based on key ethical principles will safeguard the health and well-being of children. This could mean standardizing the registration process along the lines of modern research ethics procedures: give users the key facts that they need in a format that can be read easily and quickly, rather than forcing them to wade through chapters of legal language that they

cannot understand. Improving these processes would help safeguard the mental health of children and other social media users.

II. Introduction

a) Background

Social media companies have experienced tremendous growth during the last decade; however, they have largely neglected the issues of privacy and confidentiality. In addition to connecting people, social media apps (the companies) are also tremendous data collectors, gathering a wide range of information that spans from nonsensitive to highly sensitive data. Although many data might be nonsensitive in isolation, the combination of various types of data might subsequently allow insights into sensitive health issues (Schneble et al., 2019). In fact, many studies have used social media data to gain insights into the mental state of users (Eichstaedt et al., 2018; Reece & Danforth, 2017). Moreover, with children and young adults using social media apps from a young age, companies have acquired data over long time spans, which is similar to longitudinal data used in research. Keeping this in mind and knowing that predictive algorithms will become more accurate, it is of major importance to build governance and inform users about the use of their data to foster data protection. This is all the more important given the latest scandal surrounding Cambridge Analytica (Schneble et al., 2018; The Guardian, 2018) and the sharing of data between Facebook and device manufacturers such as Apple and top-rated apps such as Spotify and Netflix (Alex Hern, 2018). These are prominent examples of misbehavior that illustrate the urgent need for a trusted partnership between users and social media companies.

Contractual law in the terms and conditions (also known as terms of services) and privacy policies define how privacy, confidentiality, and data sharing are handled. They are the predominant legal and contractual mechanisms that define the relationship between users and social media companies. These mechanisms are subject to various national and international regulations. The General Data Protection Regulation (GDPR) of the European Union (EU) (General Data Protection Regulation, 2016) sets boundaries concerning the processing of data. In the United States, the Children's Online Privacy Protection Act (COPPA) (Children's Online Privacy Protection Act, 2000) and the fair information principles issued by the Federal Trade Commission (Federal Trade Commission, 1998) are the 2 predominant regulations.

When signing up for such a service, users consent by reading or at least scrolling through the terms of service and by clicking the agree button. However, these terms and conditions are often long and written in a complex legal language. Thus, it remains questionable whether users—particularly children and young adults—truly understand the terms and conditions and are aware of the consequences of joining a network. Most of the platforms offer their service for free but require users to accept the preset package of conditions with limited privacy choices to permit access to their services.

Social media apps are ubiquitous in today's world and have changed the way we communicate, share, and interact with each other daily. They are also omnipresent in the lives of young people, and it is estimated that 1 in 3 of all internet users is under the age of 18 years (Milkaite & Lievens, 2019; UNICEF, Brian, Ed. Little, Céline, 2017). A recent study by the UK Children's Commissioner has shown that a significant number of children access social media through their parents' accounts, whereas most adolescents (71% in the United States and 85% in Europe) have one or more social media accounts or identities (Butterfill et al., 2018). When children access social media through their parents' accounts, parents might feel that they have control over their children's media use. This is problematic for 2 reasons: first, parents will not be able to control every click, and second, as the UK Children's Commissioner points out, children might be presented with explicit adult content of which their parents remain unaware.

Letting children use parents' accounts also bypasses the age requirements imposed by social media companies. In their terms of service, social media apps and services defined the minimum age at which adolescents or children can use the app or service without obtaining parental consent. With regard to age requirements, the law plays an important role by setting boundaries for protecting children's privacy, data sharing, and profiling. In the United States, COPPA defines 13 years as the minimum age to join such communities. Before that age, explicit parental consent is needed to sign up. The EU has recently introduced the GDPR, in which Article 8 defines the necessity of parental consent for all youths aged below 16 years in situations where information society services are offered directly to them. However, the member states are free to choose and adopt their own particular regulation within the age range of 13-16 years. Some countries, such as the United Kingdom, have opted for an age of 13 years, whereas others such as Germany have set the boundary at 16 years (Milkaite & Lievens, 2019). The GDPR would thus not prohibit the use of such services before the minimum age requiring children's self-consent but would instead

require parental consent to access these services and process the personal data of children, as defined in the GDPR. Most of the companies however set their minimum age requirements at the age imposed by national law, as shown in our results.

However, the efficacy of such age regulations remains to be questionable as the primary research strands in children's digital rights show that children and parents feel social pressure to join such communities (Butterfill et al., 2018) and thus might lie about their age when joining social media services (Boyd et al., 2018). Doing so is easy because normally, signing up relies only on the honesty of the user.

b) Objectives

This paper provides an ethical analysis of the most popular social media platforms and services used by children and adolescents (in the EU and the United States). It focuses on age requirements, how information about the platform is presented, how consent is obtained, how (and if) age verification is implemented, whether resources are provided to educate parents or children, and if there are community guidelines. It then discusses the emerging issues and the predominant regulations of our target countries and illustrates how experiences from research ethics could be used to develop a trusted relationship between users and companies, facilitating the ethical functioning of social media networks.

III. Methods

We conducted a thematic analysis (Braun & Clarke, 2006) of the terms and conditions of the 20 most popular social media platforms in 2019 (We Are Social, 2018) and the 2 predominant mobile phone ecosystems, Android and iOS. Within this sample of 20 platforms, we excluded all apps and social networks targeting only Chinese-speaking users (because of a lack of terms and conditions in English; WeChat, QQ, QZone, and Sina Weibo), discussion websites (Reddit), and those targeting only adults (LinkedIn or Viber), resulting in 10 platforms relevant to children. The terms and conditions were read in depth, emerging topics of ethical interest were identified, and categories for further in-depth analysis were created. The categories identified were the minimum age to join, how the consent process was handled, the age verification process, the presence of parental portals (educating parents on the use of the respective platforms), and the possibility of requesting account deletion in the cases of underaged users. Note that most of the platforms are available either as web apps or as smartphone apps. The results of this in-depth analysis are

summarized in Table 4.1, and the apps are scored according to the criteria in Table 4.2. As most of the apps are available on smartphones, we also decided to include the quasi-standard platforms such as Android and Google, as they have a gatekeeping function (in terms of age) to allow children to access those networks.

Table 4.1 Overview of the most popular social media apps.

Platform/App	Active users (in millions)	Provider	Predominant content	Viewable without signing in	Minimum age	Age verification	Possibility to request deletion of the account	Parental consent	Parents Portals/Community guidelines
Social Media									
Facebook	2234	Facebook Inc	Video/Text/Images/Social Messaging	Yes	13	Verification of official document when Account is locked	Yes (Form)	Consent by user	Yes
YouTube	1900	Google	Video creation	Yes	13 (14+/16+) ¹	Background check/Verification of official document/Credit card verification when locked	Yes	Consent by user or parents if under 13	Yes
WhatsApp	1500	WhatsApp Inc. (Facebook Inc.)	Social Messaging (Video/Text/Music)	No	13	By SMS	No	Consent by user	No
Instagram	1000	Facebook Inc.	Images/Video	Yes	13 (16) ¹	Verification of ID when locked	Yes (Form)	Consent by user	Yes
Tik Tok	500	Beijing Bytedance Technology	Music/Images	No	13 (14) ¹	No	Yes (Mail)	Yes (for certain countries)	No
Twitter	335	Twitter Inc.	Text	Yes	No	Yes for sensitive posts	Yes	Consent by user	No
Skype	300	Microsoft Corporation	Social Messaging	No	No	No	No	Consent by user	No
Snapchat	291	Snap Inc.	Video/Photo posting	No	13	By peers/Birthday can be changed only a limited number of times	Yes (Mail)	Consent by user	Yes
Pinterest	250	13	Images	Yes	13	By peer	Yes (Form)	Consent by parents if underaged use	Yes
LINE	203	LINE Corporation	Social Messaging	No	No	No	No	No	No
Ecosystems									
IOS (Apple ID)	Na.	Apple	Apps	N/A	13	Yes (Credit card/SMS)	Yes	Consent by parents if underaged users	Yes.
Android Play Store	Na.	Google	Apps	N/A.	13 (14+, 16+) ¹	Back check/Verification of official document/Credit card verif.	Yes	Consent by parents if underaged users	Yes

Table 4.2. Scoring the most popular social media apps.

Platform or app	Minimum age or age verification	Parental consent	Possibility to request deletion of the account	Parent portal or community guidelines	Total score
Facebook	Age restriction and implemented age verification present	Consent by user	Yes	Parent portal present	3
YouTube	Age restriction and implemented age verification present	Consent by parents	Yes	Parent portal present	4
WhatsApp	Age restriction and implemented age verification present	Consent by parents	No	No parent portal	2
Instagram	Age restriction and implemented age verification present	Consent by user	Yes	Parent portal present	3
TikTok	No age restriction or no age verification present	Consent by parents	No	No parent portal	1
Twitter	No age restriction or no age verification present	Consent by user	Yes	No parent portal	1
Skype	Age restriction and implemented age verification present	Consent by parents	No	No parent portal	2
Snapchat	Age restriction and implemented age verification present	Consent by user	Yes	Parent portal present	3
Pinterest	Age restriction and implemented age verification present	Consent by parents	Yes	Parent portal present	4
LINE	Age restriction and implemented age verification present	Consent by user	No	No parent portal	1

IV. Results

The results of our analysis will be discussed thematically, in turn, after presenting the results of our scoring mechanism.

a) Scoring System

On the basis of the data in Table 4.1, our scoring system (Table 4.2) awards each platform a possible score of 1 or 0 across the 4 different categories used in our analysis. The criteria are presented in Table 4.3. The category for minimum age and age verification is cumulative. One point will be awarded only if both criteria are met, because we believe this fulfills the gatekeeper function. Studies suggest that children are often happy to lie about their age and that parents even encourage their children to sign up (Boyd et al., 2011; British Office

of Communications, 2019); thus, the efficacy of a minimum age requirement in the absence of verification remains ethically questionable.

Table 4.3. Constraints of the scoring system.

Topic	Criteria for point (+)	No point (none) if
Minimum age or age verification	Age restriction and implemented age verification present	No age restriction or no age verification present
Possibility to request deletion	Yes	No
(Parental) consent process	Consent by parents (+)	Consent by user (none)
Parent portal	Parent portal present	No parent portal

b) Age Requirements and Age Verification

Table 4.1 shows that all companies except LINE have adopted a minimum age of 13 years for the use of their services. However, the Apple and Google (Android) ecosystems offer the possibility of using their various services at a younger age with parental consent. Google achieves this by integrating the child's account into the so-called Family Link (Google Inc., 2020), a platform to group and administrate family member accounts; the same applies to Apple, which has also set up an infrastructure to manage family accounts. Most service providers rely on other users reporting underage use and offer either a mailing address or a form as the only way of contact when requesting the deletion of an account created by underage children. A more sophisticated method has been adopted by Google, where a background check is performed by verifying the age entered in any one of its services whenever the user uses another service that is part of its ecosystem. Once an account is locked, Instagram and Facebook request a copy of an official document (ID card or passport) to unlock it. Android, iOS, and YouTube adopt another way of handling this issue, where the check is performed against a valid credit card, resulting in a parent giving de facto consent. In contrast, Snapchat allows users to change their date of birth only a certain number of times (Snap Inc., 2018).

c) Consent Process

Upon registration, the user was asked to accept the terms and conditions. In most cases, the user agrees to the terms and conditions by checking a checkbox and subsequently clicking the register button or even by only clicking the register button (Facebook and Instagram).

Sometimes, the link to the terms and conditions is in a smaller font (see Appendix Table 2 for an overview) so that it is hardly identifiable (Snapchat). On Instagram and Facebook, it is highlighted in bold font. Although the Article 29 Working Party (an independent European advisory body on data protection and privacy created by the EU) offers some recommendations on the consent process (Article 29 Working Party, 2018), we were not able to identify a standard presentation form or standard procedure in presenting terms and conditions. Most forms show their terms and conditions only in continuous text, whereas others have adopted a question and answer form (eg, Facebook, Instagram, and Pinterest). Pinterest is the only platform that provides a simplified version in addition to the full version of its terms (Textbox 1).

Textbox 1. Full text versus simplified terms and conditions (Pinterest).

Full text

You grant Pinterest and our users a non-exclusive, royalty-free, transferable, sublicensable, worldwide license to use, store, display, reproduce, save, modify, create derivative works, perform, and distribute your User Content on Pinterest solely for the purposes of operating, developing, providing, and using Pinterest. Nothing in these Terms restricts other legal rights Pinterest may have to User Content, for example under other licenses. We reserve the right to remove or modify User Content or change the way it's used in Pinterest, for any reason. This includes User Content that we believe violates these Terms, our Community Guidelines, or any other policies.

Simplified version

If you post your content on Pinterest, we can show it to people and others can save it. Don't post porn or spam or be a jerk to other people on Pinterest.

d) Parent Portals or Community Guidelines

Almost every platform (except social messaging platforms) offers a parent's portal or community guidelines. This ranges from simply linking to interesting articles (Snapchat) to providing an information center (Instagram and Facebook) to video sequences (Facebook) on problematic behavior along with short sequences showing a safe way to use the service.

V. Discussion

a) Principal Findings

On the basis of our scoring system (Table 4.2), most providers scored 3 out of 4 points. However, one-third of the service providers achieved poor results. This shows that the regulations that service providers comply with, either by themselves or by law, offer at least some protection for users. However, TikTok, Twitter, and LINE only scored 1 point and only 2 companies achieved the maximum score (Pinterest and YouTube).

In the following section, we will therefore discuss the categories presented in Table 4.1 and suggest possible improvements within the framework of the 4 guiding ethical principles.

b) Minimum Age to Sign Up for a Service

Our analysis reveals that most apps have adopted the minimum age of 13 years for children to sign up to use their services. This complies with the US COPPA and GDPR. In contrast with the COPPA, the GDPR provides a minimum age requirement ranging from 13 to 16 years for children to register for a service. Owing to the GDPR's extraterritorial force (as mentioned in Article 3 of the GDPR), other states and companies outside the EU have to comply with EU standards when targeting users (and children) in an EU member state.

Strongly intertwined with the definition of the minimum age is the issue of age verification. As Table 4.1 shows, the issue of age verification is currently not taken seriously by companies, and an age requirement is largely useless in the absence of verification. Therefore, we argue that a robust age verification process needs to be adopted by service providers in the coming years. However, establishing such mechanisms needs to be implemented in a way that complies with privacy and the principles of data minimization (Article 29 Working Party, 2018). The survey mentioned earlier (Boyd et al., 2011) has shown that some children lie about their age and the ease of registering for a social media service (requiring only a few minutes) does not constitute a barrier.

Currently, some providers request verification by email or phone by sending the user a short message during the registration process (the standard procedure for setting up a WhatsApp account). The latter provides an additional security layer as cell phone companies have a minimum age for issuing a contract; when a child has a cell phone, the parents have at least agreed to the use of such a device and thus are aware that the child might sign

up for such a service, even if they are potentially unaware of the services that the child subsequently signs up for. However, this might be a problem in countries where pay-as-you-go phones require no identification, either by age or by verification with an official ID card or social security card. Furthermore, implementing an age verification process by requesting verification through a text message could be seen as discriminating against children who do not possess a cell phone at all and, thus, solely have to rely on a parent to register.

Other providers delegate age verification to their users by setting up forms where one can report underage use. However, this method does not guarantee age verification and, in the absence of other measures, it suggests that the service provider is neither serious nor proactively interested in complying with the minimum age requirement.

Today's technologies could make it possible to approach the minimum age to check more proactively. For example, artificial intelligence could enable the use of techniques such as image classification algorithms or natural language processing to detect underage children by analyzing their physical face properties (such as the Amazon recognition application programming interface (Amazon, 2018)) or using written language with neurolinguistic programming for processing natural language. We are fully aware that the use of such technologies can lead to other ethical and legal concerns. Although these concerns are too complex to address in depth in this paper, we discuss them briefly in the following section.

Article 9 of the GDPR places biometric data in a special category: processing is prohibited unless special circumstances are met. However, notably Article 9 (General Data Protection Regulation, 2016) of the GDPR permits each EU member country to introduce certain derogations with respect to restrictions on processing biometric data (member states may maintain or introduce further conditions, including limitations). For instance, the Netherlands has provided an opt-out option for biometric data if necessary, for authentication or security purposes, and Croatia's new data protection law exempts surveillance security systems (Whitener & Aragon, 2019). In the United States, no federal law regulates the collection of biometric data. However, 3 states—Illinois, Washington, and Texas—have implemented regulations on biometric data (Whitener & Aragon, 2019). On the ethical side, the introduction of such technologies to tackle the issue of age verification is also potentially problematic, as appropriate consent must be obtained from the user, who should also have a full overview where the biometric data are being used, as these types

of data represent special categories that are harmful when misused. Thus, the use of such technologies should follow clear ethical guidelines. For example, such technologies should not be used to collect more information about users and data than is necessary, and they should always be used for a specific purpose. This is also because an increasing number of predictive analyses are possible (Reece & Danforth, 2017; van der Hof & Lievens, 2017) from simple social media data.

c) Obtaining Consent

Obtaining valid user consent (and in the case of children, parental consent) is one of the 6 lawful bases to process personal data, as listed in Article 6 of the GDPR. Generally, as consent is a tool that gives users data subjects control over whether personal data concerning them will be processed (Article 29 Working Party, 2018), to do so, valid consent has to meet certain criteria; it must be freely given, be specific, and be informed and include an unambiguous indication of the data subject's wishes. How consent is presented to the user, whether it is written or presented pictorially or in short video sequences, is up to the controller (company). This means that harmonization is not currently envisaged. However, the Article 29 Working Party (an advisory board of the EU on data protection issues) does lay out how data subjects (users) should provide consent. Obtaining consent by simply scrolling down and ticking a checkbox is not seen as appropriate from an ethical standpoint, although it might be sufficient from a policy perspective. Thus, the Working Party provides 2 examples of how a valid mechanism could look (outlined in Textbox 2), which is not currently met by any of the services that are subject to our investigation. As shown in our analysis, users are presented with written information on their rights and rights of companies on topics such as data protection, community rules, and minimum age. A further issue is that some of the services only provide a checkbox to tick or, in the worst case, only a button to register where the terms and conditions are not displayed during the account's creation unless the user clicks the link. This fosters a click and forget mentality and is far from providing a sustainable and respectful partnership between service providers and users. Often, the link to the terms and conditions is presented in smaller fonts and stands in contrast with the large textboxes filled during the registration process, as shown in the examples in Table 2 in the Appendix.

Textbox 2. Example of how to obtain consent (examples of the Article 29 Working Party).

Appropriate way

Swiping a bar on a screen, waiving in front of a smart camera, turning a smartphone around clockwise, or in a figure-eight motion may be options to indicate agreement, as long as clear information is provided, and it is clear that the motion in question signifies agreement to a specific request (e.g., if you swipe this bar to the left, you agree to the use of information X for purpose Y. Repeat the motion to confirm). The controller must be able to demonstrate that consent was obtained this way, and data subjects must be able to withdraw consent as easily as it was given.

Inappropriate way

Scrolling down or swiping through a website will not satisfy the requirement of a clear and affirmative action. This is because the alert that continuing to scroll will constitute consent may be difficult to distinguish and/or maybe missed when a data subject is quickly scrolling through large amounts of text and such an action is not sufficiently unambiguous.

A special category for obtaining consent is imposed for children below the age of legal maturity in their respective countries. In such cases, the GDPR and COPPA require approval from the parent or guardian. This has several positive and negative aspects. On the one hand, this regulation places the burden on the parents to protect children from potential harm, which could, in turn, be built by safeguarding mechanisms of the platforms. On the other hand, overrestrictive consent processes could be a driver of inequality, as strict parents could hinder beneficial usage. A complex consent process (such as using the parents' credit card or facial recognition) is always associated with more data being collected not only from the child but also from the parent. Thus, balancing data minimization against sufficient safeguards plays an important role in designing an ethical consent process.

Emphasizing consent is important; however, other scholars have argued that solely focusing on this aspect and implying parental consent is not enough. By making data protection impact assessment mandatory (as required by the GDPR), risks can be already identified at an earlier stage (van der Hof & Lievens, 2017). Combining these 2 approaches for

making the terms and conditions more readable and fostering data protection impact assessments would help to protect children's rights.

d) Educating Users and Parents

As the report of the UK Children's Commissioner (Butterfill et al., 2018) has shown, the safe use of social media depends on building awareness and educating children about its use and fostering digital literacy. Parents and teachers play an important role. Most of the apps we analyzed offered parents websites where the companies either provided links to useful literature (the simplest way to deal with that issue) or by providing short YouTube sequences to inform children and parents about potential harm and the security measures to take when using social media.

Given the importance of educating parents and teens (Butterfill et al., 2018), we suggest that future legislation should mandate the implementation of such parental portals. From an ethical point of view, it would be good to encourage companies to spend a reasonable amount of their revenue in educating parents and children about the potential harm resulting from the use of their services. A good example is provided in the Facebook Help Center, which offers short YouTube sequences and quizzes on the topics of data protection and possible harm.

e) Social Pressure

Social media apps have become ubiquitous among children and adolescents. It has become difficult to refuse to be part of such networks, because of both social pressure and an increasing number of institutions (such as schools) requiring such channels, resulting in social pressure to use these services for communication, regardless of whether parents regard the use of these services to be appropriate for their children. This could also be seen as a loss of autonomy concerning the freedom to decide whether and when to join. We can imagine a scenario in which children who want to participate in social media life are pressured to lie about their age on the internet by fellow schoolmates or friends because this peer group's main vehicle of social interaction is heavily mediated by online-messaging and social media, for example, children need to be on WhatsApp to be able to meet with others because all of the peer meetings are communicated that way. It is also possible that parents could incentivize their offspring to engage in online misconduct as they want their children to use online messaging services (eg, WhatsApp) out of convenience or for monitoring purposes. These phenomena can create new social inequalities. In fact, in its 2017

report, UNICEF (United Nations International Children's Emergency Fund) warned of the formation of a significant digital divide (UNICEF, 2017), highlighting the gap between children who can connect and subsequently sign up for social media networks. This divide could be the result of either having more permissive parents who agree to the use of such services or because the child is wealthy enough to purchase a pay-as-you-go phone with data to access SM services secretly. Conversely, children who are left out of social media because their parents are more law-abiding or controlling or because their socioeconomically disadvantaged background makes personal phones unaffordable or are forced to share their parents' devices. Children in the latter group feel left out of their friends' social lives and end up being ostracized by their peers or even bullied.

With the introduction of the GDPR and the adjustment of the minimum age to 16 years in certain countries, it is expected that the topic of social pressure will defuse itself at least on an institutional level because institutions must adhere to this requirement. However, social media companies' adherence to the GDPR age requirement could, on the other hand, worsen social pressure for children as the gap between the legal age at which it is possible to join social media and children's actual social practices differs (Livingstone, 2018). In medical care, children can give consent for themselves below the legal age of maturity; however, this exception does not apply in the case of compliance with GDPR.

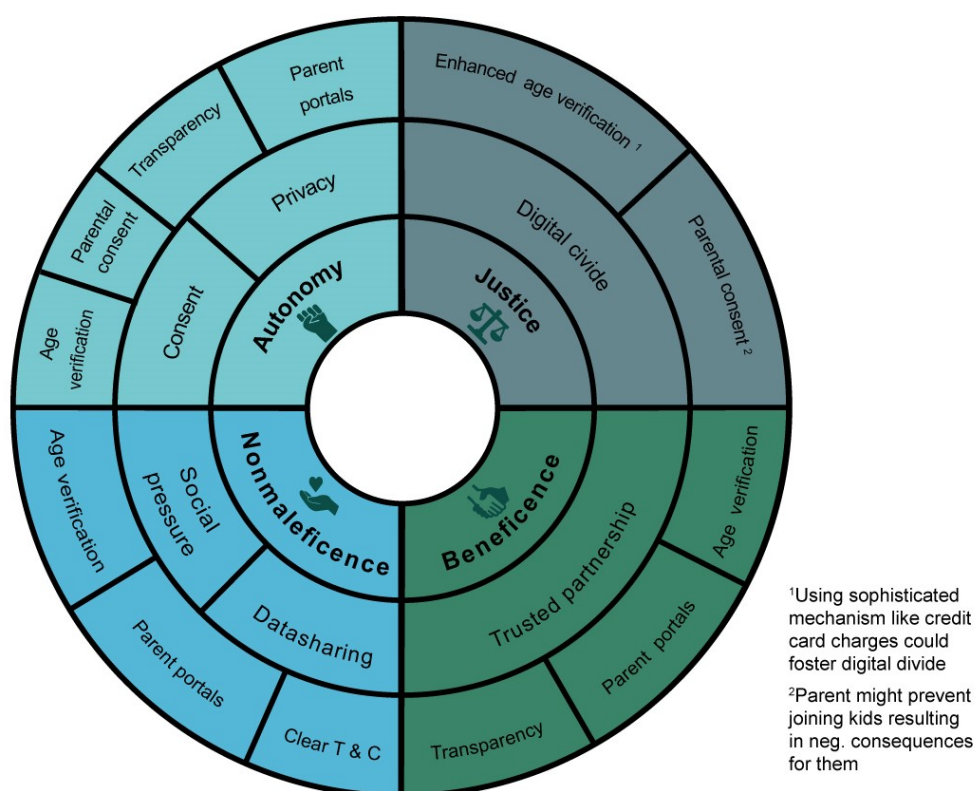
f) Research Ethics as a Model for a Trust-Based Partnership

Similar to social media today, biomedical research used to have a bad reputation in terms of involving participants. People were included in medical studies without their consent, and their data were shared without their knowledge. To prevent such unethical practices, 4 main ethical principles have become fundamental to research ethics and biomedical ethics more widely: respect for autonomy, nonmaleficence, beneficence, and justice. In the context of social media, all of these principles are relevant; however, this is particularly true of respect for autonomy and nonmaleficence. Fig. 3.1 illustrates how social media can innovate to ensure age verification, valid consent, and other aspects to make sure that these key ethical principles are respected. Fundamentally, it is an ethical imperative to ensure that children are of suitable age and understand the risks of social media to reduce the risk of harm to their emotional well-being and mental health: evidence suggests that social media can have substantial impacts in the areas of self-esteem and well-being, with issues related to cyberbullying and Facebook Depression (Richards et al., 2015).

In research ethics, the informed consent process plays a crucial role and contributes to a trusted partnership between subjects and researchers. When approached about the possibility of involvement in a clinical study (and increasingly for interviews or survey participation), potential participants are given all relevant information and time to digest and consider it before signing an informed consent form. In the past, the information provided to participants often ran to over 100 pages, thus raising the same concerns about accessibility and comprehensibility as social media terms and conditions. In recent years, however, there has been a move toward making such information much more patient- and participant-friendly, with, for example, the UK Human Research Authority supporting the use of simple information sheets in a question and answer format running to a maximum of 5-10 pages. This practice focus on communicating relevant information about risks and harms in a concise and comprehensible format could also serve as a model for building trusted relationships between social media users and companies. The problem with using terms and conditions as an information sheet is that such policies are essentially legal documents and written in dense legal language. Disentangling lengthy legal texts from the salient information required to provide informed consent is essential for social media companies. However, today's relationships are still unbalanced from the very beginning, with users required to sign up with a simple click after having to read information that is only presented in written form and complex language. This means that many users remain to be unaware of exactly what they are signing up for. Moving toward some sort of pictorial consent system would be a much more appropriate approach to informing both children and adults about the risks of social media use. This debate is not new in the legal context; Brunschwig (Brunschwig, 2001) was one of the first to show how contractual law can be exemplified with comics fostering a better understanding of otherwise complex matters. Several scholars have been working on this topic, proposing nutrition label-like terms and conditions (Kelley et al., 2009) and grid-based terms and conditions (Reeder et al., 2008). Such pictorial forms of consent are best practices in research ethics settings, especially with vulnerable or illiterate study participants. There might be some implementation issues with such solutions. Nevertheless, when we are speaking about children—a vulnerable group—such terms and conditions are a much better means of informing users about potential harm. This is not a purely theoretical discussion and approach, as Apple recently presented nutrition labels for their App Store (Morse, 2020).

Another possible solution, and a step in the right direction, is the simplified text-based rules for several social media apps developed by the UK Children's Commissioner (UK Children's Commissioner, 2018). Research ethics also requires that data can typically only be shared and processed with the consent of the persons concerned. However, recent social media scandals (Schneble et al., 2018, 2020) have shown that some social media companies have neglected this issue, which must also be addressed more clearly in terms and conditions. Another essential aspect of research ethics is the right to withdraw consent and the possibility of deleting data (or an account if research takes place via the internet) by the user. However, for underaged users (with respect to the minimum age required by the companies), it should also be possible for parents to delete an account without going through a complicated process. This could be done, for example, by specifying a parental contact when registering the account. Finally, research ethics also address the potential risks in participating in a study. Most companies in our sample address possible harms of using their services in their parent portals and community guidelines.

Figure 3.1) Different levels of Abstraction in Big Data research.



This figure maps the four biomedical ethical principles to issues arising from the use of social media and links them to possible fields of actions.

VI. Conclusions

Our analysis reveals that social media networks are still lacking in many respects with regard to adequate protection for children. Consent procedures are flawed because they are too complex, and in some cases, children can create social media accounts without sufficient age verification or parental oversight. Given the high risks of inappropriate content being shared and the targeting of children with specific advertisements, social media companies must improve their procedures to protect not only children but also all users. This can be achieved by standardizing the registration process in accordance with modern research ethics procedures described earlier: give users the key facts that they need in a format that can be read easily and quickly, rather than forcing them to wade through chapters of legal language that they cannot understand. Disentangling the practical information that users need from the complex legal language would also have the benefit of facilitating standardization; regardless of the jurisdictions, the language for consent documents should be simple and straightforward. In addition, in some cases, using pictorial versions of the terms and conditions would surely leverage the efficacy of today's mostly unread versions. The vast majority of social media users have given only uninformed consent; however, the click, consent, and forget at your peril model must be relegated to history in favor of a more transparent and ethical system. The standardization of terms and conditions is only possible if an effective political intervention is implemented. Recent developments and discussions about monopolistic large social media companies in the US Congress are a step toward harmonization. Furthermore, the role model function of the GDPR as a quasi-standard for new data protection regulations will eventually simplify standardization. Adopting measures based on key ethical principles will safeguard children's health and well-being and those of other social media users.

VII. Acknowledgments

This study was supported by the Swiss National Science Foundation under project number 167211.

VIII. References

- Alex Hern. (2018, December 5). Facebook discussed cashing in on user data, emails suggest. *The Guardian*.
<https://www.theguardian.com/technology/2018/dec/05/facebook-discussed-cashing-in-on-user-data-emails-suggest>
- Amazon. (2018). Amazon Rekognition. <https://aws.amazon.com/rekognition/>
- Article 29 Working Party. (2018). Guidelines on Consent under Regulation 2016/679. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051
- Boyd, D., Hargittai, E., Schultz, J., & Palfrey, J. (2011). Why parents help their children lie to Facebook about age: Unintended consequences of the “Children’s Online Privacy Protection Act.” *First Monday*. <https://doi.org/10.5210/fm.v16i11.3850>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- British Office of Communications. (2019). Why children spend time online. <https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2019/why-children-spend-time-online>
- Brunschwig, C. R. (2001). *Visualisierung von Rechtsnormen : legal design (Vol. 45)*. Schulthess.
- Butterfill, R., Charlotte, M.-P., Miles, A., Powell, H., Nettleton, O., Kriszner, M., Waldie, A., & De Ianno, D. (2018). *Life in ‘Likes: Children’s Commissioner Report into Social Media use in among 8-12 Year Olds*.
<https://www.childrenscommissioner.gov.uk/wp-content/uploads/2018/01/Childrens-Commissioner-for-England-Life-in-Likes-3.pdf>
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44), 11203–11208. <https://doi.org/10.1073/PNAS.1802331115>
- Federal Trade Commission. (1998). *PRIVACY ONLINE: A REPORT TO CONGRESS*.
<https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>
- Google Inc. (2020). Google Family Link. <https://families.google.com/familylink/>
- Kelley, P. G., Bresee, J., Cranor, L. F., & Reeder, R. W. (2009). A “Nutrition Label” for privacy. *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS ’09*, 1. <https://doi.org/10.1145/1572532.1572538>
- Livingstone, S. (2018). Children: A Special Case for Privacy? *InterMEDIA*, 46(2), 18–23. <http://www.iicom.org/intermedia/intermedia-past-issues/intermedia-july-2018/children-a-special-case-for-privacy>
- Milkaite, I., & Lievens, E. (2019). Children’s Rights to Privacy and Data Protection Around the World: Challenges in the Digital Realm. *European Journal of Law and Technology*, 10(1). <https://biblio.ugent.be/publication/8618586>
- Morse, J. (2020). Apple’s privacy-focused “nutrition labels” for apps are only a start. *Mashable*.

-
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- Reeder, R. W., Kelley, P. G., McDonald, A. M., & Cranor, L. F. (2008). A user study of the expandable grid applied to P3P privacy policy visualization. *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society - WPES '08*, 45. <https://doi.org/10.1145/1456403.1456413>
- Richards, D., Caldwell, P. H. Y., & Go, H. (2015). Impact of social media on the health of children and young people. In *Journal of Paediatrics and Child Health*. <https://doi.org/10.1111/jpc.13023>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2019). All our data will be health data one day: the need for universal data protection and comprehensive consent. *JMIR Preprints*. <https://preprints.jmir.org/preprint/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). Google's Project Nightingale highlights the necessity of data science ethics review. *EMBO Molecular Medicine*. <https://doi.org/10.15252/emmm.202012053>
- Snap Inc. (2018). Snapchat Support. <https://support.snapchat.com/en-US>
- Children's Online Privacy Protection Act, (2000).
- General Data Protection Regulation, Official Journal of the European Union (2016). https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf
- The Guardian. (2018). The Cambridge Analytica Files. <https://www.theguardian.com/news/series/cambridge-analytica-files>
- UK Children's Commissioner. (2018). Simplified Social Media Terms and Conditions for Facebook, Instagram, Snapchat, YouTube and WhatsApp. <https://www.childrenscommissioner.gov.uk/publication/simplified-social-media-terms-and-conditions-for-facebook-instagram-snapchat-youtube-and-whatsapp/>
- UNICEF, Brian, Ed.|Little, Céline, E. (2017). State of the Worlds Children 2017 - Children in a Digital World. In UNICEF. UNICEF. 3 United Nations Plaza, New York, NY 10017. Tel: 212-326-7000; Fax: 212-887-7465; Web site: <http://www.unicef.org/education>. <https://eric.ed.gov/?id=ED590013>
- UNICEF. (2017). The State of the World's Children.
- Van der Hof, S., & Lievens, E. (2017). The Importance of Privacy by Design and Data Protection Impact Assessments in Strengthening Protection of Children's Personal Data Under the GDPR. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3107660
- We Are Social. (2018). Most famous social network sites worldwide as of October 2018, ranked by number of active users (in millions). Statista - The Statistics Portal. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Whitener, M., & Aragon, R. (2019). How should we regulate facial-recognition technology? International Association of Privacy Professionals.
<https://iapp.org/news/a/how-should-we-regulate-facial-recognition-technology/>

Chapter 5: Informed Consent in the Age of Big Data - "A Contradiction in Terms"? A qualitative study of lawyers and data protection officers

Additional Paper (Not published)

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, David Martin Shaw¹, PhD, Jens Gaab, PhD², Arthur Caplan³, PhD, Bernice Simone Elger^{1,4}, MD, PhD,

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²Faculty for Psychology, University of Basel

³Department of Population Health, NYU

⁴University Center for Legal Medicine, University of Geneva, Switzerland

I. Abstract

The process of obtaining consent has traditionally been part of any ethical research project and is seen as essential to respecting the autonomy of participants. However, the concept of consent as traditionally conceived is threatened by Big Data science. It is thus of importance to investigate how consent can be updated and autonomy safeguarded in the era of Big Data. This paper addresses this issue by presenting the results of our interviews conducted amongst lawyers in the US.

II. Introduction

Over the last decade the rapid growth and dispersion of social media platforms and the ubiquitous use of smartphones has led to an explosion of the datafication of individuals (Bello-Orgaz et al., 2016). The advancement of computational methods for collecting, processing and analysing datasets has created new opportunities for science. However, the use of new technologies in research involving human subjects often means that consent processes must be adapted (Ioannidis, 2013). Participants must be provided with sufficient information, but this information must be easily understood to ensure valid consent.

In traditional research, those principles have been well established for decades, with a small number of participants ranging up to tens of thousands. However, with Big Data research becoming more common the notion of “the more data the better” has become more widespread, meaning that as well as large volumes of data being processed, the number participants has increased by several orders of magnitude, from a hundred thousand up to millions and millions of participants. This increase challenges the current process of a truly informing participants, as that would take month to years to inform them, applying traditional standards. Although one could argue that less harm, especially less tangible harm would justify less rigorous consent procedures, several scandals (Schneble et al., 2018, 2019; Shaw, 2016) using Big Data have revealed sensitive data about patients such as their mental disorders, health conditions, or simple data that a participant would not want to become visible or public (Kosinski et al., 2013; Reece & Danforth, 2017; Saeb et al., 2015). Shifting towards written consent where participants are presented with text-only or short video sequences or comic-like creations as has been proposed in contractual law (Brunschwig, 2001) and might be a good solution. Another solution that has found its ways into legislation is anonymization, anonymization in conjunction with broad consent has

been seen as the gold-standard for safeguarding privacy (Bernice S. Elger & Caplan, 2006; Jacobs & Popma, 2019).

Consent also plays an important role in the commercial sphere; in which users are presented with Terms and Conditions and the data policy of the service prior to signing up for a service . In contrast to consent in research those instruments have a binding character between the user and the service provider and are based on contractual law instruments. However here the same issue become present as in research with Big Data. Due to contractual law those terms and conditions are long and jargon-ridden and are seldom read thoroughly by users. Many scholars (Cate & Mayer-Schonberger, 2013; Favaretto, de Clercq, et al., 2020; van der Hof & Lievens, 2017) have argued that there should be a redesign of such instruments.

In the early days of Big Data, Anderson (Anderson, 2008) predicted the end of traditional research and advocated that research would be hypothesis free. Even though this picture has become less black and white, Big Data research is becoming more and more mainstream in the research community of various disciplines. Each of them has been confronted with domain specific problems (Favaretto, de Clercq, et al., 2020). However common to all of them is that at the moment at which data is being harvested for a specific project it is seldom clear what data will be useful and what insights could be gained from them. In terms of consent this would be only possible by obtaining broad consent, meaning being less specific on the use of the data as one would be in traditional clinical research. This is not a new phenomenon, and many researchers especially those in the field of biobanks, had to deal with the same issue (Elger et al., 2008). Especially keeping up-to-date with the patients has been proved to be cumbersome. Although many scholars have presented new ways of managing consent, such as the notion of dynamic consent (Wee et al., 2013), they are far from becoming standard.

The concept of consent as traditionally conceived is threatened by Big Data science. It is thus of importance to investigate how consent can be updated and autonomy safeguarded in the era of Big Data. This paper addresses this issue by presenting the results of our interviews conducted amongst lawyers in the US.

III. Methods

a) Sample

The data for this manuscript stems from a larger research project on the ethical and regulatory aspects of Big Data research. Within the framework of this project, we conducted interviews in the United States.

First, pilot interviews in the US were carried out upon recommendations of one of the international collaborators, a renowned expert in the field. In a further step, a purposive sample of interviewees was systematically recruited by identifying top law schools and focusing on lawyers with an involvement in the wider field of legal tech / Big Data.

b) Data Collection

A total of 15 interviews were conducted by the first author between January 2018 and September 2019. After completion of the first pilot interviews, the first author received constructive feedback on the pilot interviews from two senior researchers in order to ensure high-quality data collection.

The interviews were conducted online). Prior to the start of the interview, verbal consent was obtained from all participants and the interviewer briefly explained the purpose of the study, his role in the project, the confidential nature of the interview and participants were given space to ask questions. Each interview was framed by a semi-structured guide, which also allowed the participants to introduce their own topics.

The interviews lasted between 40 and 90 minutes and were conducted in German / French and English. This allowed the participants to choose their preferred language, which increased data quality. All interviews were recorded on tape and then transcribed word for word to allow qualitative analysis. Transcripts, when requested, were sent to participants for review. The transcripts were successively transferred to the qualitative analysis software MaxQDA (*MAXQDA: Qualitative Data Analysis Software*, n.d.) to support the analysis process (Guest et al., 2014).

c) Data Analysis

Applied thematic analysis was used for the data analysis. The aim of this method is to analyze and highlight thematic elements and patterns within the data, to organize, describe

and interpret the data set in detail (Beauchamp & Childress, 2013). Therefore, the transcripts were read in full length and independently analysed by at least two of the members of the research group. This first step of analysis consisted of open coding to explore the thematic elements in the interviews. Later, the members of the team met to explore the independent open-ended coding, discuss, and sort the identified topics.

IV. Results

For this study we analysed a subset of our findings that focused on consent, as consent can be seen as one of the predominant ethical tools used in human subject research and is entangled with several key ethical principle such as autonomy, responsibility and nonmaleficence (21). Thematic Analysis of the data sets showed two Clusters 1) Notice and Consent (Social Media / Internet Research) and 2) Consent in general Research that emerged from the data and their respective sub themes of autonomy, harm, legislation relating to consent.

a) Autonomy

Many researchers mentioned the problem that the use of the data is not foreseeable and thus the only possibility is to obtain only broad consent being valid indefinitely.

“Yes, I mean there is a lot of discussion on how problematic that is because you provide consent, and it applies for the rest of your life and you have no ideas what it will be used for what it could be used for in the future who it might hurt you etc. But there is no other way to do it you cannot require researchers to contact researchers every time they undertake a new query and so its problematic but basically, we do have consent form that says I authorize the use of my data for whatever purpose researcher might want to use it indefinitely.” (USL2)

As well as autonomy being threatened by unforeseen future use of data, complex consent terms also pose issues. In the realm of Internet research the Terms and Conditions tend to be long and difficult to understand especially for vulnerable groups.

“I don't think the users of social media are well informed in that sense and I think, if you don't try it to make them well informed that you can give them a general sense about information they are providing that may be used but it's not reasonable to ask most internet users to understand all the implications of data-processing about them.” (USL 17)

Still in the realm of Internet Research, one researcher refers to his own experience that staying up-to date is a burdensome and time-consuming task. And often in the realm of commercial products users are nudged to use a product / app.

“And frankly everyday people don't have the energy that's the main problem with notice and consent. They don't have the time to process a different choice. So, I tried to be an informed consumer, and, in the end, I had to say no I want a thermostat that I can program for my phone.”(USL 16) [The thermostat was mentioned as an example of an IOT device and the problematic of privacy and notice and consent]

Another participant underlined that consent requirements can make research difficult because participants may not volunteer and risk their privacy.

“What's important is that research is voluntary. Much of research where the participant must voluntarily agree to participate is altruistic and voluntary and you need to say yes. With the toll bridge you get something out of it you are getting the convenience of making your commute faster. If you take part of research you're not really, you're not getting much out of it. And so that's an area where people could say no I don't want to do it because I'm worried about privacy in a way that very few people would say about using the fast track to cross toll bridges. So I do think when privacy, when research involves voluntary subjects privacy will continue to be an issue as long as the subjects demand privacy.” (USL5)

Another researcher mentioned that strong philosophical focus on autonomy and subsequently the strong focus on control over data in Western DPO-Laws.

“There is a very strong philosophical basis behind a lot of DPO-law as well as in the US and EU. About people as autonomous agents who get to control their data. That's why we keep asking them constantly you consent to this, you consent to this. You consent to this etcetera etc. I don't think that model works.” (USL13)

One participant stated that, while harm is still important, it is important to focus on rights, autonomy being one of those rights.

“Yeah, I think no harm is an issue here. But I don't like to focus on harm I like to focus on, right, the rights of data subjects, and interests of all the people involved in the data in any kind of data activity. Rather than just as the American business community here, for a large part likes to talk about harm, and they say, Well, if you're not harmed, then we can do anything we want with your data. And they define harm burden naturally. And I don't think that's a way of moving forward with privacy regulations.” (USL1)

Another participant highlighted that secondary use of data poses a particular threat to autonomy, all the more so in the realm of Big Data where it is unclear how collected data will be used in future.

“And also, that that concept of future use, and try and figure out like if that really is as broad as any future use contemplated? Or if there's meaningful bounds? Like if it has to be consistent with our consent? Or has to be I don't, I don't know, I've never figured I've never seen anybody defined very meaningfully. But that is the problem. For me, I think the biggest issue is how do you meaningfully decide to find or put bounds on future uses of data so that it is consistent with what a reasonable person would expect, when they agree permitted that use? Does that make sense?” (USL9)

And mentioned that safeguards will become crucial.

“I would say probably that you, you really work to define the identification de identified and anonymization that you make sure to include security requirements, not just privacy, stringent security requirements”. (USL9)

b) Anonymization

One participant argued that, by removing the need for consent, anonymisation threatens autonomy, which could lead to a scandal.

“The thing with Big Data is and special with Big Data and the current rules and laws with respect to anonymization is privacy and research does not necessarily require volunteers anymore. And then I think it doesn't have that constraint to worry about privacy and then I think someday there's going to be a scandal And if it's part of enough person that could spark and make an outrage and make the public really unhappy and could lead to a backlash and could lead to stupid laws and bad regulation for primary research. So, I think part of it is in science's best interest to avoid that kind of backlash. And the backlash comes when reality and expectations are different and the public suddenly discovers they are different. You can change it be either making the reality different or by making the expectations different.” (USL5)

All participants agreed that anonymisation is increasingly difficult, but that nevertheless security precaution should be taken.

“ One issue with deidentification is that you never can a hundred percent prevent reidentification somebody who is motivated enough and skilled enough can reidentify at least a small percentage of records no matter what you do and there are people that have dedicated their careers in doing that. (...) So at least in the US we have laws that address anonymization the HIPPA privacy rule for entities that are covered by the HIPPA privacy rules, provides a lot of guidance as to how you deidentify information if you are going it to using it

without patient consent. And they have the eighteen identifiers they suggest removing or using some other proven techniques and so I think legal guidance is helpful. That is regulatory because a lot of people don't know how to do it. It's not a hundred percent of guarantee of anonymization no matter what you do." (USL2)

One of the respondents mentioned that one should keep in mind that the main purpose of anonymization is to prevent harm.

"Like the reason we think that anonymized, generally speaking, my understanding is the reason we think anonymization is a useful thing to do is because it's supposed to reduce the chances that use of the data is going to be harmful to the individual, like, or like traced back to the individual in a way that might cause harm to them. I don't think that that's a particularly good framing. I think that that is not true for two reasons." (USL 11)

One of the participants argued that it does not necessarily follow that anonymization should be neglected as one of the means to secure privacy in the law.

" My opinion is that anonymization is one way of protecting but it is not increasingly it is not as effective as it once was. I think anonymization is effective against the misuse of data by lay people, but anonymization is increasingly not effective as a protection of those who have themselves sophisticated effective software that can use existing publicly available databases to identify anonymized and deidentified data. So increasingly we have seen that it is possible using sophisticated software programs to take an anonymized dataset and be able to recreate the identity so that individuals by triangulating a lot of publicly available databases that contain information about identifiable individuals and increasingly I think that the thoughts of people in the data, that deal with data and risk management around data is that. The better regulatory approach. It's not to abandon anonymization because it is partially effective, and it will remain partially effective, but it is instead to penalize the misuse of data. Whether the data are anonymized or not."(USL6)

Another participant quote goes into the same direction. Anonymization as the sole mechanism is not the way to protect users' rights; it is important to set boundaries especially in the commercial field.

" If its anonymized. Again, my answer would be it depends on what the outcomes are. Right. Let's say that Facebook sells my data to another institution and that institution is known to have probably weak security and then that's institution gets hacked. And my data ends up in some Chinese database. And the Chinese government database. Right yes, I don't think Facebook should have sold that data. That data to that company not because Facebook needs to ask me permission when it sends data about me around to other companies. Because Facebook acting in concretely harm my interests and it should have known. So, you the question of third-party sharing is a bit complicated one, but my answer would always

be who is the third party or what are they going to do with that data. Is that going to be completely harming my interest and if so, let's just ban that." (USL13)

c) **Harm**

A participant stated that, even if harm cannot be prevented, legitimate interest should guide usage.

"The second principle is that they should not be using the data in ways that would in some sort of harm individuals. Unless individuals have consented to that in one way or another. And then the third thing is that if they use data to identify people and cause harm. Then there should be although there are legitimate interests such as credit rating and things like that issuance. Opening a new bank account. Because there are some credit worthy people and some not. There are legitimate business interests, in doing that. But other than that you know for example using an anonymized database to which no one has consented to identify people with particular diseases conditions or for marketing drugs for example I think that many people would think that's inappropriate. And should be penalized." (USL6)

As already mentioned, many participants linked anonymization to avoidance of harm. But one of the researchers with a strong practical focus made the point that in fact the debate seems to be theoretical rather than focusing on real world examples.

"But what I struggle with is the harm. To me it seems a lot of the harm seems theoretical vs. actual. And they are sure not physical. But it's not to say their reputational harm isn't important. So I don't know I think that anon.... I don't agree that anon. should mean that you can do whatever you want with data I guess that might be one protection that can be taken but I don't think if they remove it out of oversight of some form. Certainly anon. Is going to be a very fluid concept. And I think that participants should have the ability perhaps also with anon to request that their data is not going to be shared on a forward base. But I don't know it's a complicated question it's a great question.... You know the ability to link rep. harm and how do we incentivise people to let their data to be used." (USL9)

Some of the participants raised the issue of intangible versus tangible harm. This means that it is no less problematic as data collected and being used with the intention for benefiting the public could result in serious disadvantages for certain groups.

"I think we rejected that basic notion you know that has key experiments. No one is comfortable with that example and it's harder to see it in information because the harm isn't quite physical harm. But imagine you could think of a situation where a group of people are horrible disadvantages because of the use of their information even if there was a potential result in public good. I don't think norms have changed at this point, but I don't think no one of us are particularly comfortable with that people as guinea pigs. Even if the benefits are hard. So it's weird with data our example has already been mitigated and

researchers can take your biological sample and make it into a million dollar product.”
(USL9)

d) Legislation

Many of the respondents doubted that legislation’s focusing on notice and consent is the right approach to tackling issues around datafication.

“I am generally sceptical of the dominant American approach to DPO which is notice and consent. To the extent of the GDPR also requires this. I think it’s kind of silly because people click yes on everything. I would much rather have regulation on what types of services and products Facebook can operate rather than on what data they can access. So, I’m not terribly concerned on Facebook analysing all my data. What I am concerned with is Facebook using that data to make more addicted to the deploy them. So, I am not sure that is the response to your question. But hopefully that is some.” (USL13)

Another participant pointed out that, although safeguards are in place, research has shown that data can be moved from one entity to another without consulting data-subjects. The current HIPPA legislation has some weakness in that regard:

“There is basically no regulation so some of my work of the last few years has been showing how data brokers don’t really have to worry about HIPPA-rules. Because they can create proxies of health records completely outside of HIPPA regulation. So they can use medically inflected data. They can buy health data that has been legally allowed out of HIPPA. For example, the health departments some of them sell that data. So they can collect vast quantities of health data quite legally. They can collect medically inflected data, right.”
(USL3)

Another researcher pointed out that detailed regulations on how to obtain consent make the mechanism ineffective.

“And definition of consent is weak. If they signed a form that is 30 pages long and has long sentences saying this, well that is usually enough. So, there is not, there is some regulation particularly in the context of genomic biobanks and genomic databases. It probably provides some protection there is much less in the context of commercial operations and there is basically none when people have decided to make their information public. Even though their information is basically their family’s information. Especially if they are an identical twin. So if a twin makes its genome sequence public it made the sequence of her identical twin public. Without needing the identical twin’s consent.” (USL5)

The current US-regulatory environment is weak in terms of data protection; therefore, international declarations are very important.

“The US do not care much about them. Most Americans don't think there is such a thing and even American lawyers tend to not view it as particularly as really powerful except certain circumstances because the enforcement mechanism is usually very very weak. So the universal declaration of human rights which in most of the world is taken seriously at least someone's seriously It's not viewed certainly by real people but even by lawyers it's not viewed as having any particular legal power event though in certain circumstances it can. I don't think it's probably not novel. That powerful countries tend to ignore international law weaker countries tend to love international law.” (USL5)

Although such previously mentioned weak points exist. There are differences in various areas not only by type of data but also in terms of privacy, resulting in harm. Therefore, it is crucial to apply different standards to different fields/sectors.

“ Well, I, I don't know. I don't know about the laws regulating social data. I have not said that. I am willing to say that it's less regulated than biomedical data, because there's a definite governance structure. And when somebody sets up a biobank, there is a definite governance structure for that as well. So, I think social data is less regulated than by medical data. And having said that, I, I think both types of data should have should be regulated, published to the same extent, maybe different mechanisms. But I think the, the threat to privacy is probably equal, and not the threat to privacy. Well, the threat to well, the threat to privacy and the resulting harms that could occur from an invasion of privacy, and just, you know, just seeing how authoritative and governments are finding it. very instrumental in getting social data to control the citizens, that that alone tells you the employer. And so and then, you know, it's a whole concept people get good data, let's say government get your data or any entity, and then they could send you certain messages. Right, to control you with these messages. And that that is essentially what happened in my, in my elections. Now that it's targeted information, based on top data obtained from people. So, so the short answer to the question is, yes. All types of data need the same rigorous regulations.” (USL7)

“And so, data protection frameworks that essentially assume that the harm is to individuals rather than to sort of our code, but rather than a collective harm, are going to be not well suited to Big Data problems, because the whole thing about Big Data is actually the one piece of data that can get you anything, right. Like, that's the whole, the whole framework is the idea that like, Well, you know, we're going to do all these things, because we have all this data, and we can make all these predictions, because we have all this data and like, whatever, questions, I have a methodological rigor, rigor of that stuff, and I have a lot that,

you know, the basic framework is that like, one person's data is actually not meaningful. And so, you know, it's the collective gathered dataset.” (USL 11)

e) New Forms

Many of the subjects argued that although the burden of recontacting is problematic in traditional research, this issue becomes paramount in Big Data research.

“Yes, I mean there is a lot of discussion on how problematic that is because you provide consent and it applies for the rest of your life and you have no ideas what it will be used for what it could be used for in the future who it might hurt you etc. But there is really no other way to do it you cannot require researchers to contact researchers every time they undertake a new query and so its problematic but basically, we do have consent form that says I authorize the use of my data for whatever purpose researcher might want to use it indefinitely”. (USL 2)

Two respondents mentioned that there is still room to implement a better mechanism by automating the consent process.

“Yeah, positive notification is just going to overwhelm me. The way that cookie tracking notification or other website banners, that’s unhelpful. I agree that users should be notified at all but users how want should see information about them. And in a way that it can be automatically processed. But that depends on infrastructure supporting tools. That can receive this information in a form they can work with or act as proxy on their behalf and take appropriate actions.” (USL17)

A completely different approach is suggest by USL11 how opts in favour of setting up a board of trustees that act on behalf of the participants. As it has been successful in the field of community work.

“Therefore, you know, one of the things that we have thought about is like how you create, create advisory methods, where you can gain input from people who are in some ways representative of the people who are present in the dataset, especially the folks who are at highest risk, were present in the data set. To, you know, get it's not consent, but get some sort of like, process for getting more information about how to approach problems or what people might want. So I think that's a promising model for certain types of Big Data research, because it creates, like, because it sort of splits the difference between individual consent by all people included in the data set, which is just not feasible for some circumstances. And like, Oh, well, you know, we don't have to get individual consent, therefore, we don't have to care about what people in the dataset think at all. So I think that's a model I think is promising.” (USL 11)

V. Discussion

Consent and subsequently informing participants prior to involving them in a research project has been key for sound and ethical research. Many research guidelines, such as the Common Rule (Common Rule, 2020) and the Belmont Report (Research, 1978) define detailed and clear criteria on how consent should be sought in medical research. In 2020 the Revised Common Rule defined the following six general requirements. It shall be obtained before involving participants in research, must minimize coercion and give participants the opportunity to resonate, it shall be written in an understandable language. Undoubtedly consent plays an important role in traditional research. However, the use of Big Data and subsequently the collection of large samples of “scraped” internet data questions this mechanism for several reasons, as the number of participants hardly justify the extra burden which researchers would have to deal when collecting large samples. This leads to the effect that the participant-researcher relationship cannot be as close as it is in clinical research, resulting in potentially harmful situations for participants. If the paradigm of getting consent needs to be changed and to be effective for data-science, it is of crucial importance to discuss the ethical principles that are affected.

a) **Autonomy**

The paramount principles in the realm of consent is autonomy. Many of our participants mentioned this topic. Autonomy in this context means that participants of research as well as user should be truly informed before they take part in a study or sign up to a service. However, some participants suggested that this ideal of autonomy is unachievable when the traditional model of consent is applied to the context of Big Data, for two main reasons, one relating to users (participants) and the other to use of data.

First, many participants in Big Data research are simply users of social media who do not (really) read the terms and conditions that they are signing up to, and thus could be participating in Big Data research without even realising (Cate & Mayer-Schonberger, 2013; Kaye et al., 2015; Schneble et al., 2020a). Although such users have consented in the legal sense, ethically, this process does not safeguard their autonomy. In addition, even if such users do actually read the small print, they are unlikely to have enough of a grasp of Big Data research to understand that anonymisation might not be possible if their data is combined with other sources (Vayena et al., 2013).

Second, even when participants are actually presented with specific information and a consent form for a specific project, they are unlikely to have any grasp of the myriad possible avenues of future research projects that might use their data if the form also includes broad consent to future use. Broad consent respects autonomy inasmuch as some participants will be happy for their data to be used in all types of research, but disrespects it in the sense that it constitutes a waiver to specific consent for future projects (Mikkelsen et al., 2019; Steinsbekk et al., 2013). Even if institutional review boards or research ethics communities review any such future studies – which is less likely in Big Data and social media research than in biomedical studies that reuse data – the use of broad consent pays lip service to traditional concepts of consent and autonomy rather than respecting them properly. Overall, the broad view of our participants was that, given the twin dilemmas of failure to read/understand at the initial point of consent and the ethical failings of broad consent in this context, a different approach to consent might be required. In the following section we discuss possible alternatives to consent, that lower expectations regarding consent as the primary safeguard.

b) Consent Based on Types of Data and Anonymization

Our participants also expressed the view that, with the increasing numbers of participants in Big Data research, managing consent becomes a significant burden. Our participants suggested that it might be better to denote certain types of data which should be subject to stronger protective measures than other types of data. Thus, participants regarded data stemming from biological samples and genetic and health data as one group of data that should be subject to stronger safeguards and therefore also need stricter consent mechanisms. For other types of data, such as for example data from the social media Sphere (Twitter/Reddit) less strict consent mechanism would be considered acceptable.

However, data that is subject to less strict consent mechanism is still prone to disclosing sensitive matters as multiple studies using data from social media have shown. The same applies to anonymised data for which, in most jurisdictions, no new approval was required if the data were to be made available for further use. However, as various studies have shown (Tene & Polonetsky, 2012), this paradigm is no longer valid as it is possible to infer the identity of a person with just a few calculations or data matching. Although there are precise guidelines for anonymisation, especially in the area of HIPPA, some of the interviewees consider them too complicated and useless, as identification remains easily

possible. This fact calls for new frameworks to be developed that take into account the notion that secure anonymization has become extraordinarily complex. Newer frameworks like differential privacy (Abadi et al., 2016) could eliminate some of the concerns raised. Furthermore, when anonymization becomes obsolete it is key that it is replaced by a trust between researcher and participants. In case of Research-Projects that use data that is collected by the researchers themselves and participants consent directly this might be easier. But if researchers harvest data from public sources as in Internet Mediated research things become fuzzier; is it the responsibility of the researcher or should the platforms inform users about possible consequences, or is this just part of today's digital literacy and citizens should be aware of the risks, as they are in case of crossing a road in everyday life? In the end, however, the burden must remain on researcher and it is thus of importance for researcher to weigh out the risk of reidentification against the social goods resulting from research.

c) Bad legislation / Missing Oversight

Participants in our study repeatedly pointed out that the law also plays a decisive role. In particular, they pointed out that the Health Insurance Portability and Accountability Act (HIPAA) legislation covering medical data which requires consent when data is processed outside the protected entities has been shown to be weak in terms of respecting safeguards. Participants as well as several scholars stressed that it is possible to transfer data from within the HIPAA boundaries and build proxies legally outside the HIPAA realm (Terry, 2014).

Research with social media and mhealth is governed by the Food and Drug Administration (FDA) and Federal Trade Commission (FTC) (which regulates the big tech-companies). These regulations influence the consent and notice procedure that users of such services need to agree with. As contractual law they are not designed with an emphasis to be easily understood but rather to safeguard the interests of the company. One point emerging from our interviews is that research participants often provide legally valid consent that is nonetheless extremely ethically suspect.

Another sizable subset of Big Data research not subject to the Common Rule is research supported exclusively by private funding. The sharp distinction between publicly and privately funded research results in inconsistent oversight, as many privately funded research activities carry the same types of risks as research funded by the government. In particular,

many research projects are multinational and multicentre, but most regulatory laws are national in scope and IRB approval is even institutional. This leads to a complicated set of rules that are hardly understood and respected. The research community would thus benefit from clear guidelines of a global character.

d) Regulating Use / Stronger IRBs

As one of the participants correctly said there is “no one size fits all” approach for obtaining consent. As discussed earlier the burden of getting truly informed is becoming insurmountable or not possible for researchers. However this should not result in a Big Data Exceptionalism by watering down research ethics (Rothstein, 2015). Vayena and colleagues emphasize (Vayena et al., 2016) the need of a shared distribution between researcher, funding agencies and IRBs. Steinman and others propose a model based on different more balanced principles that should guide researcher through their design process of their research and making it them less dependent on consent.

Other forms as pointed out by our participants could be to set up proxy boards of a subset of the potential participation sample and use their positive or negative consent as base for the evaluation of a research project. In the realm of health and genetic data it might be beneficial to move data to trusted bodies that govern over the data. This would allow better control especially of secondary use of such data. Last but not least more effort should be put towards the digitalisation and automation of the consent process.

Another approach would be to regulate use, as some of our participants argued that the shift in society and the ubiquitous use of data in many of today’s products, services and research is a process that cannot be avoided. Data protection laws should therefore not regulate the collection of data but rather control data use independently of consent. This approach, mostly referred to as Big Data exceptionalism, has been widely discussed in the scholarly community (Nissenbaum, 2017). From a consent perspective, this approach has an impact on several levels. First, autonomy is taken away from the participant; while it might be beneficial for some vulnerable groups and reduce harm, it might deprive participant of making their own choice, which depends on individual beliefs and influence personal risk assessment. Second, it puts additional burden on the researcher and subsequently Institutional Review boards that have to assess research and build up their knowledge for new usage-types.

However, the situation remains complex and there are other arguments in favour of regulating use. On the one hand, there are researchers and companies using data from participants and users; putting additional burden on them might hinder research which is often not in the public’s interest and might prevent relevant issues that affect certain vulnerable groups being investigated. Research, especially human subject research has a long

tradition in IRB's reviewing research. It is in fact then a question of establishing new frameworks and defining use cases that match certain types of research especially those where harm is more severe than in other types (Favaretto, De Clercq, Gaab, et al., 2020).

VI. Conclusion

Research needs informed consent, but the effort to achieve consent for researchers should be proportionate, especially when little or no harm results for data subjects from the usage and aggregation of data. This is especially the case when it comes to public data. The use of new technologies and automation offers great potential, especially when it comes to automating the consent process. Dynamic consent portals (Budin-Ljøsne et al., 2017) have already been proposed but have not yet been able to establish themselves. Such portals as well as data cooperatives (Hafen 2015) are only a small step in the right direction. In an environment where people have to behave more and more reactively, it is about taking the burden away instead of imposing more. It will be central to automate the use of data. It would be conceivable to create individual profiles of individuals' preferences based on a questionnaire and then only release data if there is a positive match between the query and the profile. This would have advantages for the user but also for the researcher who would only have to define the danger fields and access the ethical automated data-sharing and collection fields. On the other hand, given that getting broad consent may be the only viable solution in some cases it is paramount to foster IRBs and to staff them with enough expertise to deal with upcoming projects, whether they are obtaining new data or re-using data under broad consent.

VII. References

- Abadi, M., McMahan, H. B., Chu, A., Mironov, I., Zhang, L., Goodfellow, I., & Talwar, K. (2016). Deep learning with differential privacy. *Proceedings of the ACM Conference on Computer and Communications Security, 24-28-October-2016*. <https://doi.org/10.1145/2976749.2978318>
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. <https://www.wired.com/2008/06/pb-theory/>
- Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics*. Oxford University Press.
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative*

-
- Research in Psychology, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brunschwig, C. R. (2001). *Visualisierung von Rechtsnormen : legal design* (Vol. 45). Schulthess.
- Budin-Ljøsnø, I., Teare, H. J. A., Kaye, J., Beck, S., Bentzen, H. B., Caenazzo, L., Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M. K., Jones, E., Katić, V., Simpson, A., & Mascalcioni, D. (2017). Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics*, 18(1), 4. <https://doi.org/10.1186/s12910-016-0162-9>
- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67–73. <https://doi.org/10.1093/idpl/ipt005>
- Elger, Bernice S., & Caplan, A. L. (2006). Consent and anonymization in research involving biobanks: Differing terms and norms present serious barriers to an international framework. In *EMBO Reports* (Vol. 7, Issue 7). <https://doi.org/10.1038/sj.embor.7400740>
- Elger, Bernice Simone, Biller-Andorno, N., Mauron, A., & Capron, A. (2008). Ethical issues in governing biobanks: Global Perspectives (Bernice Simone Elger, N. Biller-Andorno, A. Mauron, & A. Capron (Eds.)). Ashgate.
- Favaretto, M., De Clercq, E., Gaab, J., & Elger, B. S. (2020). First do no harm: An exploration of researchers' ethics of conduct in Big Data behavioral studies. *PLoS ONE*, 15(11 November). <https://doi.org/10.1371/journal.pone.0241865>
- Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0228987>
- Guest, G., MacQueen, K., & Namey, E. (2014). Applied Thematic Analysis. In *Applied Thematic Analysis*. <https://doi.org/10.4135/9781483384436>
- Hafen, E. (2015). MIDATA cooperatives - citizen-controlled use of health data is a prerequisite for big data analysis, economic success and a democratization of the personal data economy. *TROPICAL MEDICINE & INTERNATIONAL HEALTH*.
- Ioannidis, J. P. A. (2013). Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research. *The American Journal of Bioethics*, 13(4), 40–42. <https://doi.org/10.1080/15265161.2013.768864>
- Jacobs, B., & Popma, J. (2019). Medical research, Big Data and the need for privacy by design. *Big Data & Society*, 6(1), 205395171882435. <https://doi.org/10.1177/2053951718824352>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics : EJHG*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>
- MAXQDA: Qualitative Data Analysis Software. (n.d.). MAXQDA. Retrieved June 15,

- 2019, from <https://www.maxqda.com/>
- Mikkelsen, R. B., Gjerris, M., Waldemar, G., & Sandøe, P. (2019). Broad consent for biobanks is best-provided it is also deep. *BMC Medical Ethics*, 20(1). <https://doi.org/10.1186/s12910-019-0414-6>
- Nissenbaum, H. (2017). Deregulating Collection: Must Privacy Give Way to Use Regulation? *SSRN Electronic Journal*. <https://www.google.com/search?client=firefox-b-d&q=Deregulating+Collection%3A+Must+Privacy+Give+Way+to+Use+Regulation%3F>
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- Research, T. N. C. for the P. of H. S. of B. and B. (1978). Belmont Report.
- Rothstein, M. A. (2015). Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics. *The Journal of Law, Medicine & Ethics*, 43(2), 425–429. <https://doi.org/10.1111/jlme.12258>
- Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research*, 17(7), e175. <https://doi.org/10.2196/jmir.4273>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2019). All our data will be health data one day: the need for universal data protection and comprehensive consent. *JMIR Preprints*. <https://preprints.jmir.org/preprint/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). All our data will be health data one day: the need for universal data protection and comprehensive consent (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/16879>
- Shaw, D. (2016). Facebook's flawed emotion experiment: Antisocial research on social network users. *Research Ethics*, 12(1), 29–34. <https://doi.org/10.1177/1747016115579535>
- Steinsbekk, K. S., Kare Myskja, B., & Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem. *European Journal of Human Genetics*, 21(9). <https://doi.org/10.1038/ejhg.2012.282>
- Tene, O., & Polonetsky, J. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. *Stanford Law Review Online*, 64.
- Terry, N. P. (2014). Big Data Proxies and Health Privacy Exceptionalism. In *Health Matrix: The Journal of Law-Medicine* (Vol. 24).
- Common Rule, (2020).
- van der Hof, S., & Lievens, E. (2017). The Importance of Privacy by Design and Data Protection Impact Assessments in Strengthening Protection of Children's Personal Data Under the GDPR.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3107660

Vayena, E., Gasser, U., Wood, A., O'Brien, D., & Altman, M. (2016). Elements of a New Ethical Framework for Big Data Research. *Washington and Lee Law Review*, 72(3).

Vayena, E., Mastroianni, A., & Kahn, J. (2013). Caught in the web: informed consent for online health research. *Science Translational Medicine*, 5(173), 173fs6. <https://doi.org/10.1126/scitranslmed.3004798>

Wee, R., Henaghan, M., & Winship, I. (2013). Dynamic consent in the digital age of biology: online initiatives and regulatory considerations. *Journal of Primary Health Care*, 5(4), 341–347. <http://www.ncbi.nlm.nih.gov/pubmed/24294625>

Chapter 6: Data Protection Laws in the Age of Big Data: A Qualitative Study with Lawyers and Data Protection Officers in the United States and Switzerland

Additional Paper (Not Published)

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, David Martin Shaw, PhD¹, Mark, Rothstein², PhD, Bernice Simone Elger¹, MD, PhD,

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²Departement of bioethics, University of Louisville

I. Introduction

The use of Big Data has invaded our daily lives. Big data holds the promise of resolving many pressing questions in commercial and research settings. In fact, the use of Big Data to reveal individual or group patterns and the use of predictive analysis are already effective in several areas ranging from building smart hospitals (Mertz, 2014) and cities (Hashem et al., 2016) to fighting climate change (Faghmous & Kumar, 2014) and to predict individual health related topics. Big data thus touches the core of today's massively interlinked and connected society; this omnipresent use raises societal questions and also affects individual rights, raising many issues for regulators.

After several scandals involving Big Data, ethical aspects of the use of such large data sets have gained publicity both in academia and the public sphere. Scandals affected not only to the commercial sphere but also the research sphere. One example from the research setting involved the use of social media data from 700 thousand individuals without obtaining participants consent; this was the Facebook emotional contagion experiment (Hunter & Evans, 2016; Shaw, 2016). Another example from the commercial field, the Cambridge Analytica scandal, involved harvesting personal data without consent for political advertising (Schneble et al., 2018). In Google's "Nightingale" project, sensitive health data from 50 million patients was transferred from Ascension (one of the larger healthcare providers in the USA) to Google without the patients' knowledge or consent (Schneble et al., 2020b). Common to the above-mentioned examples of ethical misconduct is the fact that responsible actors in these cases were mainly not respecting autonomy of the individuals – thus not getting their consent.

These incidents challenge the core tenets of privacy laws around the globe. Current guidelines and legislation do not offer adequate safeguards to protect individuals' rights in the context of Big Data. Data protection law (with some exceptions) still assumes that data collection about an individual person should focus on data minimization (at least in the European Data Setting) (General Data Protection Regulation, 2016). However, Big Data methodologies use vast amounts of data (in terms of data points) and because predictive algorithms and artificial intelligence (AI) outperforms well if applied to large and interconnected data sets, this means that data and sometimes also sensitive data from various sources may be necessary and its use should be legal. One of the traditional bases has been notice and consent (Terry, 2014). However, notice and consent, including informed

consent in research, presume that participants are informed sufficiently about the use of their data. This has proven to be difficult for two reasons. First, the sheer number of participants in Big Data studies makes it difficult to get consent of all the participants and subsequently respect their choices and, if applicable, provide follow up. And second, the ambiguous nature of potential future studies makes it impossible to address issues at the initial point of consent, making broad consent the only alternative. Therefore, obtaining traditional informed consent is becoming increasingly difficult.

If the old model of notice and consent is no longer adequate in the Big Data setting, the question remains of how it should be regulate meaning how to prevent harm to participants and the public. One possibility could be *Big Data exceptionalism* (BDE), which has various meanings and applications. One often used approach is that the use of data, rather than the way it is collected, should be regulated (Nissenbaum, 2017). Another approach uses BDE to denominate the difference in regulation of conventional research versus research with Big Data (Rothstein, 2015). Finally, BDE can refer to the fact that there are different laws for different kinds of data (health data /genetic data), especially in the United States because narrow and sector-specific privacy laws foster exceptionalism (Rothstein, 2015).

The concept of exceptionalism is not new in the wider field of privacy. During the rise of genetic technologies there were discussions about genetic exceptionalism, meaning that genetic information should be regarded as different from other patient data. Although amongst academics this concept has been often criticized (Evans & Burke, 2008; Murray, 2019) and other concepts like genetic contextualisms were proposed (Garrison et al., 2019), it found its way into laws and regulations. It has been implemented, for example, in the Swiss federal Act on Human Genetic Testing¹ and by the United States Genetic Information Nondiscrimination Act (GINA)².

Exceptionalism has also been part of the discussion around health data as mentioned before. In the United States exceptionalism is one of the predominant concepts of the legislative

¹ Art of the HRA States: “This Act applies to research concerning human diseases and concerning the structure and function of the human body, which involves: a. persons; b. deceased persons; c. embryos and foetuses; d. biological material; e. health-related personal data.” It defines further in Art 3: “*Genetic data* means information on a person's genes, obtained by genetic testing”

² The act bans the use of genetic information in health insurance and employment: it prohibits group health plans and health insurers from denying coverage to a healthy individual or charging that person higher premiums based solely on a genetic predisposition to developing a disease in the future, and it bars employers from using individuals' genetic information when making hiring, firing, job placement, or promotion decisions

environment (Terry, 2014) as evidenced in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule³, which regulates the use of health, but not other, data. Critics of sector-specific regulation note that health data produced outside of covered entities (under HIPAA) are subject to little, if any, protection, despite health data often contain sensitive data about individuals and could result in harm.

The laws of the United States and continental European law tradition take a different approach. The United States sectoral approach regulates privacy and the use of data differently in various sectors while the European and Swiss umbrella approach applies the same rules regardless of whether data are processed by the government, individuals, or commercial entities.

Big data challenges the core tenets of current privacy laws, and there is a growing need to develop up-to-date laws and more specific data ethics for use in the commercial and research field. For this reason, we conducted expert interviews in Switzerland and the United States to analyze two different regimes. This paper compares the expert opinions and questions whether Big Data exceptionalism is a valuable alternative to addressing the shortcomings in the current regulatory environment or if there are other possible building blocks to develop more appropriate regulations.

II. Methods

a) Sample

The data for this manuscript stems from a larger research project on the ethical and regulatory aspects of Big Data research. Within the framework of this project, we conducted 35 semi-structured interviews with cantonal data protection officers (DPOs)⁴ and hospital lawyers of the largest hospitals in Switzerland and lawyers in the United States.

The United States was chosen as a comparator country because it has a different data protection/privacy regime. Participants selected in Switzerland were regulatory officials (DPOs at cantonal levels) or the DPOs of the major university hospital (a highly regulated environment in Switzerland).

³ The 110th United States Congress. An act to prohibit discrimination on the basis of genetic information with respect to health insurance and employment. *The 110th United States Congress* (2008) 110–233.

⁴ Like the US system the Swiss system has also a federalist approach leaving the canton (would be equal to states in the US) many freedom in having their own law in certain areas.

First pilot interviews in the US were carried out upon recommendations of the middle author, a renowned expert in this field. In a further step, a purposive sample of interviewees was systematically recruited by identifying top law schools and focusing on lawyers with an involvement in the wider field of legal tech / Big Data.

b) Data Collection

The interviews were conducted by the first author between January 2018 and September 2019. At the time of the interviews, the first author was a PhD student at the Institute for Bio- and Medical Ethics at the University of Basel. After completion of the first pilot interviews, the first author received constructive feedback on the pilot interviews from two senior researchers in order to ensure high-quality data collection.

The interviews were conducted at a time and place chosen by the interviewee (usually in their office for Switzerland, and remotely by Skype for the United States). Prior to the start of the interview, verbal consent was obtained from all participants and the interviewer briefly explained the purpose of the study, his role in the project, the confidential nature of the interview and participants were given space to ask questions.

Each interview was framed by a semi-structured guide, which also allowed the participants to introduce their own topics. The interviews lasted between 40 and 90 minutes and were conducted in German / French and English. This allowed the participants to choose their preferred language, which increased data quality. All interviews were recorded on tape and then transcribed word for word to allow qualitative analysis. Transcripts, when requested, were sent to participants for review. The transcripts were successively transferred to the qualitative analysis software MaxQDA⁵ to support the analysis process (Guest et al., 2014).

c) Data Analysis

Applied thematic analysis was used for the data analysis. The aim of this method is to analyze and highlight thematic elements and patterns within the data, to organize, describe and interpret the data set in detail (Braun & Clarke, 2006). Therefore, the transcripts were read in full length and independently analyzed by at least two of the members of the research group. This first step of analysis consisted of open coding to explore the thematic

⁵ MAXQDA: Qualitative Data Analysis Software, MAXQDA.

elements in the interviews. Later, the members of the team met to explore the independent open-ended coding, discuss, and sort the identified topics.

III. Results

In the following sections we present the topics that emerged from our thematic analysis and that are relevant to the discussion of BDE. While we found some country specific differences, interviewees from both the US and Switzerland were divided amongst themselves as to the best oversight of data and used various arguments for and against regulating data use. Our investigation reveals a complex situation: the majority of the participants in the United States are in favor of regulating the use of data and letting entities collect whatever and as many data they want, while many of the Swiss participants prefer the paradigm of notice and consent and data minimization.

a) Regulating Data Use

Many of the participants think that regulating the use rather than the collection is the base of a good regulation to address misconduct. One definition of BDE is given by the following participant, who is also in favor of setting up a body that controls and approves the data. This approach could provide oversight similarly to the way institutional review boards set up by various universities provide ethical clearance of other types of research:

“I think the most important issue is how data is used. We need to focus not on what can be extracted but on how the data is used. [...] Because we know that data can be used for good purposes and for bad purposes, so if you know. If you focus on extraction, it’s very difficult to ensure that the good uses are done and the bad uses are not. Because if you focus on extraction, you’re too far removed on how the data will be used to be able to regulate it intelligently. So the focus should be on uses. And to that end, my view is there should be a regulator tasked specifically with approving uses of data across the economy. And if we had something like that, we would really be able to leverage the benefit of data without the harm. So I think the whole focus on privacy is misguided.” (USL-14⁶)

Strongly intertwined with defining bad and good use is the topic of harm, which is a major focus of the discussion on BDE. One of the participants pointed out that harm in terms of Big Data is collective rather than individual:

⁶ Participants interview are pseudonymized according to the following scheme <country><sample group><consecutive number> - In the case of CHLR07 this means Swiss group, lawyers, interview 07, in the case of USL07 this means US group, lawyers, interview 07

“The overall thing that I could think about especially in the usage of Big Data is harm. The harm that comes from Big Data is not essentially individualized nor preventable by anonymization. And so a data protection framework that essentially assumes that the harm is individual, rather than a collective harm, is not well suited to Big Data.” (USL-11)

The same participant pointed out that harm cannot be managed simply by putting efforts on standard techniques of anonymization and subsequently the minimization of data collection as opponents of BDE argue. Another nuance was mentioned by USL-2, who advocates that harm is mainly caused by discrimination and commercial interests.

“Another area that I’m always concerned about which is not privacy per se, it’s discrimination because the big concerns, and we haven’t really talked about that, but the reason you don’t want people to use your data is because they might deprive you of opportunities. It might hurt your employment prospects, insurance prospects, business prospects, or whatever so you also have to have point of use regulations that prohibit use. That’s an area you can’t ignore because just saying don’t collect this information from ethicists point because it’s simply out there it is so widespread so making sure don’t use it to hurt other people is really important.” (USL-2)

His conclusion is to identify uses that harm people and prohibit those, adding that the information is already out and not collecting information would be simply infeasible.

The Swiss approach focuses on the collection of data rather than its use. One of the Swiss participants sees this as a weak point and advocates stronger mechanisms for data sources to control what happens with their data instead:

“The law should always reflect the current technological state and just strengthen the power of the person so that they can decide more about what happens with their data.” (CHL-4)

Another US participant emphasizes that it is important to regulate use. Especially with the ever-growing volume of the data and the linkage of different data sources the collection of today’s non-sensitive data could turn out to be sensitive (e.g., health data).

“Right, so point of use regulations are important in simply understanding that if we are talking about health information it’s not only information for doctors or hospitals. People can infer all sorts of things from all sorts of data so almost anything can be health data. So we have to be really thoughtful about regulations and not assume that the world was the way it was 20 years ago where you just say ok doctors can disclose information you covered it because your Facebook information can be very illuminating health data.” (USL-2)

b) Skepticism on Regulating Data Use

One participant argues that it would be beneficial to define deidentification and anonymization.

“You should really define deidentification and anonymization. Be sure to include security from the beginning and not only privacy, also the concept of future use and try to figure out things and broaden any future use or if there are useful bounds or if they are consistent. I have never seen somebody defining meaningfully define the balance on future use of data consistent with what a person wants to expect when they agree to permit that use. But I don’t know how you do that.” (USL-9)

Another participant points out that there are some use cases that are highly problematic. The collection of such data to be used for that purpose should simply not be allowed, meaning that the commercial providers should set high security standards.

“For something like the facial recognition system, there are many uses that are relatively innocuous and maybe you want to allow those. What we really want to prohibit is a facial recognition database of three million images being used to mark every single person to walk out on the street in the suburbs. But think also what you want to prohibit programming ten friends in your smart door. Maybe you do, that’s a choice you want to make. But don’t make it by default as you prohibit other things. So it’s an area where law has to be nuanced because people can’t do much for themselves. And it’s an area where there is a ton of value to allow certain use of data.” (USL-16)

c) Building blocks of an ethical Big Data law

At the end of the interview participants were asked what the building blocks of a privacy / data protection law should be. Their answers also shed some light on BDE. Within the sample we could identify different themes including 1) Institutional aspects 2) Harm 3) Notice and consent / Informed Consent 4) Societal Aspects / Freedom rights 5) and the issue of staying up to date with the law. In the following subsections, results will be presented for each of those categories.

Another participant pointed out that the concept of individual harm is outdated with respect to today’s use of Big Data and predictive analytics. He concluded that we have to think about new models to integrate those aspects.

“A data protection framework that essentially assumes that the harm is individual, rather than a collective harm, is not well suited to Big Data. Because the whole thing to Big Data is actually not one piece of data that’s the whole thing [...] The basic framework is that one person’s data is actually not neutral, it’s the collective gathered dataset. And so I think,

thinking about if I were proposing data protection and particular ethical obligations with regard to Big Data usage, I would try and work it whenever there are other types of collective harms.” (USL-11)

One of the participants points out that the perception of harm really depends on the cultural perceptions of individual beliefs. Defining generalizable harm as it would be the case for “bodily” harm is difficult in the Big Data setting.

“Every value is different, that’s all about ethics. It’s very community based. Thankfully we live in diverse world. That’s why it’s difficult to enforce. When they are harmed, they have their own rights according to their own beliefs.” (USL-9)

Notice and consent still plays an important role, especially when it comes to research, but also in commercial settings.

“One of the most important aspects is informed consent. What does a person need to know about the data in order [for] consent to be informed. Like 30 years [ago] we realized the basic blocks of informed consent for biomedical research. And when genetics research came, we added more elements of information a participant needed to know before they gave consent for their biological sample to be used. Now we need to understand better what those people need to know in the realm of informed consent for Big Data.” (USL 7)

USL-13 describes the tension between freedom /self-determination and paternalistic protection of citizens. Ultimately, it’s all about preventing harm:

“Treat data as you treat any other consumer good. Like let as a general matter people to make whatever choices they want to make, but don’t let them buy exploding toasters.” (USL-13)

Finally, the right to informational self-determination was mentioned by two Swiss participants:

“I think the most important thing is still that I have this self-determination, so the right to informational self-determination must not be left out of the data protection law, or, the principle is that I should know transparently what others do with my personal data.” (CHLR03)

Another Swiss participant took this thought further, arguing that citizens should have a participative right:

“You should get to the point where you say that the owner is still the patient who gives his data. But at the moment we have the regulation that the institution is responsible for the data, but I think the patient should have a participatory right, which goes beyond that, simply giving consent for any processing - it goes on, after all, so we also want to

commercialize and that we also want to use it for research, where the patient should also participate.” (CHLR10)

d) Institutional and Societal Aspects

Data protection clearly depends on the institutional setting as many of the participants in the United States and Switzerland state:

“I would build a clear agency staffed with experts including those in psychology who can value the effect of certain kind of data use on people’s wellbeing and regulate accordingly.” (USL-13)

On the one hand, it is a matter of funding and staffing how well the DPO office is equipped but also an organizational issue and a question of how enforcement mechanisms are set up. Is the DPO simply consulted when a new project is set up, or do they have enforcement mechanisms?

“But having a guideline does not mean that it is followed. And there is a lack of resources to implement it well in the organizations and also to take measures if the guidelines are not followed.” (CHLR08)

One of the interviewees points out that it is fundamental to have societal discourse where the development should head. This gives legislators the foundation to build privacy laws and finally to protect data.

“Figure out what your people care about and then starting with that in discussion with government, researchers, agencies, corporations, etc., to figure out the best way to protect as far as possible what your people care about it. [...] So privacy is a topic that varies a lot from place to place and I think you need to start with what do your people care about. (USL-5)

The high pace in development of new technologies demands the law to keep up to date. The lifecycle thus remains shorter. The following quote of USL8- illustrates this issue.

“That’s a Big Data question. I don’t know how viable it is, but I think the idea is that the laws need to be flexible enough to try to keep up with the technological developments. If you write law for today’s technology, it is hard to look ahead at what tech could be up in five years or ten years. But if you only write the law for today it will be obsolete in a couple of years. The challenge is how to write the law, so it has real power beyond a couple of years. Because otherwise you will only catch up and people will always be scrutinizing the laws. You have to have shorter lifecycles, but in the world of politics, you know it’s difficult.” (USL-8)

IV. Discussion.

As Nissenbaum describes in her article “Deregulating Collection: Must Privacy Give Way to Use Regulation,” (Nissenbaum, 2017) proponents of BDE use the narrative that in the age of Big Data, collection cannot be prevented, and harm arises primarily by the (mis)use of this data, for example, by surveillance and racist profiling or discrimination based on socioeconomic factors. It is thus not the presence of data and or its collection that should be regulated, but rather the use of that data. As Nissenbaum argues, however, it is difficult to draw a clear line between collection and use. BDE being a valuable concept depends on the distinction between collection and use, which is difficult to delineate.

a) The Swiss, the EU and US Context

Many of our participants, especially those from the U.S., were in favor of regulating the use rather than collection. They argue that in today’s interconnected world there is no way back and data collection continues. The fact that data becomes valuable only after being processed and aggregated and built in to predictive and or AI-models is a strong argument for BDE. As data are used multiple times the secondary use of data needs more attention in the future. One point that needs to be addressed is to guarantee oversight. Data processors in the commercial as well as the research settings need to think about new strategies for how users can keep track of where their data is used. A solution could be the use of dynamic consent models also in the commercial sphere (Kaye et al., 2015).

There is also another facet to consider, which may explain the differing perspectives between our two groups of participants: digitalization and the use of Big Data for commercial purposes in Switzerland is far behind such use in the United States. Thus, one could argue that the Swiss system has not been exposed to the questions of Big Data as much as the U.S. has. Whilst many of the participants agreed that the introduction of the GDPR has had a major impact on US companies and that the GDPR can be seen as progressive in terms of its concept and protection of the data subject, experts were unanimous in their opinion that such a system in the U.S. would not be possible primarily because of the political system. They also pointed out that the federalist approach, in which each state is responsible for data protection, makes it difficult to find a uniform regulation. However some states as California with its California protection Act have recently introduced a quasi-GDPR like data protection. As far as the federalism is concerned, Switzerland and the USA have parallels. The Swiss data protection regulation is subdivided in two parts

each state (canton) is responsible for its institutional law (that governs, hospitals, universities and other public institutions) and the commercial part that is overseen by the federal government. Many experts argued that this subdivision associated with few financial and human resources is one of the major challenges.

From an ethical point of view BDE may seem appealing. The massive collection of data makes it very difficult to determine whether data is sensitive or not as simply the interlinkage of non-sensitive data as, for example, pollution data combined with more sensitive data, such as location data, could generate highly sensitive health data (Schneble et al., 2019b). However, practically speaking, it is difficult to see how use of data can be overseen effectively.

Institutional review boards evaluating research in academia have been proved to be highly efficient assessing ethical aspects of research projects. Over the years they have been dealing with several new technologies such as genetics and nanotechnology. However, IRBs and research ethics committees still tend to review research projects before they begin, and it would be a major paradigm shift for these or similar bodies to evaluate data use on an ongoing basis. More generally, moving away from notice and consent poses threats to transparency; if consent is not obtained and data subjects not informed, any review bodies will have to satisfy themselves that they are aware of all ongoing data processing, which places a great deal of trust in those actually using and processing the data.

While there is an ongoing discussion on the pros and cons of BDE, regulatory frameworks have often used the concept of distinct categories and have made exceptions for all sorts of data. In Europe, the General Data Protection Regulation (GDPR)(General Data Protection Regulation, 2016) delineates health data as special category. Such sensitive health data is protected by default by both laws. The United States, with its sectoral approach has applied high safeguards for health data as long as the data is within the HIPAA sphere. Data being transferred out of the sphere of covered entities is no more subject to higher safeguards and is usually only protected by state laws, which are generally quite weak.

For decades the main harm from research was harm concerning physical integrity as, for example, side effects resulting from new medication (Rid et al., 2010). Participants mentioned that with the use of Big Data there is a shift towards intangible harms, such as breach of privacy, discrimination, and potential bias. Another angle that was mentioned by

the participants is the move from individual towards group harm, meaning that predictive algorithms are of their nature susceptible to harm arising from discrimination. In fact, if one looks at research guidelines such as the Belmont Report, harm resulting from bad use (of data) is systematically being overlooked (Raymond, 2019).

Many of the participants highlighted that what is considered as harm in the era of Big Data is culturally dependent or even motivated by individuals. This could be seen as a contradiction to regulate the use as assessing if a use should be prohibited or not is also highly belief dependent. A one size fits all approach therefore will not do the job.

b) BDE and Policy

The fundamental policy issue is whether Big Data should be treated differently from other data. It could be argued that BDE is necessary to support development of Big Data by establishing special rules more attuned to and solicitous of the needs of sophisticated predictive analytics. In other words, Big Data is a discrete and socially valuable technology that ought to be free of constraints imposed on traditional data uses and disclosures. The objection to this argument is that it should be subject to the same rules as other data acquisition, uses, and disclosures (Rothstein, 2015).

On the other hand, it could be argued that BDE is necessary because of Big Data's extraordinary potential to produce harm, even when using commonplace, but disparate data. For example, in a much-discussed incident in the United States, Target, a large chain of discount department stores, used its analytics team to see if it could discover customers' pregnancies through their purchasing patterns so they could market various products. The analytics team reviewed the shopping histories of women who signed up for the store's baby gift-registry. These women bought unscented lotion around their third month of pregnancy, and a few weeks later they often bought supplements, such as magnesium, calcium, and zinc. Using this and similar data they were able to develop accurate predictions for due dates so that the store could send appropriate coupons for each stage of pregnancy. The problem arose when one day when an angry man stormed into a Target store in Minnesota to see the manager. His daughter, a high school student, received coupons for baby clothes and cribs, and the man was greatly offended.(Kashmir, 2012)

This incident raises a variety of important questions. Initially, was there harmful conduct in the use of the data and, if so, should there be legal intervention? If yes, should the focus be on the data collection, aggregation, or dissemination? We believe that the collection

was not problematic, but the method of disseminating the data certainly was because of a lack of consent to receive such notifications and the method of notification.

The European law, although also focusing on consent and notice approaches such cases differently. Under the GDPR, information that would lead to the detection of pregnancy would be classified as health data requiring explicit consent of the user to process such data. In fact, in April 2019 the UK Information Commissioner Office (ICO) fined pregnancy and parenting advice service Bounty UK Ltd £400,000 for sharing the personal data of over 14 million individuals to a number of organizations including credit reference and marketing agencies without informing the individuals that they would do this. (This breach actually occurred before implementation of GDPR, but under the new regime this would clearly be a serious leak of health data.)

The Swiss data protection law makes a distinction between commercial and federal/cantonal entities, resulting in slight differences in the underlying legislation. For the federal/national entities the lawful ground for processing and collecting data is the presence of a permission defining the use and the need for every case. The commercial entities, however, only need a specific interest to collect and process data. Concerning the case mentioned above the Swiss law also defines health data as a special sensitive category needing explicit consent to process such data. As in the European case, processing data without any consent would have been prohibited.

Privacy laws in the United States are notoriously weak and sectoral. The collection, aggregation, indefinite storage, analysis, and possible disclosure of vast troves of sensitive personal information with presumed, but unknown, value is extremely troublesome. Informed consent notwithstanding, individuals might not realize the potential harms of such a program. For example, one of the elements of the program is return of results to participants. To benefit from any research findings, individuals might have their data uploaded into their electronic health records stored by their physician. Then, in applying for a life insurance policy, they might be required to sign an authorization disclosing all their medical records, and the research data could be used to deny coverage (Rothstein & Talbott, 2017).

The European and especially the Swiss context differ from the USA regarding large accumulations of health data. Although there has been an initiative to establish a national EHR System in Switzerland; implementation gives the patient a bundle of rights to control who

can access the data, though implementation is far from complete. A case like the one mentioned above would not be possible for basic insurance coverage because of these rights.

When it comes to health data BDE in the European and Swiss setting might on the first glimpse seem to weaken the already strong protection of those categories. But with the ever-increasing number of well-being apps and all sort of aggregation of non-health data leading to potential health revealing insights (Schneble et al., 2019a) a new notion of BDE meaning that the use is also evaluated could result in a better ethical oversight of such apps and prevent potential backlashes. In fact, the GDPR has adopted a similar approach for privacy related matters by requesting a data protection impact analysis for every “data-product” (General Data Protection Regulation, 2016). In our view, the most difficult issue involves data aggregation, because it goes to the heart of Big Data. Certain types of data aggregation, such as using AI to evaluate disparate clues in a crime investigation or to make a complicated diagnosis of disease based on unrelated symptoms, are socially desirable. Drawing distinctions between all of the good and bad applications of Big Data, however, is extremely difficult and the ethical landscape is likely to change constantly. Should individuals have a right to prevent the aggregation of their data and, if so, when would it apply? The complexity and variability of these questions might result in sectoral application of Big Data laws and a complicated regulatory scheme. For the United States, BDE could produce many of the same problematic qualities of existing privacy and data protection laws.

V. Conclusion

The results of this study illustrate the plurality of opinion regarding Big Data exceptionalism in Switzerland and the United States. Participants expressed a variety of views on how Big Data should be regulated, and conclusions regarding the acceptability of BDE depend on which definition of BDE one is using. As stated in the Introduction, there are at least three different interpretations of BDE; it can mean that use rather than consent should be regulated, that data research should be treated differently from other research, or that different laws exist for different types of data. Regulating consent rather than use of data might seem attractive in principle, but legislating to that effect alone will be insufficient in the absence of extremely robust mechanisms for review and oversight of how data is used. Given Switzerland’s strong legal emphasis on the notice and consent model, BDE is not particularly compatible with the Swiss context; the Swiss focus on data

minimization is a particular challenge. In contrast, the fragmented legislative and regulatory landscape in the USA might better lend itself to treating Big Data as an exceptional case; however, the caveat about robust oversight would apply even more strongly in this context, given precisely that specific context.

VI. References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Evans, J. P., & Burke, W. (2008). Genetic exceptionalism. Too much of a good thing? In *Genetics in Medicine* (Vol. 10, Issue 7, pp. 500–501). Nature Publishing Group. <https://doi.org/10.1097/GIM.0b013e31817f280a>
- Faghmous, J. H., & Kumar, V. (2014). A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data*, 2(3), 155–163. <https://doi.org/10.1089/big.2014.0026>
- Garrison, N. A., Brothers, K. B., Goldenberg, A. J., & Lynch, J. A. (2019). Genomic Contextualism: Shifting the Rhetoric of Genetic Exceptionalism. *The American Journal of Bioethics*, 19(1), 51–63. <https://doi.org/10.1080/15265161.2018.1544304>
- Guest, G., MacQueen, K., & Namey, E. (2014). Applied Thematic Analysis. In *Applied Thematic Analysis*. <https://doi.org/10.4135/9781483384436>
- Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, 36(5), 748–758. <https://doi.org/10.1016/j.ijinfomgt.2016.05.002>
- Hunter, D., & Evans, N. (2016). Facebook emotional contagion experiment controversy. *Research Ethics*, 12(1), 2–3. <https://doi.org/10.1177/1747016115626341>
- Kashmir, H. (2012). How target figured out a teen girl was pregnant before her father did. *Forbes*. <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics : EJHG*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Mertz, L. (2014). Saving lives and money with smarter hospitals: Streaming analytics, other new tech help to balance costs and benefits. *IEEE Pulse*, 5(6), 33–36. <https://doi.org/10.1109/MPUL.2014.2355306>
- Murray, T. H. (2019). Is Genetic Exceptionalism Past Its Sell-By Date? On Genomic Diaries, Context, and Content. In *American Journal of Bioethics* (Vol. 19, Issue 1, pp. 13–15). Routledge. <https://doi.org/10.1080/15265161.2018.1552038>
- Nissenbaum, H. (2017). Deregulating Collection: Must Privacy Give Way to Use Regulation? *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282

-
- Raymond, N. (2019). Safeguards for human studies can't cope with big data. In *Nature* (Vol. 568, Issue 7752, p. 277). NLM (Medline). <https://doi.org/10.1038/d41586-019-01164-z>
- Rid, A., Emanuel, E. J., & Wendler, D. (2010). Evaluating the risks of clinical research. In *JAMA - Journal of the American Medical Association*. <https://doi.org/10.1001/jama.2010.1414>
- Rothstein, M. A. (2015). Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics. *The Journal of Law, Medicine & Ethics*, 43(2), 425–429. <https://doi.org/10.1111/jlme.12258>
- Rothstein, M. A., & Talbott, M. K. (2017). Compelled disclosures of health records: Updated estimates. In *Journal of Law, Medicine and Ethics* (Vol. 45, Issue 1, pp. 149–155). SAGE Publications Inc. <https://doi.org/10.1177/1073110517703109>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2019a). All our data will be health data one day: the need for universal data protection and comprehensive consent. *JMIR Preprints*. <https://doi.org/10.2196/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2019b). All our data will be health data one day: the need for universal data protection and comprehensive consent (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). Google's Project Nightingale highlights the necessity of data science ethics review. *EMBO Molecular Medicine*. <https://doi.org/10.15252/emmm.202012053>
- Shaw, D. (2016). Facebook's flawed emotion experiment: Antisocial research on social network users. *Research Ethics*, 12(1), 29–34. <https://doi.org/10.1177/1747016115579535>
- Terry, N. P. (2014). Big Data Proxies and Health Privacy Exceptionalism. In *Health Matrix: The Journal of Law-Medicine* (Vol. 24).
- General Data Protection Regulation, Official Journal of the European Union (2016). https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf

Chapter 7: “Eigentum” an Patientendaten – Wer meint eigentlich was?”

Full Reference: Widmer, M., Schneble, C.O., Egli, P., Elger, B.S. Eigentum” an Patientendaten – Wer meint eigentlich was?“Eigentum” an Patientendaten – Wer meint damit eigentlich was? Pflegerecht – Pflege in Politik, Wissenschaft und Ökonomie, Issue 02, 2021

Full name(s) of author(s): Widmer, Michael^{†1}, Dr. iur, LL.M, Christophe Olivier Schneble^{†2}, MSc, Philipp, Egli¹, Dr. iur., Bernice Simone Elger^{2,3}, MD, PhD

[†]Equal Author contributionship

Full affiliation(s):

¹ZHAW School of Management and Law

²Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

³University Center for Legal Medicine, University of Geneva, Switzerland

I. Einleitung

Im Gesundheitswesen werden Daten immer wichtiger. Im Rahmen des Forschungsprojekts Big Data Research Ethics des NFP 75 Big Data wurden kantonale Datenschützer und Spitaljuristen/ Datenschützer interviewt. Ein Thema, welches in den Gesprächen aufkam, war die Diskussion um ein untechnisch gemeintes "Dateneigentum" speziell im Umfeld von Personendaten von Patienten. Dabei hat sich gezeigt, dass dieser Begriff unscharf und unklar ist und diesbezüglich diverse Missverständnisse bestehen. Der vorliegende Beitrag analysiert deshalb die Frage, was die Akteure unter einem "Eigentum an Daten" verstehen, wie diese Themen vom Recht reflektiert werden und soll für Nicht-Juristen gewisse Missverständnisse ausräumen, die in diesem Zusammenhang oft bestehen.

II. Ausgangssituation: Es gibt kein (sachen-)rechtliches Dateneigentum

Das Thema "Dateneigentum" ist Gegenstand einer intensiven und kontrovers diskutierten rechtswissenschaftlichen Diskussionen. Überblickartig lässt sich die Diskussion wie folgt zusammenfassen:¹

Das (sachen-)rechtliche Eigentum ist in der Schweiz wie folgt geregelt (Art. 641 ZGB²):

"Wer Eigentümer einer Sache ist, kann in den Schranken der Rechtsordnung über sie nach seinem Belieben verfügen."

Eigentum im rechtlichen Sinn besteht demnach grundsätzlich (nur) an Sachen. Daten als solche stellen keine Sachen dar, sondern sind immateriell: Daten sind weder räumlich fassbar noch körperlich greifbar. Während teilweise argumentiert wird, Daten seien Sachen

¹ Siehe zu diesen Themen insbesondere: Eckert, Martin, Digitale Daten als Wirtschaftsgut: Besitz und Eigentum an digitalen Daten, SJZ 112 (2016), 265 ff. (zit. "Eckert, Besitz und Eigentum"); Eckert, Martin, Digitale Daten als Wirtschaftsgut: digitale Daten als Sache, SJZ 112 (2016) 245 ff. (zit. "Eckert, Digitale Daten"); Fröhlich-Bleuler, Gianni, Eigentum an Daten? In: Jusletter 6. März 2017, https://jusletter.weblaw.ch/juslissues/2017/883/eigentum-an-daten-_8ba966db21.html__ONCE; Früh, Alfred / Lombard, Alexandre / Thouvenin, Florent (Hrsg.), Eigentum an Sachdaten: Eine Standortbestimmung, SZW (2017), 25 ff.; Gordon, Clara-Ann, Personal Information Management Systems (PIMS), in: Jusletter IT 11. Dezember 2017, https://jusletter-it.weblaw.ch/flash/flash/11-dezember-2017/personal-information_b00b0b7ab1.html; Hürlimann, Daniel / Zech, Herbert, Rechte an Daten, sui-generis (2016), 89 ff.; Picht, Peter Georg, Dateneigentum und Datenzugang, Schutz von Geschäftsgeheimnissen als Alternative, in: Jusletter IT 11. Dezember 2017, https://jusletter-it.weblaw.ch/flash/flash/11-dezember-2017/dateneigentum-und-da_06cb18f419.htm; Schmid, Alain / Schmidt, Kirsten Johanna / Zech, Herbert, Rechte an Daten – zum Stand der Diskussion, sic! (2018), 627 ff.; Thouvenin, Florent, Wem gehören meine Daten? Zu Sinn und Nutzen einer Erweiterung des Eigentumsbegriffs, SJZ 113 (2017), 21 ff.; Weber, Rolf H. / Thouvenin, Florent, Dateneigentum und Datenzugangsrechte – Bausteine der Informationsgesellschaft?, ZSR 137 (2018) 43 ff.

² Schweizerisches Zivilgesetzbuch vom 10. Dezember 1907.

gleichzustellen³, vermögen die diesbezüglichen Argumente nach geltendem Recht nicht zu überzeugen, da der Begriff der "Sache" zwar im Gesetz nicht definiert, in der Lehre aber schon lange als "körperlicher (...) Gegenstand" bezeichnet wird.⁴ Würde der Begriff der "Sache" um Daten erweitert, verlöre er jegliche Konturen.

Es ist deshalb davon auszugehen, dass ein "Eigentum" an Daten im eigentlichen (sachen-) rechtlichen Sinne nicht existiert.⁵ Aus diesem Grund ist der Ausdruck "Dateneigentum" ungenau und könnte sogar als unzutreffend bezeichnet werden.⁶

Dennoch scheint die Tatsache, dass es kein (sachen-)rechtliches Eigentum an Daten gibt, in weiten Kreisen noch immer unbekannt. Nach wie vor wird über "Dateneigentum" diskutiert und der Begriff verwendet, um verschiedene völlig unterschiedliche Konzepte und Ideen auszudrücken.

In den nachfolgenden Ausführungen wird, basierend auf den Ergebnissen der durchgeführten Experteninterviews, aufgezeigt, welche Akteure aus ihrer jeweiligen Perspektive mit dem Ausdruck "Dateneigentum" welche rechtlichen Aspekte bezeichnen. Zugleich wird jeweils auf die rechtlichen Instrumente verwiesen, die unter geltendem Recht zur Verfügung stehen und dem untechnisch verstandenen "Dateneigentum" am nächsten kommen.⁷

³ Eckert, Besitz und Eigentum, 265; Eckert, Digitale Daten, 245.

⁴ Schmid / Schmidt / Zech, 630; Honsell, Heinrich / Vogt, Nedim Peter / Geiser, Thomas, Basler Kommentar ZGB II, Basel 2015, ZGB vor Art. 641 ff. N 6.

⁵ Siehe dazu u.a. Schmid / Schmidt / Zech, 630.

⁶ In der juristischen Literatur wird ferner darüber diskutiert, ob (und in welcher Form) allenfalls ein Dateneigentum als eigenes Rechtsinstrument eingeführt werden soll. WEBER/THOUVENIN gelangten indessen in einer ausführlichen Analyse zu Recht zum Schluss, dass keine ausreichenden Gründe für die Einführung einer neuen Rechtsfigur "Dateneigentum" vorliegen. Thouvenin / Früh / Lombard, 31 ff.; Weber/Thouvenin, 64. Sie gelangen auch zum Schluss, dass in diesem Bereich zwar gewisse Regelungslücken bestehen. Diese sollen indessen durch eine punktuelle Anpassung der Rechtsordnung in einzelnen Bereichen gefüllt werden (bspw. im Schuldbetreibungs- und Konkursrecht).

⁷ Darüber hinaus wird in den oben Beiträgen erläutert, mit welchen rechtlichen Instrumenten verschiedene Aspekte eines "Dateneigentums" unter geltendem Recht verwirklicht werden können und welche Änderungen des Rechts möglich oder notwendig wären, um ein dem "Eigentum" ähnliches Recht an Daten einzuführen. Neben dem Dateneigentum wird in diversen Beiträgen als weiterer Aspekt jeweils das "Datenzugangsrecht" erwähnt. Das Thema des "Datenzugangsrechts" wird im vorliegenden Artikel jedoch nicht näher beleuchtet, weil es in den Interviews nicht zur Sprache kam. Zu diesem Thema siehe bspw. Weber/Thouvenin, 43 ff.; Schmid / Schmidt / Zech, 634 ff.

III. "Dateneigentum" aus unterschiedlichen Perspektiven

a) Zu unterscheidende Perspektiven

Wenn in den durchgeführten Interviews⁸ die Befragten jeweils von "Dateneigentum" sprachen, waren damit offenbar jeweils unterschiedliche Aspekte gemeint. Der Begriff wurde zunächst einmal schon von den Befragten je nach Perspektive unterschiedlich aufgefasst. Ausserdem berichteten die Experten auch über ihre Beobachtungen, wie kontrovers verschiedene Akteure Dateneigentum interpretieren. Es können dabei im Wesentlichen die folgenden drei Blickwinkel unterschieden werden:

Forschende:

Aus der Sicht (privater oder öffentlicher) Institutionen bzw. Unternehmen, die mit Daten forschen wollen, kann sich u.a. die Frage stellen, wie sie eine gewisse "Verfügungshoheit" über Daten behalten resp. sich an kommerziell verwertbaren Forschungsergebnissen beteiligen können – unabhängig davon, ob es sich um Personendaten oder um nicht personenbezogene Daten, so genannte Sachdaten, handelt (siehe nachfolgend Ziff. 4.2). Hierbei kann zwischen einer positiven Seite und einer negativen Seite unterschieden werden. Die positive Seite betrifft die umfassende "Herrschaft", also bspw. Gebrauch, Verwertung und Übertragung der Daten; die negative Seite das Recht, die Daten von Dritten heraus zu verlangen und Einwirkungen von Dritten abzuwehren.⁹

Betroffene Personen:

Aus Sicht der betroffenen Person, über welche Daten bearbeitet werden (Art. 3 lit. b DSGVO¹⁰), wird die Bezeichnung "Dateneigentum" oft mit der Frage in Verbindung gebracht, wie weit die betroffene Person über die Verwendung ihrer Daten bestimmen kann und/oder inwieweit es ihr möglich sein soll, daran (wirtschaftlich oder anderweitig) zu partizipieren (siehe nachfolgend Ziff. V).

Leistungserbringer:

Aus Sicht eines Spitals oder eines Arztes stellen sich schliesslich Fragen wie: wer (ein Patient als betroffene Person, ein Spital oder der Arzt) hat woraus welche Rechte an den

⁸ Die Interviews betrafen im Wesentlichen Themen der medizinischen Forschung und der Tätigkeit von Ärzten und Spitalern.

⁹ Siehe hierzu Thouvenin, 26 f.

¹⁰ Bundesgesetz über den Datenschutz vom 19. Juni 1992.

vom Arzt bearbeiteten Personendaten, welche in einer Krankengeschichte oder Ähnlichem gesammelt werden (siehe nachfolgend Ziff. VI)?

Daneben gibt es noch weitere Perspektiven, z.B. diejenige eines Kranken- oder Unfallversicherers, worauf aber an dieser Stelle nicht näher eingegangen wird.

IV. Sicht von Forschenden

a) Ausgangslage

Abb. 1 enthält Aussagen aus geführten Interviews, welche den Blick von forschenden Unternehmen und Institutionen auf das Thema "Dateneigentum" beleuchten. Dabei stellen sich u.a. die Fragen, (a) wie Daten verwertet werden können ("positive" Wirkung des "Eigentums"¹¹), die im Rahmen von Forschungsvorhaben bearbeitet werden oder wenn aus Primärdaten etwas Neues entsteht; und (b) wie diese Daten vor Einwirkungen von Dritten geschützt werden können ("negative" Wirkung des "Eigentums").

Da das Schweizer Recht wie eingangs erwähnt kein (sachen-)rechtliches Eigentum an Daten kennt, stellt sich die Frage, wie die aus dieser Perspektive mit dem Begriff "Dateneigentum" verbundenen Verwendungs- und Verwertungsinteressen (kommerzielle Interessen) rechtlich geschützt werden können. Dazu gibt es folgende Lösungsansätze.

b) Eigentum am Datenträger?

Im Gegensatz zu Daten stellt *ein konkreter Datenträger* (bspw. Server oder anderes Medium), auf welchem Daten gespeichert werden, eine Sache im rechtlichen Sinn dar (Art. 641 ZGB): Der Datenträger ist greifbar. An solchen Datenträgern kann (sachenrechtliches) Eigentum bestehen. Es ist demnach aus sachenrechtlicher Sicht zu unterscheiden zwischen (a) dem *Speichermedium*, an welchem *Eigentum* bestehen kann, und (b) den darauf *gespeicherten Informationen (Daten)*, welche nicht Sachen im rechtlichen Sinne darstellen und an denen demnach *kein Eigentum* besteht.

Wird diese Unterscheidung bewusst eingesetzt, kann das Eigentum am Datenträger für eine Institution bereits einen gewissen Schutz ihrer Interessen ermöglichen, indem bspw. in einem Konkursverfahren unter gewissen Voraussetzungen ein Datenträger herausverlangt werden kann. Bei der Durchsetzung des Eigentums am Datenträger besteht aber nach wie vor das Problem, dass sich dieses nicht auf die gespeicherten Informationen bezieht.

¹¹ Siehe dazu Abb. 1, CHLR 8.

Deshalb kann bspw. ein Herausgabeanspruch des Eigentümers gegenüber einem Dritten von diesem auch durch Herausgabe des leeren (gelöschten) Datenträgers erfüllt werden. Der Dritte könnte auch nach Belieben die Daten kopieren und den Datenträger anschliessend herausgeben.

c) **Immaterialgüterrechtlicher Schutz?**

Weil Daten immaterielle Güter sind, ist neben der Möglichkeit des sachenrechtlichen Eigentums zu prüfen, wie weit Daten allenfalls durch *Immaterialgüterrechte* geschützt werden können.¹²

Ein immaterialgüterrechtlicher Schutz für "Daten" ganz allgemein besteht in der Schweiz nicht. Im Einzelfall kann es indessen sein, dass Daten rechtlich als "geistige Schöpfungen mit individuellem Charakter" eingestuft werden und damit *urheberrechtlichen Schutz* geniessen.¹³

Gemäss Art. 2 Abs. 2 URG¹⁴ stellen "geistige Schöpfungen der Literatur und Kunst, die individuellen Charakter haben" unabhängig von ihrem Wert oder Zweck urheberrechtlich geschützte Werke dar. Als "Werke" geschützt sind ferner *Sammlungen*, "sofern es sich bezüglich Auswahl oder Anordnung um geistige Schöpfungen mit individuellem Charakter handelt" (Art. 4 Abs. 1 URG).

Im Zusammenhang mit Daten und Zusammenstellungen resp. Sammlungen von Daten ist deshalb zuerst zu fragen, ob ein gewisses "Datum" urheberrechtlich geschützt ist. Dabei ist die Schwelle zum Urheberrechtsschutz nicht besonders hoch. Voraussetzung ist aber immerhin, dass es sich um eine "geistige Schöpfung mit individuellem Charakter" handeln muss. Reine Ideen oder Informationen sind demnach nicht urheberrechtsgeschützt, ebenso wenig blosse Tatsachen, wie Labor- und Forschungsdaten oder Messergebnisse.¹⁵ Aus diesem Grund mag zwar im Einzelfall das Urheberrecht manchmal greifen. Oftmals fallen

¹² Neben der Möglichkeit, die Rechtsfigur eines "Dateneigentums" als sachenrechtliche Eigentum anzulehnen, wird auch die Einführung eines Immaterialgüterrechts "sui generis" diskutiert; siehe dazu neben anderen Härting, Niko, "Dateneigentum" - Schutz durch Immaterialgüterrecht? Was sich aus dem Verständnis von Software für den zivilrechtlichen Umgang mit Daten gewinnen lässt, CR 2016 646–649.

¹³ Auf die Frage eines patentrechtlichen Schutzes wird vorliegend nicht näher eingegangen. Ein solcher dürfte in der Regel daran scheitern, dass Daten meist die Voraussetzungen für patentrechtlichen Schutz (technischer Charakter der Erfindung; Erfindungshöhe; technischer Fortschritt; Neuheit; und gewerbliche Anwendbarkeit) nicht erfüllen.

¹⁴ Bundesgesetz über das Urheberrecht und verwandte Schutzrechte vom 9. Oktober 1992.

¹⁵ BGE 113 II 306; Fröhlich-Bleuler, Rz. 10, Cherpillod, Ivan, in: Müller, Barbara K. / Oertli, Reinhard (Hrsg.), Handkommentar, Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 2. A., Zürich 2012, N 44 zu Art. 2 URG.

Daten aber durch die "Maschen" des Urheberrechts und geniessen keinen Schutz – entweder weil sie keine "geistige Schöpfung" darstellen oder nicht "individuellen Charakter" haben.

In einer zweiten Stufe stellt sich sodann die Frage, ob eine Sammlung von Daten bzw. Zusammenstellung von Daten, wie sie bspw. bei Ergebnissen von Forschungsprojekten oftmals vorliegen, geschützt sind. Für solche Sammelwerke gilt dieselbe Schwelle wie für einzelne Werke. Das heisst, die Auswahl oder Anordnung der Daten muss ebenfalls das Erfordernis einer *"geistigen Schöpfung mit individuellem Charakter"* erfüllen (z.B. aufgrund ihrer Struktur oder Darstellungsform).¹⁶ Oftmals wird diese Schwelle des Urheberrechts von solchen Zusammenstellungen von Daten nicht überschritten, womit wiederum ein entsprechender Schutz fehlt.

Soweit im Einzelfall tatsächlich ein Werk oder eine als Werk geschützte Sammlung vorliegen sollte, hat der Urheber u.a. *das ausschliessliche Recht am eigenen Werk* und damit das ausschliessliche Recht zu bestimmen, ob, wann und wie das Werk verwendet wird (Art. 9 Abs. 1 und Art. 10 Abs. 1 URG). Deshalb könnte der Urheber einer im konkreten Einzelfall geschützten Datensammlung (bspw. von Forschungsdaten) Dritten verbieten, diese (gesamte) Sammlung ohne seine Zustimmung zu gebrauchen. Das wäre auch dann möglich, wenn zwischen dem Urheber und dem Dritten kein Vertrag besteht. Es wäre eine Verletzung des Urheberrechts, wenn man Exemplare einer solchen Datensammlung vervielfältigen, anbieten oder sie Dritten zugänglich machen würde. Dem Urheber wäre es sodann möglich, seine Rechte durch Lizenzierung (oder auch Verkauf) der entsprechenden Urheberrechte zu kommerzialisieren.

Allerdings erstreckt sich der Schutz eines Sammelwerks durch das Urheberrecht lediglich auf die Struktur der Sammlung (Auswahl und Anordnung), *nicht jedoch* auf die einzelnen Daten. Die Verwendung oder Verwertung *einzelner Daten* aus einer solchen Sammlung wäre damit urheberrechtlich zulässig. Ein möglicherweise im Einzelfall bestehender Schutz ist für die Institution deshalb nicht wirklich befriedigend und wenig sinnvoll.

Zusammenfassend lässt sich der Zweck, welcher von Institutionen bei Verwendung des Begriffs "Dateneigentum" offenbar (mit)gemeint ist, mit dem Mittel des Urheberrechts

¹⁶ Egloff, Willi, in: Egloff, Willi / Barrelet, Denis (Hrsg.), Das neue Urheberrecht, Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. A., Bern 2020, N 5 zu Art. 4 URG.

kaum erreichen. Denn die Verwendungs- und Verwertungsinteressen sind über das Urheberrecht nur teilweise geschützt.

d) Schutz von Datenbanken?

Wie sich aus diesen Ausführungen ergibt, ist eine Datenbank aus urheberrechtlicher Sicht nur dann gegen eine Übernahme der darin enthaltenen Daten durch Dritte geschützt, wenn sie die oben dargelegten Voraussetzungen erfüllt.¹⁷ Der urheberrechtliche Schutz erstreckt sich – wenn überhaupt – also auch in diesem Fall nur auf die Struktur einer Datenbank und in der Regel nicht auf deren Inhalt.

Der in der EU bestehende so genannte "Sui-generis"-Datenbankschutz¹⁸ soll hingegen Investitionen in Datenbanken schützen. Er besteht zwar unabhängig davon, ob der Inhalt einer Datenbank urheberrechtlichen Schutz genießt, erstreckt sich aber ebenfalls nur gegen die Entnahme und Weiterverwendung *wesentlicher Teile* der Datenbank.¹⁹ Auch dieser Datenbankschutz erstreckt sich somit nicht auf einzelne Daten.

Im Schweizer Recht besteht ohnehin keine Regelung, welche einen den EU-Regelungen entsprechenden Schutz von Datenbanken – unabhängig vom Urheberrechtsschutz – gewährt.²⁰

e) Schutz gemäss UWG und Schutz von Fabrikations- und Geschäftsgeheimnissen?

Das Bundesgesetz gegen den Unlauteren Wettbewerb könnte in Einzelfällen ebenfalls einen gewissen Schutz bieten. So ist es, um ein Beispiel herauszugreifen, unlauter, "das marktreife Arbeitsergebnis eines andern ohne angemessenen eigenen Aufwand durch technische Reproduktionsverfahren als solches" zu übernehmen und zu verwerten (Art. 5 lit. c UWG²¹). Allerdings wird hier keine Eigentumsposition geschaffen, sondern einzig normiert, dass ein gewisses Verhalten widerrechtlich ist.

¹⁷ Egloff, Willi, in: Egloff, Willi / Barrelet, Denis (Hrsg.), Das neue Urheberrecht, Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. A., Bern 2020, N 6 zu Art. 4 URG m.w.H.

¹⁸ <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:DE:HTML>, zuletzt besucht am: 20.06.2020.

¹⁹ https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_de.htm, zuletzt besucht am: 20.06.2020.

²⁰ Cherpillod, Ivan, in: Müller, Barbara K. / Oertli, Reinhard (Hrsg.), Handkommentar, Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 2. A., Zürich 2012, N 6 zu Art. 4 URG.

²¹ Bundesgesetz gegen den unlauteren Wettbewerb vom 19. Dezember 1986.

Darüber hinaus dürften im Einzelfall oftmals auch die Voraussetzungen von Art. 5 lit. c UWG gar nicht erfüllt sein. So ist neben den bereits genannten Punkten zu berücksichtigen, ob der Ersteller die Möglichkeit hatte, seinen Aufwand zu amortisieren. Wenn das der Fall ist, besteht in der Regel kein Schutz nach Art. 5 lit. c UWG (mehr). Demnach bietet auch diese Bestimmung des UWG für denjenigen, welcher einen rechtlichen Schutz für seine "Datenhoheit" sucht, meist keine verlässliche Rechtsgrundlage.

Ferner könnte Art. 6 UWG zur Anwendung gelangen, welcher die unrechtmässige Verwertung von Fabrikations- und Geschäftsgeheimnissen verbietet. Dies würde allerdings voraussetzen, dass es sich im konkreten Fall um Geheimnisse handelt, und dass diese "ausgekundschaftet oder sonst wie unrechtmässig erfahren" wurden. In solchen Fällen ist eine Verwertung der entsprechenden Daten verboten. Auch diese Regelung wird wohl nur in "extremen" Fällen zur Anwendung gelangen.

f) Strafrechtlicher Schutz?

Auch das schweizerische Strafrecht lässt sich zu einem gewissen Grad dazu verwenden, die gewünschte "Datenhoheit" zu verteidigen. So ist die Verletzung der beiden soeben genannten Bestimmungen des UWG auf Antrag strafbar (Art. 23 UWG).

Darüber hinaus sind Fabrikations- und Geschäftsgeheimnisse durch die Regelung in Art. 162 StGB²² strafrechtlich geschützt. Gemäss dieser Norm ist es strafbar, solche Geheimnisse für sich auszunützen, wenn sie verraten wurden, obschon eine gesetzliche oder vertragliche Pflicht bestanden hätte, sie zu bewahren. Das Strafrecht enthält noch diverse weitere Normen, welche zwar einen gewissen Schutz von Daten bezwecken,²³ aber ebenfalls keine "Eigentumsposition" zu begründen vermögen.

Schlussendlich wäre es im vorliegenden Zusammenhang für denjenigen, welcher seine "Datenhoheit" behaupten will, ohnehin unbefriedigend, auf solche allgemeinen strafrechtlichen Bestimmungen zurück geworfen zu werden, die in der Regel nur in Ausnahmefällen greifen würden und langwierige Strafverfahren voraussetzen. Immerhin kann ein drohendes Strafverfahren eine gewisse abschreckende Wirkung haben.

²² Schweizerisches Strafgesetzbuch vom 21. Dezember 1937.

²³ Als Beispiele seien hier genannt: Unbefugte Datenbeschaffung (Art. 143 StGB); Unbefugtes Eindringen in eine Datenverarbeitungsanlage (Art. 143^{bis} StGB); Betrügerischer Missbrauch einer Datenverarbeitungsanlage (Art. 147 StGB).

g) Schutz durch Vertrag?

Aus den angeführten Gründen bleibt den Betroffenen in der Regel nichts anderes übrig als zu versuchen, die Hoheit über die Daten vertraglich zu schützen. Der erhebliche Nachteil dieses Vorgehens ist, dass die entsprechenden Vertragsbestimmungen nur gegenüber dem *konkreten Vertragspartner* wirken und – wenn überhaupt – auch nur gegenüber diesem durchgesetzt werden können. Allerdings sind solche Regelungen besser als nichts und können eine Durchsetzung der Rechte nach UWG oder Art. 162 StGB unterstützen. Dies insbesondere, wenn dem Vertragspartner eine vertragliche Pflicht zur Geheimhaltung auferlegt wird.

Bei der Gestaltung der entsprechenden Verträge ist insbesondere zu überlegen, welche Punkte mit dem Vertragspartner geregelt werden müssen, um eine zumindest "eigentumsähnliche" Position zu erlangen. Zentral können in dieser Hinsicht und je nach Art der Zusammenarbeit u.a. folgende Themen sein – jeweils faktisch zusätzlich durch technologische Massnahmen abgesichert:

- (a) Geheimhaltungspflicht;
- (b) Verfügungs- und Zugangsrecht für den "Inhaber";
- (c) Verwertung/Bearbeitung regeln, insbesondere Nutzungsrecht durch den "Inhaber" und Nutzungsbeschränkungen für den Vertragspartner;
- (d) Möglichkeit zur Löschung durch den Inhaber.

h) Faktische Kontrolle

Neben einer rechtlichen Zuordnung von Daten ist auch eine faktische Zuordnung möglich. Eine solche erfolgt, so lange eine Institution durch technische und organisatorische Massnahmen *rein faktisch* die Kontrolle über Daten hat und Dritte vom Zugang zu den Daten ausschliessen kann.

Weil der rechtliche Schutz von Daten nicht umfassend ist (bspw. nicht so ausgeprägt wie sachenrechtliches Eigentum), erscheint es für Institutionen umso wichtiger, gewisse Daten faktisch zu kontrollieren, also die "faktische Verfügungsgewalt" zu bewahren und durch technische und organisatorische Massnahmen zu verhindern, dass Daten an Dritte gelangen, besonders wenn diese Dritten nicht vertraglich gebunden sind.

V. Sicht der betroffenen Person

a) Ausgangslage

Eine zweite Perspektive ist diejenige der betroffenen Personen. Diese wird gut ersichtlich aus der Formulierung in der Aussage: "*Diese gehören nicht mehr dem Patienten.*"²⁴

b) Datenschutz

Begriffe und Anwendungsbereich

Betrachtet man den Begriff "Dateneigentum" aus Sicht einer betroffenen Person, über die Daten bearbeitet werden, bezieht er sich oftmals auf die *Persönlichkeitsrechte* der betroffenen Person. Die Schweizer Bundesverfassung kennt ein Grundrecht auf Datenschutz (Art. 13 Abs. 2 BV). Gemäss Bundesgericht garantiert dieses Grundrecht, "dass grundsätzlich ohne Rücksicht darauf, wie sensibel die fraglichen Informationen tatsächlich sind, jede Person gegenüber fremder, staatlicher oder privater Bearbeitung von sie betreffenden Informationen bestimmen können muss, ob und zu welchem Zweck diese Informationen über sie bearbeitet werden".²⁵ Dieses sog. *Recht auf informationelle Selbstbestimmung* ist sehr weit gefasst und gibt der betroffenen Person eine Art Verfügungsrecht über "ihre" Daten. Ein absoluter Schutz vermag dieses Grundrecht aber nicht zu liefern. Gesetzliche Einschränkungen des Grundrechts auf Datenschutz sind rechtlich möglich.

Diese Persönlichkeitsrechte werden durch die Bestimmungen des Datenschutzrechts konkretisiert. Durch diese Normen werden indessen nicht die "Daten" an sich geschützt, sondern im Wesentlichen die Persönlichkeitsrechte der betroffenen Personen, über welche Daten bearbeitet werden.²⁶

Das Datenschutzrecht kommt nur dann zur Anwendung, wenn eine Bearbeitung von *Personendaten* in Frage steht. Als Personendaten gelten alle Angaben, die sich auf eine bestimmte oder bestimmbare natürliche oder juristische Person beziehen. Besteht gar kein Personenbezug (bspw. bei Sachdaten) oder wird ein bestehender Personenbezug definitiv entfernt (bspw. im Rahmen einer Anonymisierung), liegen keine Personendaten mehr vor und das Datenschutzrecht ist nicht anwendbar.²⁷

²⁴ Siehe dazu Abb. 1, CHLR 17.

²⁵ Statt vieler BGE 148 I 11 E. 3.1.1. S. 13.

²⁶ Schmid / Schmidt / Zech, 635.

²⁷ Detailliertere Ausführungen zum Begriff der Personendaten finden sich bspw. bei Rosenthal, David / Jöhri, Yvonne (Hrsg.), Handkommentar zum Datenschutzgesetz, N 1 ff. zu Art. 3 DSG.

Dieses Verständnis kommt in der Aussage des Teilnehmer CHLR11 zum Ausdruck, in welcher ausgeführt wird, dass *weil es sich um anonymisierte Daten handle, diese nicht mehr dem Patienten gehören*. Hier wird also das "Dateneigentum" verknüpft mit der Frage, ob ein Personendatum vorliegt und deshalb das DSG anwendbar ist.

“Für mich gehören sie nicht dem Patienten, weil es sich genau um anonymisierte Daten handelt, Es ist klar, dass wir danach wirklich sicherstellen müssen, dass die übertragenen Daten es uns nicht ermöglichen, die Verbindung zum Patienten herzustellen.” (CHLR11)

Im Zusammenhang mit dem Datenschutzrecht wird ferner der Begriff des "*Bearbeitens*" äusserst breit verstanden. Nicht nur das Beschaffen, Verwenden, Umarbeiten, Bekanntgeben und Archivieren, sondern bereits das Aufbewahren und auch die Vernichtung von Personendaten gelten als Bearbeiten in diesem Sinne (Art. 3 lit. e DSG).

Hinzuweisen ist auf das Bestehen von teilweise *unterschiedlichen* datenschutzrechtlichen Vorschriften bezüglich der Datenbearbeitung durch private Personen, derjenigen durch Bundesorgane (bspw. ETH) und durch kantonale Organe (bspw. Kantonsspitäler oder Universitäten).²⁸

Widerrechtliche Persönlichkeitsverletzungen und Rechtfertigungsgründe

Werden Personendaten von einer privaten Person bearbeitet (bspw. einer privaten Forschungsinstitution), so ist diese Bearbeitung gemäss Art. 12 Abs. 1 DSG nur dann verboten, wenn dabei die Persönlichkeit der betroffenen Person widerrechtlich verletzt wird. Eine Persönlichkeitsverletzung kann insbesondere vorliegen, wenn (a) Personendaten entgegen gewisser im DSG festgelegter Grundsätze bearbeitet werden; (b) Daten einer Person gegen deren ausdrücklichen Willen bearbeitet werden; oder (c) besonders schützenswerte Personendaten oder Persönlichkeitsprofile Dritten bekannt gegeben werden. Sogar in diesen Fällen ist die Bearbeitung aber erlaubt, wenn ein Rechtfertigungsgrund vorliegt (Art. 12 Abs. 2 DSG). Die Rechtfertigung kann sich aus einer Einwilligung der betroffenen Person, einem überwiegenden (privaten oder öffentlichen) Interesse oder einer Gesetzesbestimmung ergeben.

Daraus ergibt sich, dass nicht für jede Bearbeitung eine Einwilligung der betroffenen Person benötigt wird. In denjenigen Fällen, in denen eine solche Einwilligung erforderlich ist

²⁸ Auf die Bearbeitung von Personendaten durch private Personen oder Bundesorgane ist das Bundesgesetz über den Datenschutz ("DSG") anwendbar, auf die Bearbeitung durch kantonale Organe grundsätzlich das kantonale Datenschutzrecht. Die nachfolgenden Ausführungen beziehen sich im Wesentlichen auf das DSG und die Bearbeitung durch Private.

(bspw. weil Gesundheitsdaten an Dritte bekannt gegeben werden, ohne dass dafür ein anderer Rechtfertigungsgrund vorliegt), wird der betroffenen Person zumindest theoretisch die Möglichkeit eröffnet, im Zusammenhang mit der Einwilligung einen Vertrag zu schliessen. In diesem könnte sie versuchen, sich ebenfalls einen (kommerziellen oder anderen) Nutzen aus der Bearbeitung zu sichern. Allerdings ist eine solche Einwilligung jederzeit frei widerruflich, weshalb sie für die Bearbeiter keine Sicherheit zu schaffen vermag.

Diese jederzeitige freie Widerrufbarkeit der Einwilligung räumt der betroffenen Person zwar Rechte ein und stärkt ihre Position. Aus der Perspektive der Forschenden und anderer Beteiligter erzeugt aber das Erteilen einer Einwilligung zur Bearbeitung von Personendaten deshalb deutlich weniger Sicherheit als bspw. die Übertragung sachenrechtlichen Eigentums.²⁹ Deshalb steht die freie Widerrufbarkeit der Einwilligung letztlich der konkreten Verwirklichung eines eigentumsähnlichen Rechtsinstituts entgegen.³⁰

Rechte und Pflichten aus dem Datenschutzrecht

Das DSG auferlegt dem Bearbeiter diverse Pflichten, welche er bei der Bearbeitung zu beachten hat. So dürfen beispielsweise Datenbearbeitungen grundsätzlich nur im Einklang mit den allgemeinen Datenschutzgrundsätzen erfolgen und beim Beschaffen von Personendaten durch Bundesorgane sowie bei der Beschaffung von besonders schützenswerten Personendaten (darunter fallen bspw. Gesundheitsdaten) oder Persönlichkeitsprofilen durch Private bestehen spezifische Informationspflichten gegenüber den betroffenen Personen (Art. 14 und Art. 18a DSG). Auch aus weiteren der betroffenen Person zustehenden Rechten ergeben sich im Umkehrschluss für die Bearbeiter weitere Pflichten, auf welche hier nicht näher einzugehen ist.

Die Ansprüche der betroffenen Person gegen eine widerrechtliche Datenbearbeitungen durch Private ergeben sich im Wesentlichen aus der Regelung in Art. 15 Abs. 1 DSG. Diese sieht u.a. die Sperrung, Berichtigung oder Löschung vor sowie die Möglichkeit von Klagen auf, Schadenersatz und Genugtuung oder Gewinnherausgabe³¹.

²⁹ Eine solche Übertragung könnte grundsätzlich nicht mehr widerrufen werden.

³⁰ Schmid / Schmidt / Zech, 636 m.w.H.

³¹ Klagen zum Schutz der Persönlichkeit richten sich gemäss Art. 15 Abs. 1 nach den Artikeln 28, 28a sowie 28l des Zivilgesetzbuches.

Ferner hat die betroffene Person das Recht, über die bearbeiteten Daten Auskunft zu verlangen. Dies gibt ihr die Möglichkeit, vom Datenbearbeiter Kopien der sie betreffenden Daten zu erhalten.³²

Ausserdem wird der betroffenen Person das Recht verliehen, bei der Bearbeitung durch Private den Willen zum Ausdruck zu bringen, dass ihre Personendaten nicht bearbeitet werden sollen. Ein solcher Widerspruch hätte zur Folge, dass der private Datenbearbeiter einen Rechtfertigungsgrund benötigt, sofern er die Personendaten trotzdem weiterhin bearbeiten will (Art. 12 Abs. 2 lit. b DSGVO). Dieses Verständnis wird in der Aussage des Teilnehmer CHLR19 deutlich, in welcher gesagt wurde, dass der Patient natürlich immer das Recht habe, die Verwendung seiner Daten zu widerrufen. Er habe immer gewisse Rechte, auf die er einwirken könne.

“Der Patient hat natürlich immer das Recht die Verwendung von seinen Daten zu widerrufen, Einsicht zu nehmen was gemacht wird. Also er hat ja gewisse Rechte, die er... ich meine auf die er Einwirken kann.” (CHLR19)

Folgerung

Die obigen Ausführungen zeigen summarisch auf, dass sich aus dem Datenschutzrecht zwar im Einzelfall eine Möglichkeit der betroffenen Person ergibt, auf die Bearbeitungen ihrer Daten einzuwirken, bspw. in gewissen Fällen eine konkrete Datenbearbeitung zu verbieten. Es ist ihr aber kaum möglich, daraus tatsächlich eine Möglichkeit zur Teilhabe an der Kommerzialisierung ihrer Daten abzuleiten.

Demnach wird auch durch das Datenschutzrecht für die betroffene Person weder ein "Dateneigentum" noch ein zumindest eigentumsähnliches Recht geschaffen.³³

VI. Sicht des Arztes oder des Spitals

a) Ausgangslage

³² Thouvenin sieht damit den ersten Aspekt der "negativen" Seite des Eigentums verwirklicht, nämlich das Recht, die Sache von Dritten herauszuverlangen; Thouvenin 27. In dieser Hinsicht liesse sich allerdings argumentieren, dass beim (sachen)rechtlichen Eigentum dieser Aspekt auch zur Folge hat, dass der Dritte die Sache nicht mehr im Besitz hat. Bei Daten ist dies indessen nicht zwingend der Fall. So kann der Bearbeiter eine Kopie der Daten herausgeben, aber selbst ebenfalls einen Satz der Daten behalten.

³³ Thouvenin, 26, welcher darauf hinweist, dass Datenschutzrecht zwar die negative Seite des "Eigentums" (Abwehr der Einwirkung von Dritten) teilweise verwirklichen mag, nicht jedoch die positive Seite (bspw. Gebrauch, Verwertung und Übertragung).

Wenn hingegen im Verhältnis zwischen einem Arzt oder einem Spital und den Patienten von "Dateneigentum" die Rede ist, so spielen unterschiedliche Aspekte mit.

Gewisse in diesen Verhältnissen betreffend "Dateneigentum" interessierende Fragen ergeben sich aus der folgenden Antwort auf die Frage, wem die Daten gehören:

"Der Patient sagt ihm, und wir sagen natürlich uns. Ich glaube aber es ist eine geteilte, ein geteiltes Eigentum. Erschaffen tun wir die Daten, wir sind die Urheber der Daten. Weil unsere Ärzte schreiben die Berichte, die machen die Befundung. Es ist aber natürlich auch ein Auftrag, den wir für den Patienten machen. In der Regel ist das Auftragsergebnis gehört dem Auftraggeber." (CHLR8)

b) Dokumentation und Auftragsrecht

Auch in diesem Zusammenhang sind Fragen der Rechte von Institutionen und des Datenschutzes der betroffenen Person angesprochen (siehe dazu oben Ziff.IV und V).

Darüber hinaus ist das Verhältnis zwischen Arzt und Patient zu beachten. Es ist vornehmlich auftragsrechtlicher Natur, während im Verhältnis Patient – Spital öffentlich-rechtliche, oftmals kantonale Vorschriften zur Anwendung gelangen.

Soweit im Verhältnis zwischen Arzt und Patient Auftragsrecht anwendbar ist, hat der Arzt eine primäre Pflicht zur Erhebung von Patientendaten sowie eine vertragliche Nebenpflicht zur Dokumentation. Das (sachenrechtliche) Eigentum an der erstellten Dokumentation liegt dabei zunächst grundsätzlich beim Arzt (wobei dem Patienten die oben unter Ziff. V erwähnten datenschutzrechtlichen Ansprüche zukommen). Der Arzt hat indessen aufgrund seiner Pflicht zur Rechenschaftsablegung eine "Ablieferungspflicht"; das bedeutet, dass er u.a. das in Auftrags erledigung Erhaltene (bspw. Röntgenbilder) und Geschaffene (Berichte, Telefonaufzeichnungen) sowie Erlangte (bspw. Laborberichte) dem Patienten abzuliefern hat. Für gewisse Dokumente gilt aber, dass diese zwar dem Patienten (auf Verlangen) zur Einsichtnahme vorgelegt und davon Kopien erstellt werden müssen, das Original aber beim Arzt verbleibt und nicht ausgehändigt werden darf. Die Herausgabe einer vollständigen Krankengeschichte im Original durch einen Arzt kann von einem Patienten in der Regel nicht verlangt werden.³⁴

Auch aus dem kantonalen Recht, welches im Verhältnis Patient – Spital teilweise zu Anwendung gelangt, können sich Grenzen der Herausgabepflicht ergeben (bspw. aus einer kantonalrechtlichen Aufbewahrungspflicht).

³⁴ Aebi-Müller RE, Fellmann W, Gächter T, Rüttsche B, Tag B. Arztrecht. Bern: Schulthess; 2016. 43ff p.

Kann "Dateneigentum" aus dem Patientengesetz abgeleitet werden?

Im Verhältnis zwischen Patienten und Spitälern kommt statt Auftragsrecht kantonales öffentliches Recht zur Anwendung. Dieses enthält oftmals ähnliche Anforderungen an die Dokumentation, wie sie im Auftragsrecht zu finden sind.

Im Kanton Zürich beispielsweise verlangt § 17 Abs. 1 des Patientinnen- und Patientengesetzes (PatientenG), dass über jede Patientin eine laufend nachzuführende Patientendokumentation über die Aufklärung und Behandlung anzulegen sei. Diese Dokumentation kann entweder schriftlich oder elektronisch geführt werden (§ 17 Abs. 2 PatientenG). Sodann hält § 18 Abs. 1 PatientenG ausdrücklich fest, dass *Patientendokumentationen Eigentum der Institution seien*.

Diese Bestimmung weist demnach das sachenrechtliche Eigentum an der Patientendokumentation der Institution zu. Allerdings kann der kantonale Gesetzgeber dies aus rechtlicher Sicht nur insoweit tun, als "Eigentum" nach Bundesrecht überhaupt besteht. Da an einer "elektronisch" geführten Patientendokumentation, wie oben ausgeführt gar kein "Eigentum" bestehen kann (höchstens am Datenträger), ist die entsprechende Bestimmung wohl dahingehend auszulegen, dass eine elektronisch geführte Dokumentation so weit als möglich gleich behandelt werden soll, wie eine schriftlich geführte. Dies zeigt, dass selbst Gesetzgeber den Begriff des "Eigentum" an Daten teilweise aus juristischer Sicht nicht ganz klar verwenden.

VII. Fazit

Auch aus diesem letzten Beispiel geht hervor, dass unter "Dateneigentum" jeweils die unterschiedlichsten rechtlichen Aspekte verstanden werden. Die mit diesem Begriff aus unterschiedlichen Perspektiven angesprochenen Themen dürften derzeit am ehesten durch spezialgesetzliche Bestimmungen in den jeweiligen Sachgesetzen zu lösen sein, wie sie bspw. im Humanforschungsgesetz (HFG) teilweise zu finden sind. Bemerkenswert ist in diesem Zusammenhang, dass das Krankenversicherungsgesetz die Forschung mit Daten nicht besonders regelt.

Gerade im Rahmen von interdisziplinären Diskussionen und insbesondere beim Ausarbeiten von Verträgen zwischen den Akteuren sollte der Begriff "Dateneigentum" nicht oder nur mit Vorsicht verwendet werden und die mit diesem Konzept verbundenen Problematik thematisiert und geklärt werden, um Missverständnisse zu vermeiden. Jedenfalls ist

jeweils Klarheit darüber zu schaffen, aus welcher Sichtweise der Begriff verwendet wird bzw. welcher rechtliche Aspekt damit gemeint ist.

Abb. 1: Aussagen im Institutionellen Kontext

<p>CHLR8: Schwierig wird es dann wenn aus diesem Forschungsvorhaben etwas entsteht das man kommerziell verwerten kann. Da kommen dann auch wieder Fragen rein von Unternehmern die Fragen wie können wir uns da Beteiligen. Um ein Produkt zu entwickeln.</p>
<p>CHLR17: Nun, sehen Sie, es kommt darauf an, es gibt grosse Debatten, weil es Forschungsdaten gibt, die von Forschern produziert werden, sie gehören nicht mehr dem Patienten. Es ist die Aggregation von Daten, die Forschungsdaten liefert. Diese gehören nicht mehr dem Patienten.</p>

Abb. 2: Aussagen im Fokus der befragten Person betreffend "Dateneigentum"

<p>CHLR11: Für mich gehören sie nicht dem Patienten, weil es sich genau um anonymisierte Daten handelt, Es ist klar, dass wir danach wirklich sicherstellen müssen, dass die übertragenen Daten es uns nicht ermöglichen, die Verbindung zum Patienten herzustellen.</p>
<p>CHLR19: Der Patient hat natürlich immer das Recht die Verwendung von seinen Daten zu widerrufen, Einsicht zu nehmen was gemacht wird. Also er hat ja gewisse Rechte, die er...ich meine auf die er Einwirken kann.</p>
<p>CHLR10: Man sollte soweit kommen, dass man sagt, der Eigentümer ist immer noch der Patient, der seine Daten gibt. Aber momentan haben wir die Regelung, dass die Institution verantwortlich ist für die Daten, aber ich finde der Patient sollte ein partizipatives Recht haben, welches darüber hinausgeht einfach die Zustimmung zu geben für irgendwelche Bearbeitung es geht ja auch weiter, also man will dann auch kommerzialisieren und dass man eben auch für die Forschung, dort sollte eben auch der Patient partizipieren.</p>

Abb. 3: Aussagen aus der Perspektive eines partizipativen/geteilten "Dateneigentum"

CHLR8: Der Patient sagt ihm, und wir sagen natürlich uns. Ich glaube aber es ist eine geteilte, ein geteiltes Eigentum. Erschaffen tun wir die Daten, wir sind die Urheber der Daten. Weil unsere Ärzte schreiben die Berichte, die machen die Befundung. Es ist aber natürlich auch ein Auftrag, den wir für den Patienten machen. In der Regel ist das Auftragsergebnis gehört dem Auftraggeber.

VIII. Referenzen

- Bartlett, D., & Egloff, W. (2020). Das neue Urheberrecht.
- Cherpillod, I. (2012). Handkommentar, Bundesgesetz über das Urheberrecht und verwandte Schutzrechte (B. Müller & O. Reinhard (eds.)).
- Eckert, M. (2016). Digitale Daten als Wirtschaftsgut: digitale Daten als Sache. SJZ, 112(10).
https://www.mme.ch/fileadmin/files/documents/Publikationen/160515_sjz_beitrag_digitale_daten_als_wirtschaftsgut_-_digitale_daten_als_sache.pdf
- Gordon, C. A. (2017). Personal Information Management Systems (PIMS). Jusletter IT Flash.
- Hürlimann, D., & Zech, H. (2016). Rechte and Daten. Sui Generis.
- Picht, P. G. (2017). Dateneigentum und Datenzugang – Schutz von Geschäftsgeheimnissen als Alternative? Jusletter IT Flash.
- Rosenthal, D., & Jöhri, Y. (2021). Handkommentar zum Datenschutzgesetz (D. Rosenthal & Y. Jöhri (eds.)).
- Schmid, A., Schmidt, K. J., & Zech, H. (2018). Rechte and Daten - zum Stand der Diskussion. Sic!
- Thouvenin, F. (2017). Wem gehören meine Daten? Zu Sinn und Nutzen einer Erweiterung des Eigentumsbegriffs. Schweizerische Juristen-Zeitung, 113, 21–32.
- Thouvenin, F., Früh, A., & Lomard, A. (2017). Eigentum an Sachdaten: Eine Standortbestimmung. Schweizerische Zeitschrift Für Wirtschafts- Und Finanzmarktrecht, 25–34.
- Weber, R., & Thouvenin, F. (2018). Dateneigentum und Datenzugangsrechte – Bausteine der Informationsgesellschaft? Zeitschrift Für Schweizerisches Recht (ZSR).

Chapter 8: Social-Media und Social Messaging im Spital

Full Reference: Schneble, C.O, Martani A, Shaw, D.M., Elger, B.S. Social-Media und Social-Messaging im Spital – Ethische und Rechtliche Leitlinien, Jusletter, 2020;16(11), DOI: 10.38023/daa0b2fd-c9f1-4185-93ba-3bfd763938b

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, Andrea Martani¹, MSc, David Martin Shaw¹, PhD, Bernice Simone Elger^{1,2}, MD, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²University Center for Legal Medicine, University of Geneva, Switzerland

I. Einleitung

Wie jüngst eine Studie von CommonTime mit 823 Teilnehmenden gezeigt hat nutzen 43% der ÄrztInnen WhatsApp und andere consumer-orientierte Plattformen, bei Ihrer täglichen Arbeit (Martin, 2015). Sei es um eine(n) KollegIn um Rat bei einer Wundinfektion zu bitten oder auch um eine konsiliarische Meinung zu einem Röntgenbild einzuholen, ÄrztInnen nutzen Social-Messaging Apps vielfältig. Schon 2015 wurde in einer anderen Studie aus dem Vereinigten Königreich aufgezeigt, dass Social-Messaging Apps in der klinischen Arbeit von ÄrztInnen verbreitet sind, auch, um patientenbezogene Daten zu senden (Mobasher et al., 2015). Eine weitere Untersuchung aus Irland zeigte, dass die Gesundheitsfachpersonen eines Spitäles sehr oft WhatsApp für ihre Arbeit nutzten, obwohl sie sich bewusst waren, dass der Informationsaustausch durch solche Apps eine Gefahr für die Privatsphäre der Patienten darstellen (O’Sullivan et al., 2017). Obwohl solche Untersuchungen zum Gebrauch von WhatsApp in der Schweiz nicht vorhanden sind, ist die Vermutung gross, dass aufgrund der Einfachheit und Schnelligkeit solche Tools auch in Schweizer Spitälern verbreitet sind. Dies ist problematisch, weil die Verwendung von online Dienstleistungen wie WhatsApp möglicherweise eine Datenschutzverletzung verursachen können, da personenbezogene Daten der PatientInnen ausserhalb des sicheren IT-Systems des Spitäles verarbeitet werden (Kropp & Günther, 2017).

Richtlinien von Berufsverbänden, namentlich die Empfehlungen der FMH (FMH Swiss Medical Association, 2016), beschreiben extensiv wie der Umgang von ÄrztInnen mit Social-Media erfolgen sollte. In diesen Richtlinien wird umschrieben, wie ÄrztInnen Facebook nutzen sollen, beispielsweise im Rahmen der Arzt- Patienten-Beziehung oder in der Kommunikation mit Arbeits- und BerufskollegInnen. In der Schweiz fehlen jedoch spezifische Richtlinien zum Umgang mit Social-Messaging, wie sie jüngst der National Health Service (NHS) in England herausgegeben hat. Diese Richtlinien gehen intensiv auf den Umgang mit Social-Messaging ein und geben auch Sicherheitsempfehlungen für den Einsatz der gängigen Social-Messaging Apps wie z.B. “WhatsApp” oder “Telegram”. Die Empfehlungen der FMH fokussieren sich, im Gegensatz, auf soziale Medien, und, obwohl sie WhatsApp nennen, enthalten sie keine Empfehlungen für den Umgang mit Social-Messaging¹.

¹ In der Liste der Kategorien sozialer Medien, worauf sich die Empfehlungen richten (S. 8), fehlen Social-Messaging Apps wie WhatsApp oder Telegram. Fmh Verbindung der Schweizer Ärztinnen und Ärzte, Umgang mit Sozialen Medien (FMH Swiss Medical Association, 2016)

Im Rahmen unseres qualitativen Forschungsprojekts “Big Data Research Ethics” (8), führten wir Interviews mit SpitaljuristInnen und kantonalen Datenschutzbeauftragten durch. Ein Fokus der Interviews lag auf der Verknüpfung von verschiedenen Spitaldaten miteinander. In diesem Zusammenhang wurde auch der Gebrauch von Social-Media und Social-Messaging von mehreren Interviewten angesprochen. Zusätzlich zu den qualitativen Experteninterviews wurden auch die rechtlichen Rahmenbedingungen analysiert; diese spielen gerade im Umgang mit sensiblen Daten, wie es Gesundheitsdaten sind, eine entscheidende Rolle.

Der vorliegende Beitrag hat zwei Ziele: Zum einen werden die wichtigsten Erkenntnisse aus den Interviews bezüglich des Themas Social-Messaging/Social-Media im Spital analysiert und eingeordnet. In einem zweiten Schritt soll aufgezeigt werden, wie der Einsatz solcher Instrumente aus ethischer und rechtlicher Sicht gerechtfertigt werden kann.

II. Methode

a) Sampling

Die Studie ist Teil des Forschungsprojekts “Big Data Research Ethics”. Im Rahmen des Projekts wurden Interviews mit Forschenden, kantonalen Datenschutzbeauftragten und SpitaljuristInnen in der Schweiz und in den USA im Zeitraum von 2017 bis 2019 durchgeführt. Für die vorliegende Studie wurden alle 20 in der Schweiz getätigten Experteninterviews mit Datenschutzbeauftragten und SpitaljuristInnen ausgewertet, jeweils zur Hälfte kantonale Datenschutzbeauftragte (n=10) und SpitaljuristInnen/Datenschutzverantwortliche (n=10). Bei der Rekrutierung wurde Sorge getragen, dass ExpertInnen aus den unterschiedlichen Sprachregionen der Schweiz und Kantone vertreten sind, damit die interviewten Personen ihre Erfahrungen mit kantonal unterschiedlichen Datenschutzgesetzen einbringen konnten, die für die öffentlichen Institutionen inklusive Universitäts- und Lehrspitälern relevant sind.

Da es sich beim vorliegenden Projekt um eine nationale Studie handelt, wurde die Zuständigkeit via Ethikkommission Zentral- und Nordwestschweiz (EKNZ) abgeklärt. Diese erklärte sich als “nicht zuständig”, was bedeutet, dass die Studie nicht unter das Humanforschungsgesetz fällt und in allen Kantonen ohne weitere Auflagen durchgeführt werden konnte.

b) Interviews

Bei den Interviews handelte es sich um offene semistrukturierte Interviews, die im Durchschnitt ca. eine Stunde dauerten. Sie umfassten Fragen zum Thema Dateneigentum, zur Unterscheidung Bio-Samples versus reine Daten, zum Umgang mit Social-Media Daten in der Forschung sowie diversen Fragen zum Thema Datenschutz und zu Big Data.

c) Datenanalyse

Die Interviews wurden mittels der Software MAXQDA codiert und mittels thematischer Inhaltsanalyse (Braun & Clarke, 2006) ausgewertet. Hierzu wurden die Interviews gelesen, in verschiedene Teile unterteilt und mit verschiedenen Codes versehen, um Ähnlichkeiten in den Interviews hervorzuheben. Ein weiterer Forscher (D.S.) überprüfte die Codes, um sie zu präzisieren und zu verfeinern. Falls es Unterschiede in der Codierung gab, wurden diese diskutiert und ein Konsens gesucht. Für den vorliegenden Beitrag wurden nur die Abschnitte der Interviews untersucht, welche zum Themenkomplex Social-Messaging und Social-Media im Spital gehören. Das Thema Social-Messaging wurde nicht von allen Teilnehmenden angesprochen. Die vorliegende Analyse wurde durchgeführt aufgrund der erheblichen ethisch-rechtlichen Relevanz des Themas. Um die Aussagen in den Kontext der gängigen Praxis zu stellen, wurde zudem eine Analyse der gängigen Richtlinien und des rechtlichen Kontexts durchgeführt (siehe Diskussion).

III. Resultate

Nachfolgend werden die Resultate der thematischen Analyse des Themenkomplexes Social-Messaging und Social-Media im Spital detailliert vorgestellt.

a) Nutzung von Social-Messaging / Datenaustausch im Spital

Die Nutzung von Social-Messaging in der täglichen Arbeit ist nach Ansicht der interviewten ExpertInnen auch im klinischen Bereich ein Bedürfnis für ÄrztInnen. Insbesondere für die Terminplanung und für den unkomplizierten Tausch von Diensten sei Social-Messaging sehr beliebt. Hierbei seien in der Regel keine besonders schützenswerten Patientendaten involviert und der Einsatz wird daher von den im Spital tätigen ExpertInnen als wenig heikel angesehen. Kritischer war die Sicht der Datenschutzbeauftragten beim Einsatz von Social-Messaging auch zu Konsiliarzwecken:

“Ich bin jetzt gerade daran, eine Weisung zu schreiben über Social-Media und sage im Patientenkontext im Arbeitskontext: <No WhatsApp>. Jetzt sagen die [Anmerkung: die

Ärzte] aber um Dienste abzutauschen muss das aber möglich sein und so. Jetzt hat dann die Diskussion angefangen, dass es sowieso gang und gäbe ist, dass man Röntgenbilder ... [dass] mit dem Handy einen Screenshot macht – und es dem Kollegen schnell weiterschickt. Und ähhh, das ist dann wirklich datenschutzmässig heikel. (CHLR07)²“

Einer der Teilnehmenden weist jedoch darauf hin, dass das Teilen von Daten, wenn nicht zumindest mit WhatsApp, dann aber lokal möglich sein müsse. Insbesondere erwähnt er den Gesichtspunkt, dass Vertrauen eine wichtige Rolle im Arzt-Patienten-Verhältnis spielt und dem Arzt a priori auch ein besonnener Umgang mit Daten attestiert werden sollte, ähnlich wie das Vertrauen des Patienten in den Chirurgen anlässlich eines Eingriffs:

“Wir müssen sicherstellen, dass unsere Ärzte nicht missbräuchlich mit diesen [Daten umgehen], einerseits, dass Sie damit sorgfältig umgehen. WhatsApp ist wirklich ausser Diskussion, aber wir müssen dem Arzt vertrauen, dass er bei Bedarf einmal eine Datei extern abspeichern darf und er das nicht irgendwo verhökert und so. Wir vertrauen dem Chirurgen ja auch, dass er den Patienten in der Operation ja auch nicht umbringt. Also muss ich Ihm auch vertrauen, dass er mit den Daten allfällige missbräuchliche Sachen eben nicht macht.” (CHLR07)

In welche Richtung die künftige Nutzung von Social-Messaging gehen könnte, beschreibt ein weiterer Teilnehmer. Er schlägt die Integration von Third-Party Apps in den Klinik-Alltag vor, weil der Einsatz von eigenen Geräten der Angestellten (BYOD; Bring Your Own Device) eine Herausforderung darstelle. Durch die Bereitstellung von dedizierten Applikationen, die durch die Klinik zur Verfügung gestellt würden, könnten Risiken minimiert und mit den dadurch erhöhten Sicherheits-Mechanismen die Privatsphäre garantiert werden:

“Wir haben einige von denen, die wir managen können. Aber es ist die Realität, dass jeder ein Handy in der Tasche hat. Wir schreiben gerade eine Richtlinie, die besagt das man nur Apps nutzen darf, die von uns autorisiert sind. (...) Bei uns, ist es illegal, es zu benutzen, aber die Leute haben es auf ihrem Handy. So haben wir unsere eigene Version eines Krankenhaus WhatsApp, aber es ist eine App. Und wir können dort alles kontrollieren. Ich meine, wir kontrollieren, wir überwachen, dass der Schutz und die Privatsphäre des Patienten und der Mitarbeitenden respektiert wird.” (CHLR36)

² Die Pseudonymisierung der Teilnehmer erfolgt nach dem folgenden Schema <Land><Sample Gruppe><Fortlaufende Nummer> – Im Falle von CHLR07. Bedeutet dies Schweizer Gruppe, Juristen, Interview7.

b) Codes of Conducts

Klare Regeln, insbesondere im Umgang mit Gesundheitsdaten, spielen in den Augen der interviewten ExpertInnen eine wichtige Rolle. Neben den rechtlichen Regularien spielen auch zusätzliche institutionsspezifische Richtlinien (Soft Law) eine entscheidende Rolle für einen ethisch verantwortungsvollen Umgang mit Social-Media. Es erstaunt somit nicht, dass viele der Befragten institutionelle “Codes of Conducts” entwickelt haben. Diese sind oftmals sehr breit angelegt und umfassen Vorschriften von der Kleidung bis zu Social-Media.

“Ja. Auf jeden Fall. Wir haben Verhaltensregeln, sagen wir mal, die von der Kleidung über den Umgang mit Patienten bis hin zum Fotografieren reichen. Social-Media und so weiter.”
(CHLR10.)

Gleichzeitig wird festgehalten, dass die Umsetzung einer (Informatik-/Social-Media-) Richtlinie von zentraler Bedeutung wäre, aber dass in einem restriktiven finanziellen Umfeld oftmals die Ressourcen, und auch die Weisungsbefugnisse und Sanktionsmassnahmen fehlen, um die Richtlinie optimal umzusetzen und sicherzustellen, dass die darin enthaltenen Regeln auch befolgt werden.

“Wir haben auch eine Social-Media Richtlinie, wo wir darlegen was man sollte oder eben nicht sollte. Aber eine Richtlinie haben, heisst noch lange nicht, dass diese gelebt wird. Und dort fehlt es dann eben an den Ressourcen, um diese gut zu implementieren im Unternehmen und auch an Massnahmen wenn dies nicht eingehalten wird.” (CHLR08)

IV. Diskussion

Die Resultate der Interviews zeigen, dass die befragten SpitaljuristInnen und Datenschutzbeauftragten sich über die Verbreitung von Social-Messaging Apps im Klinikalltag und der damit verbundenen Risiken bewusst sind. Sie nehmen es als ein Bedürfnis der ÄrztInnen war, Social-Messaging nicht nur zur Planung der Dienste und zum Abtauschen derselbigen zu nutzen, sondern auch im Umgang mit schützenswerten Daten wie zum Beispiel zu Konsiliarzwecken. Im Nachfolgenden werden die Resultate aus unserer Untersuchung nochmals Punkt für Punkt aufgegriffen und in den Kontext der bestehenden Verbandsrichtlinien und den rechtlichen Kontext gesetzt. Schlussendlich werden ethische Aspekte diskutiert um dann mögliche “Best Practices” im Umgang mit Social-Messaging im Spitalkontext zu formulieren.

a) **Richtlinien**

Richtlinien spielen, wie auch die Teilnehmenden der Studie festgestellt haben, eine wichtige Rolle für den verantwortungsvollen Umgang mit Social-Media/-Messaging Technologien. Neben den in den Interviews angesprochenen institutionellen Richtlinien haben auch Berufsverbände Richtlinien im Umgang mit den sozialen Medien veröffentlicht, z.B. der Berufsverband der Schweizer Ärztinnen und Ärzte (FMH). Diese Richtlinien zielen in die gleiche Richtung, wie die von den Teilnehmenden angesprochenen institutionellen Richtlinien und es sind drei Stossrichtungen auszumachen: 1) Allgemeine Verhaltensempfehlungen, 2) Technisch/organisatorische Massnahmen 3) Kontakt zu Patienten. Die Richtlinien der FMH gehen sehr detailliert auf den Gebrauch von Social-Media ein und liefern den ÄrztInnen auch konkrete Modellbeispiele, an welchen sie sich orientieren können. Auf der anderen Seite sind Richtlinien, wie Sie der NHS zum Thema Social-Messaging in sehr detaillierter Weise entwickelt hat und die insbesondere auf die Vor- und Nachteile der einzelnen Lösungen fokussieren, in der Schweiz bislang nicht auszumachen (NHS Digital Health, 2018). Die Empfehlungen zum Umgang mit Sozialen Medien der FMH sind ein guter Start, sind aber eher generell gehalten und nur schwer auf die den speziellen Fall des Social-Messaging anzuwenden, des Weiteren lassen sie auch praktische Hinweise vermissen. Die Richtlinie des NHS bieten im Gegenteil eine Liste von konkreten Kriterien, welche den Gesundheitsfachpersonen helfen können, die sicherste Social-Messaging App auszuwählen³. Darüber hinaus gibt die englische Richtlinie auch zusätzliche praktikable Vorschläge, wie die App vom Fotoarchiv des Handys zu trennen (unlink), oder die Teilnehmer einer beruflichen Messaging-Gruppe regulär zu kontrollieren/prüfen, vor allem wenn man Administrator der Gruppe ist (Martin, 2015).

b) **Sicherstellung der Einhaltung der Richtlinien**

Eine grosse Problematik, die sich im gesamten Datenschutzzumfeld stellt, ist die Frage der Sicherstellung der Einhaltung solcher Richtlinien. Limitierend sind häufig die fehlenden Ressourcen, dies wird von den interviewten ExpertInnen mehrfach erwähnt. Die Problematik wird vor allem auf der kantonalen Ebene von den Datenschutzbeauftragten mit Besorgnis wahrgenommen, die für die öffentliche Verwaltung zuständig sind und somit auch

³ Die fünf Kriterien sind konkrete Eigenschaften, die Sicherheit (oder Unsicherheit) einer Social-Messaging App bestimmen: 1) End-to-End Encryption; 2) End User Verification; 3) Passcode Protection; 4) Remote-wipe; 5) Message retention. Eine Tabelle ist auch Verfügbar, die zeigt, welche der meistverwendeten Social-Messaging App (Whatsapp, Viber, Telegram und Signal) die Eigenschaften darbietet. Siehe NHS England, Fussnote 1, Seite 3–4.

die Oberaufsicht über Datenschutz im Spital haben. Die Zunahme der Datenmenge und der damit gekoppelte Einsatz neuer Technologien (wie im vorliegenden Beispiel Social-Messaging, aber auch maschinelles Lernen) wird zukünftig einen vermehrten Einbezug von Datenschutzbeauftragten erfordern.

Gerade aber auf der Ebene der kantonalen Datenschutzbeauftragten die letztendlich eine Kontrollfunktion über die öffentlichen Institutionen haben, hat sich gezeigt, dass diese oftmals personell schlecht ausgestattet sind, und dass ihnen die notwendigen Ressourcen oft fehlen (Privatim, 2018). Die Ausstattung reicht jeweils von einer Person in Teilzeit bis hin zu grösseren Teams. Erschwerend kommt hinzu, dass Spitäler aufgrund ihrer juristischen Form jeweils auch dem eidgenössischen Datenschutzbeauftragten unterstellt sein können.

c) **Gesundheitsdaten**

Gesundheitsdaten sind besonders schützenswerte Daten im Sinne des Datenschutzgesetzes (DSG, Art. 3). Das DSG bezeichnet die folgenden Kategorien als besonders schützenswert 1) die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten oder Tätigkeiten, 2) die Gesundheit, die Intimsphäre oder die Rassenzugehörigkeit, 3) Massnahmen der sozialen Hilfe und 4) administrative oder strafrechtliche Verfolgungen und Sanktionen⁴. Dies steht im Einklang mit den meisten kantonalen Datenschutzgesetzen, die ähnliche Normen enthalten. So definiert zum Beispiel das Gesetz über die Information und den Datenschutz (IDG) (17) des Kantons Zürich in § 3. ebenfalls Gesundheitsdaten als besondere Personendaten. Dies gilt ebenfalls für das IDG (18) des Kantons Basel-Stadt (§3 Abs. 4), sowie das Genfer Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD, Art. 4) (19). Eine Ausnahme bildet das Datenschutzgesetz des Kantons Bern (KDG) (20) das in Artikel 3 lit. B indirekt Gesundheitsdaten unter dem Begriff "geistigen oder körperlichen Zustand" subsumiert.

Zur Bearbeitung dieser Daten verlangt das DSG in manchen Fällen eine Rechtfertigung, zum Beispiel in Form einer ausdrücklichen Bewilligung des Datensubjekts (DSG, Art. 4 Abs. 5). Überdies regelt das DSG in Art. 12 die Weitergabe von Daten durch private Personen und vermerkt, ein Datenbearbeiter "darf nicht: (...) ohne Rechtfertigungsgrund

⁴ Die revidierte Fassung des Datenschutzgesetzes, welche am 25. September 2020 gutgeheissen wurde (die 100-tägige Referendumsfrist läuft), fügt zusätzliche Kategorien von besonders schützenswerten Personendaten hinzu, wie z.B. genetische Personendaten (Art. 5 lit. C Abs. 3). Ein Text des neuen Datenschutzgesetzes ist hier verfügbar: <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaefte?AffairId=20170059>.

besonders schützenswerte Personendaten oder Persönlichkeitsprofile Dritten bekanntgeben“. Im Fall von Social-Messaging wäre WhatsApp (bzw. sein Besitzer Facebook) als Dritter anzusehen, und wenn eine private Person die Daten einer anderen bekanntgeben wollte, müsste ein Rechtfertigungsgrund vorliegen. Wenn ein Arzt/Ärztin als private Person die Daten eines Patienten durch eine Social-Messaging App weitergeben wollte, müsste er z.B. die Einwilligung des Patienten einholen. In einem öffentlichen Spital kann aber der Arzt/Ärztin nicht als private Person agieren, sondern handelt als Teil des öffentlichen Organs (Rütsche, 2012), d.h. der Arzt/Ärztin untersteht in diesem Fall dem kantonalen Datenschutzgesetz. Im Kanton Basel-Stadt würde das zum Beispiel heissen, dass die Bekanntgabe der Gesundheitsdaten eines Patienten durch den Arzt/Ärztin einer der folgenden Anforderungen entsprechen müsste: 1) der Arzt/Ärztin ist durch ein Gesetz dazu ermächtigt; 2) die Bekanntgabe ist notwendig zur Erfüllung einer in einem Gesetz klar umschriebenen Aufgabe; 3) der Patient hat zugestimmt oder 4) der Patient ist nicht in der Lage seine Zustimmung zu geben, aber die Bekanntgabe liegt in seinem Interesse und seine Zustimmung darf in guten Treuen vorausgesetzt werden (§ 22 IDG BS). In Art. 35 regelt das DSG zudem die Verletzung der beruflichen Schweigepflicht, welche insbesondere im Arzt-Patienten-Kontext von Relevanz ist. Ein Verstoss gegen die berufliche Schweigepflicht wird auf Antrag mit Busse bestraft. Aufgrund der kantonalen Leistungsaufträge, die ein Spital erhält – und dies gilt auch für privat-rechtliche Körperschaften – sind zudem die entsprechenden kantonalen Datenschutzgesetze zu berücksichtigen (Rütsche, 2012).

d) Arztgeheimnis

Neben dem Datenschutzgesetz spielt, wie bereits erwähnt, auch die berufliche Schweigepflicht gemäss Art. 321 StGB eine Rolle. Während Daten an Dritte nur weitergegeben werden dürfen, wenn dazu ein Rechtfertigungsgrund vorliegt – z.B. hier die Einwilligung des Patienten (oder bei dessen Urteilsunfähigkeit einer berechtigten Vertretungsperson) – darf bei der Betreuung eines Patienten durch ein Ärzteteam (was in der Klinik der Fall ist) davon ausgegangen werden, dass eine stillschweigende Einwilligung für den Informationsaustausch innerhalb des Teams besteht (EDÖB, 2020). Diese Regelung, so wie sie der eidgenössische Datenschützer formulierte, lässt demnach einen Interpretationsspielraum für den Gebrauch solcher Apps im Klinik-Alltag offen. Dies steht im Widerspruch zum DSG da im Falle von Social-Messaging der Anbieter als “Dritter” angesehen werden könnte, da dieser die Daten bearbeiten könnte. Das Einholen einer Zweitmeinung bei

einem/r KollegIn ausserhalb des Betreuungsteams stellt allerdings eine Weitergabe von Daten dar, zu welcher der Patient einwilligen muss⁵, sofern die Anfrage nicht anonymisiert wird.

e) **Daten in der Cloud**

Bei Social-Messaging Apps⁶ werden die Daten oftmals im Ausland gespeichert und verarbeitet. Die Haftung zwischen dem Anwender und dem Anbieter wird in den allgemeinen Geschäftsbedingungen geregelt, oftmals auch nach ausländischem Recht. Dieser Umstand erschwert eine Durchsetzung allfälliger Haftungsansprüche. Teilweise stehen die Regelungen auch im Konflikt zu den lokalen Regeln (Tag, 2018). Anders sieht es beim Datenaustausch durch das geplante elektronische Patientendossier (EPD) aus; hier wird die Infrastruktur der Implementation von Patientendossiers durch verschiedene Stamm-Gemeinschaften betrieben, welche dem Schweizer Recht unterstehen⁷. Mögliche Haftungsgrundlagen im Fall eines rechtswidrigen Datenspeicherns in der Cloud sind Art. 97 des Obligationenrechts (OR), die Deliktshaftung nach Art. 41 OR sowie die Produkthaftung gemäss dem Bundesgesetz über die Produkthaftpflicht (PrHG) vom 18. Juni⁸.

f) **Ethische Aspekte**

Neben der in den vorangehenden Paragraphen dargelegten rechtlichen Situation spielen auch ethische Überlegungen beim Einsatz von Social-Messaging für den Austausch von Patientendaten eine wichtige Rolle. Abgeleitet von der in der modernen Bioethik weitverbreiteten Prinzipienethik (Beauchamp & Childress, 2013) sind die wichtigsten ethischen Grundprinzipien für den verantwortungsvollen Umgang mit sensiblen Gesundheitsdaten: a) Respektierung der Autonomie eines Menschen, b) Rechenschaftspflicht, c) Datenschutz und d) Datenfairness (Vayena, 2017). Diese werden im Nachfolgenden kurz mit einem spezifischen Fokus auf ihre Umsetzung in Bezug auf das Teilen von Daten durch Social-Messaging Apps vertieft. Der Respekt vor der Patientenautonomie, eines der wichtigsten Prinzipien der modernen biomedizinischen Ethik (Beauchamp & Childress, 2013), verlangt, dass die Wünsche des Patienten respektiert und die Würde der Person im Behandlungskontext oder in der Forschung geachtet werden. In Bezug auf den Umgang mit Daten

⁵ Vergleiche vorheriger Paragraph.

⁶ Social-Messaging Apps werden definiert als Apps welche den Austausch von Text aber auch andern Dokumenten zwischen einer oder mehreren Personen erlauben.

⁷ Vgl. Bundesgesetz über das elektronische Patientendossier vom 19. Juni 2015 (EPGD), SR 816.1.

⁸ Ibid.

bedeutet dies, dass die Datenbearbeitung und ein späteres data-sharing nur mit Einwilligung des Betroffenen erfolgen kann, solange es nicht gesetzlich definierte Ausnahmen, wie z.B. die obligatorische Weitergabe von Daten, z.B. wegen Meldepflichten im Rahmen des Epidemiengesetzes (Art. 12 EpG)⁹, gibt. Im Forschungskontext ist immer eine informierte Einwilligung (Informed Consent)¹⁰ für die Teilnahme an einem Forschungsprojekterforderlich, die verlangt, dass Studienteilnehmende genügend und mit grösstmöglicher Sorgfalt aufgeklärt werden. Eine Einwilligung des Probanden ist auch oft dazu erforderlich, um die Bearbeitung seiner Daten in Bezug auf das Forschungsprojekt zu rechtfertigen. Ausgenommen sind Studien in Notfallsituationen, wo spezifische Regelungen gelten und Studien mit bereits gesammelten Daten (Sekundärdaten), wo eine explizite Einwilligung nicht immer erforderlich ist (HFG Art. 32–34). Im Behandlungskontext gilt ebenfalls das Erfordernis der informierten Einwilligung, wobei auch hier für dringliche Situationen (Intensiv-, Notfall- und Rettungsmedizin) spezifische Regelungen gelten.

Das Prinzip der Rechenschaftspflicht verlangt, dass die Daten fair, transparent und gesetzesgemäss verarbeitet werden. Gerade im Behandlungskontext spielt dieses Prinzip eine wichtige Rolle. Behandlungen müssen im Rahmen eines Behandlungsprozesses aber besonders im Falle eines Schadens nachvollziehbar sein.

Der Datenschutz umfasst die Vertraulichkeit und den Schutz der Privatsphäre. Die rechtliche Umsetzung des Datenschutzes lehnt sich eng an den Begriff der Freiheit und somit an die Grundrechte an. Ebenfalls ist dieses Prinzip eng mit der Datensicherheit und Anonymisierung verbunden, somit sind vor allem Fragen technischer Natur oder der Governance der Institutionen verbunden. Datenfairness wird aus dem Gerechtigkeitsprinzip (Beauchamp & Childress, 2013) abgeleitet und umfasst die Nutzung von Daten für die Forschung und die Nutzung von Sekundärdaten für weitere Forschung.

Alle diese Prinzipien geben den Rahmen vor, in welchem sich ein verantwortungsvoller Einsatz von datenintensiven Technologien, wie zum Beispiel der Einsatz von Big Data und Social-Messaging, bewegen sollte. Denn oftmals können solche Technologien den Arbeitsalltag in der Klinik verbessern. Die Evidenzen für den hier diskutierten Fall, den

⁹ Vgl. Bundesgesetz über die Bekämpfung übertragbarer Krankheiten des Menschen (Epidemiengesetz, EPG) vom 28. September 2020, SR 818.101.

¹⁰ Mit dem von der SAMW entworfenen Generalkonsent, steht ein von der SAMW entwickeltes Instrument zur Verfügung, das den Gebrauch von Forschungsdaten und die Weiterverwendung (data sharing) schweizweit vereinheitlicht werden soll. Eine einheitliche Regelung ist schlussendlich wichtig, da eine solche das Vertrauen eines professionellen Patienten- /Arztverhältnis stärkt.

Einsatz von Social-Messaging, lassen sich durch diverse Untersuchungen zum Beispiel im Umfeld der Notfallmedizin (Johnston et al., 2015) oder als Zusatz in der Telemedizin (Giordano et al., 2017) untermauern. Eine Abwägung für oder gegen einen Einsatz hat auch immer die ebengenannten ethischen Aspekte zu berücksichtigen.

g) Empfehlungen für einen bewussten Einsatz von Social-Messaging im Spital-Kontext

Eine zentrale Rolle beim Einsatz von Social-Messaging im Spital-Kontext spielen Transparenz und eine klare Governance. Dies bedeutet, dass die grundlegenden ethischen Regeln wie Vertraulichkeit, Datenschutz, Einwilligung und Sicherheit eingehalten werden müssen. Gegen einen Einsatz von Social-Messaging im Spitalkontext gibt es somit aus rechtlicher und ethischer Sicht wenig Einwände. Im Gegenteil, durch die Verkürzung der Kommunikationswege und die Einfachheit der Handhabung solcher Tools können Verbesserungen für die Pflege resultieren, wie einzelne Studien bereits belegen (Kaliyadan et al., 2016; Kamel Boulos et al., 2016). Ein Einsatz solcher Tools sollte deshalb nach den untenstehenden formulierten Best Practices erfolgen. Hier schlagen wir eine vorläufige Liste vor, die auf den Resultaten unserer Studie basiert. Es soll so weit wie möglich auf Inhouse Lösungen gesetzt werden, die für Spitäler bereits existieren. Der Einsatz von “consumer” Tools wie WhatsApp, Telegram u.a. ist mit zu vielen, auch rechtlichen, Unsicherheiten belastet. WhatsApp und nicht speziell für diese Zwecke konzipierte Social-Messaging Apps bieten hierfür keinen genügenden Schutz und die Nachverfolgbarkeit aus Sicht der Institution kann mit diesen Tools nicht gewährleistet werden. Nachverfolgbarkeit ist im Spital-Kontext von grösster Bedeutung zum Schutz der Patienten und zur Vermeidung von Schäden. Der Einsatz von Social-Messaging sollte klar in Richtlinien geregelt werden. Richtlinien sollen sich hierbei nicht nur auf die Technik beschränken, sondern müssen auch Community Regeln definieren, so zum Beispiel wie die Kommunikation untereinander erfolgen soll oder in welchen Fällen es im Klinik-Alltag besser ist konventionelle Kommunikationsmittel zu nutzen. Die ethischen Prinzipien müssen beachtet werden. Der Patient muss so oft wie möglich seine Einwilligung gegeben haben, Daten müssen vertraulich behandelt werden und dürfen den Spitalkontext nicht verlassen. Transparenz im täglichen Umgang steht an oberster Stelle. Institutionelle Datenschützer müssen über genügend Ressourcen und Befugnisse verfügen, um Richtlinien durchzusetzen. Dies bedeutet, dass sie auch organisatorisch direkt der Spital-Leitung unterstellt sein sollten, damit die Unabhängigkeit innerhalb der Institution gewährleistet ist.

V. Limitationen

Da es sich um eine explorative qualitative Untersuchung handelt ist die Datenlage aufgrund der Methodik auf ein sogenanntes purposive sample beschränkt und stellt keine repräsentative Stichprobe dar (Braun & Clarke, 2006). Ziel und Vorteil dieses methodischen Ansatzes ist, eine möglichst grosse Bandbreite verschiedener Sichtweisen abzubilden, u.a. dadurch, dass ExpertInnen aus verschiedenen Regionen und Kontexten der Schweiz herangezogen wurden. Im Interview wurde der Gebrauch von Social-Media spontan nicht von allen Befragten angesprochen, so dass uns nicht von allen Beteiligten eine ausführliche Stellungnahme vorliegt. Aufgrund der ethischen Relevanz erachten wir es aber als sinnvoll diese Thematik sowohl aus ethischer aber auch aus rechtlicher Sicht in diesem Artikel eingehender zu beleuchten mit dem Ziel, eine breitere Diskussion und Lösungsfindung, z.B. durch schweizerische Richtlinien, anzuregen.

VI. Referenzen

- Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics*. Oxford University Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- EDÖB. (2020). *Schweigepflicht*. <https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/gesundheit/schweigepflicht.html>
- FMH Swiss Medical Association. (2016). *Umgang mit Sozialen Medien –Empfehlungen für Ärztinnen und Ärzte*. https://www.fmh.ch/files/pdf17/FMH-Empfehlungen_Social_Media_Langfassung_D.pdf
- Giordano, V., Koch, H., Godoy-Santos, A., Dias Belangero, W., Esteves Santos Pires, R., & Labronici, P. (2017). WhatsApp Messenger as an Adjunctive Tool for Telemedicine: An Overview. *Interactive Journal of Medical Research*, 6(2). <https://doi.org/10.2196/ijmr.6214>
- Johnston, M. J., King, D., Arora, S., Behar, N., Athanasiou, T., Sevdalis, N., & Darzi, A. (2015). Smartphones let surgeons know WhatsApp: An analysis of communication in emergency surgical teams. *American Journal of Surgery*, 209(1). <https://doi.org/10.1016/j.amjsurg.2014.08.030>
- Kaliyadan, F., Ashique, K. T., Jagadeesan, S., & Krishna, B. (2016). What's up dermatology? A pilot survey of the use of WhatsApp in dermatology practice and case discussion among members of WhatsApp dermatology groups. In *Indian Journal of Dermatology, Venereology and Leprology* (Vol. 82, Issue 1). <https://doi.org/10.4103/0378-6323.171638>
- Kamel Boulos, M. N., Giustini, D. M., & Wheeler, S. (2016). Instagram and WhatsApp in health and healthcare: An overview. In *Future Internet* (Vol. 8, Issue 3).

<https://doi.org/10.3390/fi8030037>

- Kropp, H., & Günther, U. (2017). Digitalisierung: Wie Kliniken digitale Patientendaten am besten schützen. *Deutsches Ärzteblatt, 1*.
<https://www.aerzteblatt.de/archiv/188218/Digitalisierung-Wie-Kliniken-digitale-Patientendaten-am-besten-schuetzen>
- Martin, C. (2015). *Instant Messaging in the NHS – Healthcare provider responses to the NHS consumer messaging & unsanctioned data sharing crisis*.
<https://alertiveworkforce.com/uploads/webpage-documents/5a9b043c-fa3b-4d6c-9a7c-39b7901dddbc.pdf>
- Mobasher, M. H., King, D., Johnston, M., Gautama, S., Purkayastha, S., & Darzi, A. (2015). The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. *BMJ Innovations, 1*(4), 174 LP – 181.
<https://doi.org/10.1136/bmjinnov-2015-000062>
- NHS Digital Health. (2018). *NHS issues guidance on use of instant messaging apps*.
<https://www.digitalhealth.net/2018/11/nhs-issues-guidance-use-of-instant-messaging-apps/>
- O’Sullivan, D. M., O’Sullivan, E., O’Connor, M., Lyons, D., & McManus, J. (2017). WhatsApp Doc? In *BMJ Innovations* (Vol. 3, Issue 4).
<https://doi.org/10.1136/bmjinnov-2017-000239>
- Privatim. (2018). *Digitaler Staat braucht Datenschutz*.
<https://www.privatim.ch/de/digitaler-staat-braucht-datenschutz/>
- Rütsche, B. (2012). Datenschutzrechtliche Aufsicht über Spitäler. *Digma, 6*.
- Tag, B. (2018). Gesundheitsdaten für die klinische Forschung und „Datenökonomie“ in der Zukunft (Interview). *Thema Im Fokus: Die Zeitschrift von Dialog Ethik, 137*, 4–7.
- Vayena, E. (2017). Ein ethischer Rahmen für den Austausch von Gesundheitsdaten. *Schweizer Ärztezeitung, 98*(36), 1138–1140.
<https://doi.org/10.4414/saez.2017.05941>

Chapter 9: The Cambridge Analytica affair and Internet-mediated research

Full Reference: Schneble, C.O., Elger, B.S., Shaw, D.M. The Cambridge Analytica affair and Internet-mediated research, 2018;19(8), DOI: 10.15252/embr.201846579

Full name(s) of author(s): Christophe Olivier Schneble¹, David Martin Shaw¹, PhD, Bernice Simone Elger^{1,2}, MD, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²University Center for Legal Medicine, University of Geneva, Switzerland

I. Opinion

In recent years, the Internet has become an essential source of data for research. A vast array of information can be collected via platforms, such as Amazon Mechanical Turk (Keith et al., 2017) and Survey Tools for specific research questions, or from harvesting social networks such as Twitter or Facebook (Gosling & Mason, 2015). Questions about data protection, consent and confidentiality will therefore become increasingly important (Rothstein, 2015), not only for users, but also for researchers and providers of such research and social media services. The European General Data Protection Regulation (GDPR) (European Union, 2016), with its paradigm of security and privacy by default, is a step in the right direction.

The recent scandal surrounding Facebook and Cambridge Analytica (Lewis et al., 2018) shows that these aspects of security and privacy are often not taken into account. Cambridge Analytica, a British consulting firm, was able to collect data from as many as 87 million Facebook users without their consent. The company gained access to 320,000 user profiles and their friends' data through the "thisisyourdigitallife" app developed by psychologist Alexandr Kogan of Cambridge University, UK, when he sold it to the company. Although the 320,000 Facebook users gave their consent for the app to use their data and that of their friends, the latter were not asked for consent and none consented to passing on their data to Cambridge Analytica. Though the information was anonymized and aggregated, the fact that app users were able to consent to the use of their friends' data is very unusual, both in terms of research ethics and social media terms and conditions. When it turned out that Cambridge Analytica had received this data in contravention of its rules, Facebook demanded that the company simply delete the data, without taking any further action to alert the public or warn users. There is still some ambiguity in the media coverage, and even Cambridge University remains unclear about exactly what happened (University of Cambridge, 2018). Nonetheless, the Analytica affair makes it very clear that Internet-mediated research requires much closer ethical oversight.

In traditional human subjects research in psychology and medicine, the ethical evaluation and approval of research projects at the institutional level have become accepted best practice for ethically correct research. However, the relevant guidelines were often designed for medical research, and the issues are substantively different from those in data science. For example, every participant has to give explicit consent for use of his or her personal

data in primary research and in much secondary research. Exemptions can be made for anonymized data and when obtaining consent is disproportionately difficult. But today's data science deals with vast amounts of data from various sources, some anonymized, some de-identified and some fully identifiable. Obtaining traditional informed consent from all participants—which normally requires a personal encounter between research personnel and the participant to inform them about use of their data (whether primary or secondary)—is not feasible because of the extremely high number of participants.

We recently conducted a review of ethical guidelines for Internet-mediated research at the top 10 Universities in the USA, the UK and Switzerland (Schneble et al.). The results clearly show that only a small minority of academic institutions has developed guidelines for data science. This in turn means that most universities are simply not prepared to perform ethical evaluations of research proposals that make use of vast amounts of data collected from social media, secondary apps and the Internet. In general, individual institutions, or at least major research associations, need to start developing appropriate guidelines. We are fully aware that the complex and fast-evolving field of data science does not make this an easy task. One possible way to overcome this problem is to adopt guidelines that serve more as critical reasoning advice rather than making specific suggestions for a single platform or technology, as those of the Association of Internet Researchers (Markham & Buchanan, 2012). However, this might also turn out to be problematic as ethical considerations in this field of research need a deep understanding of the legal and technical background surrounding it. If the aim remains to develop specific guidelines for Internet-mediated Big Data research, what are the most important issues that need to be addressed?

Institutional review boards (IRB) need to pay special attention to several issues that may not be adequately covered by existing guidelines. First, commercial or scientific value is often not obvious at the time of data collection (Cate & Mayer-Schonberger, 2013); if there is any possibility that future commercial use of data is possible, this should be mentioned in the initial consent. Second, data which are considered public can turn out to have a private character when being aggregated with other datasets. This calls into question the claim made in many guidelines that consent is only needed when data are private: in some situations, combinations of public data might also lead to data being revealed that participants or identifiable groups (especially if they are vulnerable) would want to be kept private. Third, researchers have to pay close attention to the sites from where data are

collected from and to the properties of the data: is the data protected/publicly available, and what were the “Terms and Conditions” the users agreed when sharing their information? A related point is that data that are anonymized today might be made re-identifiable tomorrow. Furthermore, Alexandr Kogan also worked outside Cambridge University: thus, another important aspect on the Facebook data breach is academics’ other jobs. For example, models developed in a research project might be used commercially owing to technology transfer, which could be ethically problematic.

One important practical issue is that IRBs in many countries are not required by law to review such research. However, while IRBs are more used to dealing with health data, the Analytica scandal illustrates vividly that people care deeply about other types of data that are usually subject to national data protection regulations, even if it is not legally regarded as “sensitive data”. If data science is to be conducted ethically, IRBs should not wait for the law to catch up, but should review such studies even if legislation does not mandate this. We also believe that social media companies should take the protection of user's data more seriously and deal with this issue more transparently. Currently, issues of privacy and data protection are listed in the terms and conditions but might not be comprehensible to members of the public. At the European level, the GDPR that came into force on 25 May 2018 demands in Article 7 that “the request for consent shall be presented in a manner which is clearly distinguishable from other matters, in an intelligible and easily accessible form, using clear and plain language” and that “any part of such a declaration, which constitutes an infringement of the Regulation shall not be binding”. However, it remains questionable whether the GDPR would in practice prevent the common “click and forget” consent systems common to Internet interfaces. This means that IRBs must remain vigilant regarding the information and consent options used in IMR research, particularly when using secondary data and considering waivers of further consent.

A more prominent and understandable way of presenting those issues, as is common practice in traditional clinical research, would prevent further scandals and make the tremendous amount of data that could be used for research more accessible without harming users, but as part of a trusted partnership triangle of social media companies, users and researchers. Facebook has already introduced a new way for users to control their data in a more user-friendly way; research institutions also need to find new ways to effectively guide researchers towards ethical Internet-mediated research.

II. Acknowledgements

This publication was produced as part of the National Research Program Big Data, Project Regulating Big Data research: A new frontier, financed by the Swiss National Science Foundation. Project Big Data Research Ethics No. 167211.

III. References

- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67–73. <https://doi.org/10.1093/idpl/ipt005>
- Gosling, S. D., & Mason, W. (2015). Internet Research in Psychology. *Annual Review of Psychology*, 66(1), 877–902. <https://doi.org/10.1146/annurev-psych-010814-015321>
- European Union, General Data Protection Regulation, Official Journal of the European Union (2016). https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf
- Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems Perspective of Amazon Mechanical Turk for Organizational Research: Review and Recommendations. *Frontiers in Psychology*, 8, 1359. <https://doi.org/10.3389/fpsyg.2017.01359>
- Lewis, P., Grierson, J., & Weaver, M. (2018). *Cambridge Analytica academic's work upset university colleagues*. <https://www.theguardian.com/education/2018/mar/24/cambridge-analytica-academics-work-upset-university-colleagues>
- Markham, A., & Buchanan, E. (2012). Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee. *Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. <https://doi.org/Retrieved> from www.aoir.org
- Rothstein, M. A. (2015). Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics. *The Journal of Law, Medicine & Ethics*, 43(2), 425–429. <https://doi.org/10.1111/jlme.12258>
- Schneble, C.O., Shaw D., Zimmermann, F., Wangmo T., Gaab, J., Elger B., Ethical issues in Internet Mediated (IMR) and Big Data Research in Psychology: A scoping review of guidelines. Under Review
- University of Cambridge. (2018). *Statement from the University of Cambridge about Dr Aleksandr Kogan*,. <https://www.cam.ac.uk/notices/news/statement-from-the-university-of-cambridge-about-dr-aleksandr-kogan>

Chapter 10: Google's Project Nightingale highlights the necessity of data science ethics review

Full Reference: Schneble, C.O., Elger, B.S., Shaw, D.M. Google's Project Nightingale highlights the necessity of data science ethics review, 2020;12(3), DOI: 10.15252/emmm.202012053

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc, Bernice Simone Elger^{1,2}, MD, PhD, David Martin Shaw¹, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²University Center for Legal Medicine, University of Geneva, Switzerland

I. Opinion

On November 14 last year, the British *Guardian* published an account from an anonymous whistleblower at Google, accusing the company of misconduct in regard to handling sensitive health data. The whistleblower works for Project Nightingale, an attempt by Google to get into the lucrative US healthcare market, by storing and processing the personal medical data of up to 50 million customers of Ascension, one of America's largest healthcare providers. As the *Wall Street Journal* had already reported 3 days earlier, and as the whistleblower confirmed, neither was the data anonymized when transmitted from Ascension nor were patients or their doctors notified, let alone asked for consent to sharing their data with Google (Copeland, 2019; Pilkington, 2019). As a result, Google employees had full access to non-anonymous patient health data. Google Health chief David Feinberg commented that all Google employees involved had gone through medical ethics training and were approved by Ascension (Feinberg, 2019).

Although Nightingale violated no laws regarding health data, the case raised fears over privacy. Ethically speaking, Ascension and Google ignored various standards for handling personal and sensitive data (Zook et al., 2020) First, this was clearly a breach of confidentiality as patients trusted the hospitals that their health data would be managed with the greatest respect for privacy. Second, patients were not asked for consent to share their data with Google for storage and processing. Third, patients' privacy was seriously disrespected, because data were not anonymized prior to its transfer to Google.

II. Complex Legal Regulations

Many privacy/data protection laws regard health-related data as special category that requires a higher level of protection than conventional data. In the USA, the Health Insurance Portability and Accountability Act (HIPAA) offers some legal protection. However, the law contains a loophole as it allows hospitals or healthcare providers to disclose information to business associates for further processing or quality improvement. In the EU, the General Data Protection Regulation (GDPR) defines health data as a special category of data, the processing of which is prohibited outside the EU unless explicit consent has been given (Art. 9 lit. 1). Overall, the legal situation regarding the protection of health data remains complex and further clarification through court cases will take time.

In the meantime, ethics continues to play an important role to protect the rights of subjects whose data are being collected, processed, shared and used. Some basic principles provide a stopgap measure until the law catches up with broader conceptions of health data. Indeed, years ago the collection and storage of personal health data were very limited, compared to today's omnipresent data gathering by institutional and commercial entities, social media and smartphone apps. The concept of health data as a distinct category is also being challenged, as various research projects have used “social media data” to derive and predict health-relevant issues such as risk of depression (Reece & Danforth,2017). This raises the question of whether all personal data should be regarded as health data (Schneble et al., 2020). In any case, the broadening notion of health data makes it all the more important that researchers respect ethical principles and that appropriate oversight mechanisms for data science are in place.

III. Towards Responsible Data Science

Scientists need to keep several points in mind when using personal health data for research. Below is some minimal guidance that can be used to evaluate such research projects and to stay clear of potential legal challenges.

a) Transparency

It must be clear for patients and participants for which purposes their data are being used and where and by whom they are used and processed. Thus, health data derived in the clinic need to remain in the context of treatment unless the patient has agreed to share it for further research.

b) Explicit Consent

The storing and processing of health data are always subject to prior explicit consent by the data subject (patient). Explicit consent means that patients need to be informed for which purposes their data are being used, where it is being stored and how their data will be used in the future before the project starts. Patients also have the right to decide whether data are shared in anonymized or identifiable form. Last but not least, patients should be able to have their data corrected or to withdraw from a study.

c) Data Anonymization

Traditional anonymization techniques are increasingly being challenged by novel technologies such as machine learning, so scientists need to pay more attention to this topic. Simply cutting out birth date, zip code and sex have been proven to be ineffective. Using more up-to-date and complex methods such as k-Anonymity is a better solution, albeit still not an absolutely certain method for ensuring anonymity.

d) Ethical Reflection

Researchers should ask themselves broader ethical questions about their research. These questions are not new and have been the focus of ethically sound research for decades: Have patients agreed to their data being shared? Does the patient benefit from his or her data being shared? A negative answer does not prevent data sharing and processing per se, but direct patient benefits make it easier to justify the research. Who else would benefit from data sharing and do these benefits justify the risks? Is it necessary to share identifiable data or can research be carried out with pseudonymized/anonymized data? Has enough effort been put into maintaining privacy, including a strict anonymization regime? Is trust/confidentiality between health institution and the patients maintained?

The Association of Internet Research has recently published new ethical guidelines that offer further challenging questions to consider (Franzke et al.,2019).

IV. The Urgent Need for IRB Review

While researchers should do more to consider ethical issues themselves, Project Nightingale accentuates the urgent need for institutional review boards and research ethics committees (IRBs and RECs) to evaluate data-driven research. While discussion continues and the ethical principles that should govern best practices in data science and AI research are being developed, such research should not be exempted from IRB evaluation. In the past, such bodies have always been challenged by novel technologies, but have proven their effectiveness at achieving a balance between the interests of science and those who participate directly in research or whose data and samples are used for research. Furthermore, depending on the jurisdiction and the topic, health research has to undergo mandatory IRB evaluation in both academic and commercial settings. However, in many cases data science lies outside the scope of it. Universities have therefore set up IRB structures to

evaluate this type of research, and many journals also request to see the approval by an IRB or REC.

In cases where research at the corporate level is not subject to IRB review—either because the jurisdiction does not require it or because such bodies do not exist—it might lead to a fundamental inequality between public and private research entities. It therefore remains dubious that many large “data companies” have not introduced such review bodies in their organization. Implementing ethical review is important to ensure transparency and a long-term trusted partnership between companies and the “customers” that we all are.

V. Acknowledgements

This publication was produced as part of the National Research Program Big Data, Project Regulating Big Data research: A new frontier, financed by the Swiss National Science Foundation. Project Big Data Research Ethics No. 167211.

VI. References

- Americans. The Wall Street Journal. <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790>
- Feinberg, D. (2019). Tools to help healthcare providers deliver better care. Google Blog. <https://blog.google/technology/health/google-health-provider-tools-launch/#!/%23click=https://t.co/4GSK17oVOc>
- Franzke, aline shakti, Bechmann, A., Zimmer, M., & Ess, C. M. (2019). Internet Research: Ethical Guidelines 3.0 Association of Internet Researchers.
- Pilkington, E. (2019, November). Google’s secret cache of medical data includes names and full details of millions – whistleblower. The Guardian. <https://www.theguardian.com/technology/2019/nov/12/google-medical-data-project-nightingale-secret-transfer-us-health-information>
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). All our data will be health data one day: the need for universal data protection and comprehensive consent (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/16879>
- Zook, M., Barocas, S., boyd, danah, Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. In *PLoS Computational Biology*. <https://doi.org/10.1371/journal.pcbi.1005399>

Chapter 11: Data Protection during the coronavirus crisis

Full Reference: Schneble, C.O., Elger B.S., Shaw, D.M., Data protection during the coronavirus crisis, 2020; 21(9), DOI: 10.15252/embr.202051362

Full name(s) of author(s): Christophe Olivier Schneble¹, MSc Bernice Simone Elger^{1,2}, MD, PhD, David Martin Shaw¹, PhD

Full affiliation(s):

¹Institute for Biomedical Ethics, Faculty of Medicine, University of Basel

²University Center for Legal Medicine, University of Geneva, Switzerland

I. Opinion

The SARS-CoV-2 virus has caused a worldwide pandemic with many deaths (WHO, 2020b) and healthcare systems being pushed to their limits. This makes it all the more important to identify infected people early on and to ensure that people comply with public health measures so as to reduce the spread of the virus. In contrast to most previous pandemics, we can now use smartphone and other digital data. This is not the first case of using smartphone data for public health: The WHO's go.data initiative successfully used those technologies to fight Ebola (WHO, 2020a). Common to all digital tracking methods is the fact that we deal with different types of data, such as geo-localization data or, via Bluetooth, close-contact data, that under normal circumstances would fall within the scope of data protection laws. However, there is growing evidence that governments that using such technologies in conjunction with other basic hygiene measures are more successful in fighting COVID-19 (Ferretti et al. 2020; Normile, 2020). The question remains of how data protection regimes should react to such states of emergency.

Many data protection regulations are based on individual liberty rights. Legislation such as the EU's General Data Protection Regulation (GDPR) safeguards the privacy rights of citizens, ensuring that data are only processed if there are reasonable grounds for doing so. The GDPR mentions public health in particular as such a reason, but the Regulation also ensures that any processing follows rigorous procedures to minimize privacy breaches. In times of pandemic, though, another key value is saving lives, which can be in tension with privacy rights and individual autonomy. Given that people have already sacrificed their physical liberty by staying at home, we must consider whether and to which extent their right to privacy may also have to be compromised to facilitate the public health response. As more measures are taken to restrict liberty in order to protect the population, it makes also sense that data protection should follow this shift and allow governments or researchers to use smartphone data.

One question raised is whether highly identifiable geo-localization data are needed to track people's movements or whether it is sufficient to use anonymized proximity data using Bluetooth to notify those who have been in contact with someone infected by the coronavirus. Another important consideration is to use only the minimum amount of data to achieve effective contact tracing.

Several countries have now launched tracing apps most of which only acquire proximity data (Table 11.1). The data are encrypted on the phone, and many apps require user consent to share the data. Google and Apple have been developing an API embedded in their operating systems that enable homogeneity between apps and countries that rely on the API and its privacy specifications (<https://www.apple.com/covid19/contacttracing>).

Table 11.1. Overview of selected contact tracing apps (Howell et al., 2020)

App (Country)	Developer	Data collected	Data collected	Google/Apple API
TraceTogether (Singapore)	Government Agency of Singapore/Ministry of Health	Proximity Data (Bluetooth)	Only possible once exposed to a case. Sharing data to find others	No
Pan-European Privacy-Preserving Proximity Tracing	Scientists/Non-profit Initiative	Proximity Data (Bluetooth)	Proximity Data (Bluetooth)	No
Tracking App (Korea)	Korea Government	Health Monitor Data provided by the patient	With Government (Self-health status assessment)	No
Corona-Warn-App (Germany)	Robert Koch Institute	Proximity Data (Bluetooth)	Alerts others in contact with positive Tested Person upon consent	Yes
Swiss Covid (Switzerland)	Federal Office of Public Health FOPH	Proximity Data (Bluetooth)	Proximity Data (Bluetooth)	Yes
Stop Covid (France)	Gouvernement Francais	Proximity Data (Bluetooth)	Alerts others in contact with positive Tested Person upon consent	No
COCOA (Japan)	Ministry of Health and Labour and Welfare	Proximity Data (Bluetooth)	Alerts others in contact with positive Tested Person upon consent	Yes

The EU parliament stressed that any digital measures against the pandemic must conform with current data protection and privacy legislation. Fundamental principles include the

voluntary use of such apps and sunset clauses to stop usage of the app once the pandemic is over (European Commission, 2020). The legal basis for processing is in line with the GDPR if the user gives consent, which should be “freely given”, “specific”, “explicit” and “informed” within the meaning of the GDPR. It should be expressed through a clear affirmative action of the individual; this excludes tacit forms of consent (e.g., silence; inactivity)” (European Commission, 2020).

In the USA, the CDC developed a guidance for case identification and contact tracing plans (US HHS, 2020). However, legal regulation across the USA remains heterogenous, mainly owing to the fact that federal law in form of the HIPAA safeguards only applies to covered entities: healthcare providers, health plans, and healthcare clearinghouses. Various states such as New York, New Jersey, and California have therefore set up their own regulations.

Tracing apps offer benefits on multiple levels: If they help to avoid exponential spread, fewer people will die and the economy will be less affected, thus preventing job losses. The basic ethical issues remain the same and have been discussed extensively for many years (Mittelstadt & Floridi, 2016).

It is essential that as many people as possible use tracing apps and it is therefore questionable whether consent on an individual basis is the right mechanism. Some governments have claimed the right to compel people to install the app and see this as a lower degree of interference with individual rights than other forms of confinement that many governments have imposed.

It goes without saying that *confidentiality, privacy, and transparency* are important. Data should only be used for the defined purpose of preventing further spread of SARS-CoV-2. Transparency is key to acceptance and use of an app.

Data minimization is one of the fundamental principles of data protection. Most of the current apps adhere to this principle by collecting only the random key broadcasted by the API. However, public authorities might want to access location data to monitor and reconstruct movement patterns. This could help to prevent a lockdown of smaller areas as it happened recently in Germany after infection of meatpackers in a slaughterhouse.

Strongly intertwined with *consent* to process individual data is *data sharing*. Fighting COVID-19 relies on detecting cases and informing others about potential exposure. This results in an ethical dilemma: From a public health perspective, it would be beneficial to

automatically share a positive result, but this would require identifying individuals. Balancing the risk of spreading the virus, it remains questionable why exposed persons and the authorities are not automatically notified especially as some disease prevention laws enable mandatory quarantine.

Although many people see the use of cell phone data as the first steps on a slippery slope toward surveillance of citizens, it can be argued that the use of these data and the abandonment of informational self-determination are justified in view of the public good. However, data use and access must remain proportional to the degree of the emergency and threat of loss of lives. Thus, the use of such data should only be allowed under certain conditions delineated in the following.

First, the government should limit the timeframe during which data can be used for monitoring or contact tracing (*sunset clause*). Using such an invasive tracing of large groups must remain an exception and should not be used to pursue a political agenda. Therefore, a periodic reevaluation is required to guarantee the sole use of data for the specific purpose of monitoring infections.

Second, using identifiable cell phone data for monitoring without consent should be limited to governments under the aforementioned premises, and any such use must be justified and proportional: The benefits resulting from tracking individuals must be significantly greater for society than the potential loss of privacy. In addition, any such use of data must not contravene any national laws.

Third, apps need to implement state-of-the-art consent mechanisms or have to be democratically endorsed before implementation. Uptake and efficiency will highly depend on transparency and user confidence. Transparency is a key for success; without it, app penetration will not reach sufficient numbers to defeat the virus. Decentralized data excluding geo-localization should be used as long as this permits efficient contact tracing and people comply with quarantine measures on their own (Thüsing et al, 2020). Lastly, apps should be justified by strong scientific arguments. If any of these conditions are not fulfilled, use of the app should be terminated. Ethical digital contact tracing is possible, but certain standards must be met.

II. Acknowledgements

This publication was produced as part of the National Research Program Big Data, Project Regulating Big Data research: A new frontier, financed by the Swiss National Science Foundation. Project Big Data Research Ethics No. 167211.

III. References

- European Commission. (2020). Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection.
- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Bonsall, D. G., & Fraser, C. (2020). Quantifying dynamics of SARS-CoV-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing. MedRxiv. <https://doi.org/10.1101/2020.03.08.20032946>
- Howell O'Neill, P., Ryan-Mosley, T., & Johnson, B. (2020). A flood of coronavirus apps are tracking us. Now it's time to keep track of them. MIT Technology Review.
- Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics*, 22(2), 303–341. <https://doi.org/10.1007/s11948-015-9652-2>
- Normille, D. (2020). Coronavirus cases have dropped sharply in South Korea. What's the secret to its success? *Science*. <https://doi.org/10.1126/science.abb7566>
- Thüsing, G., Kugelmann, D., & Schwartzmann, R. (2020, April 9). Datenschutz-Experten beurteilen Corona-App. *Frankfurter Allgemeine*.
- US HHS. (2020). Interim Guidance on Developing a COVID-19 Case Investigation & Contact Tracing Plan. <https://www.cdc.gov/coronavirus/2019-ncov/downloads/case-investigation-contact-tracing.pdf>
- WHO. (2020a). Go.Data: Managing complex data in outbreaks. <https://www.who.int/godata>
- WHO. (2020b). WHO Director-General's opening remarks at the media briefing on COVID-19. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>

Chapter 12: Discussion

I. Major Findings

This thesis aims to examine the ethical and legal implications that come with the use of Big Data in research. This objective has been empirically investigated with an interview study and several ethical and legal studies.

Autonomy is a key in research, and participants should freely choose whether they participate in research. The decision should be based on facts and figures to foresee the consequences and risks of participation. Consent has been the predominant mechanism for safeguarding autonomy in traditional research (Beauchamp, 2011). Big Data Research, however, questions this mechanism.

First with the ever-increasing datafication the boundaries of the distinction between sensitive data and non-sensitive¹ data become blurred. *Chapter 3* investigates this aspect and argues that this distinction might be problematic, especially if potentially non-sensitive data might become sensitive one day². *Chapter 4* gives the commercial perspective. Although both mechanisms informed consent and notice and consent cannot directly be compared as the latter is a form of contractual law, from an ethical point of view, both should have the goal of minimising harm to research subjects / user and in the case of Big Data to data subjects. It shows that information mechanisms of social media companies often neglect good ethical practices, especially when it comes to vulnerable subjects. This is also important for research as today social media data is subject to many research projects (Kaplan, 2015; Kitchin, 2014)

The findings of *chapter 5* illustrate the interplay of law and consent as part of the interview study. It denotes the complex and often multifaceted interplay of different sector-specific laws and how IRB boards could unify the mechanism at least at the level of soft law. A possible way to mitigate the importance of consent as a primary gatekeeper mechanism without neglecting ethical principles is the concept of Big Data exceptionalism, which has been subject to investigation in *chapter 6*. However, such a paradigm shift towards

¹ In addition to the distinction between personal data and factual data, the FADP also defines data requiring special protection. The FADP defines personal data as: "All information relating to an identified or identifiable person" Art. 3 FADP. There is often also a categorization of particularly sensitive data. The Swiss DPA defines particularly sensitive data (Art. 3. DPA) as data: 1) on religious, philosophical, political or trade union beliefs or activities, 2) on health, privacy or racial origin, 3) on social assistance measures, 4) on administrative or criminal prosecutions and sanctions.

² Traditional framework don't take temporal issues serious. For example, mining existing data-pools are seldomly regarded to problematic especially in the case of anonymous data. However big Data questions the concept of anonymization. Relying only on traditional consent mechanism is therefore insufficient.

regulating the use rather than collection might be tentative but might be incompatible with the Swiss approach and problematic in the case of highly sensitive data. Many of the participants still rely on consent as the predominant to respect the autonomy and lastly to avoid harm, as users and participant have the theoretically being informed about potential drawbacks. But the question whether the use rather than the regulation should governed add an important notion to the discussion which could be realized by implement IRB-Review at company-level.

Data ownership is often discussed as a possible solution to address shortcomings and increase control of data subjects over their data. *Chapter 7* exemplifies this issue as part of the interview study conducted in Switzerland. It shows how stakeholders and various laws address this aspect and argues that this term is fuzzy and unclear, and that diverse misunderstanding exists. On predominant belief is still associated with larger control over the data, a comm misconception in the public discourse, as many regulations especially in the health environment (scope of this study) provide enough safeguards for the data such as intellectual property rights, criminal law protections and generally in the case of the swiss data protection law by the concept of informational self-determination.

The findings of *chapter 8-10* and the whole thesis focus on governance and show how important it is to oversee and assess Big Data studies. One solution is to intensify IRB-oversight of Big Data research and implement clear guidelines on how such research should be implemented. Such clear guidelines would lead to a trusted partnership of researchers and participants and commercial entities and their users.

Finally, *Chapter 11* addresses the dilemma of different freedom rights, the right to data protection and the right to health as an example of the pandemics. It argues that in times of pandemic, data-protection could be lowered to implement tracking measures to fight covid but only if governed by a clear set of rules and framed by sunset clauses.

II. Consent in the Era of Big Data

a) Challenges and the Importance of Consent

The study's findings (*chapter 5, chapter 6*) make it clear; consent is still essential for research (*chapter 3,5-6,8-11*) and the use of data in the commercial field. This finding is in line with the interviews with researcher being part of the NRP 75 project (Favaretto, De Clercq, et al., 2020). Consent is also essential in various guidelines on Internet Mediated

research (Franzke et al., 2020). Current law such as the GDPR (General Data Protection Regulation, 2016) has introduced explicit consent³ to process data as an important safeguard furthermore the European Data Protection Board (former Article 29 Working Party) (The European Parliament, 2020) has defined guidelines on consent standardising the different aspects (see footnote 65). From an ethical point of view, it is remarkable that the guideline also address issue as the imbalance of power.

Besides being a safeguard, consent is one of the paramount mechanisms of autonomy. Respect for autonomy in this context means that participants of research and user should be genuinely informed before they take part in a study or sign up for a service. In the study in *chapter 5*, some participants voiced that this ideal of autonomy is unachievable when the traditional model of consent is applied to the context of Big Data. For two main reasons, one relating to users (participants) and the other to the use of data.

First, many participants in Big Data research are simply users of social media who do not (really) read the terms and conditions that they are signing up to, and thus could be participating in Big Data research without even realising (Cate & Mayer-Schonberger, 2013; Kaye et al., 2015; Schneble et al., 2020a). Although such users have agreed in the legal sense, ethically, this process does not safeguard their autonomy. Also, even if such users read the small print, they are unlikely to have enough information to grasp the consequences of Big Data research. To understand for example, that anonymisation might not be possible if their data are combined with other sources (Vayena et al., 2013). Further users might be coerced into agreeing if they want to use the service even if they do read terms and conditions, either because they get for example rewards from companies only available through an app, or for example many peers are using the app making forcing them to sign up as it is the sole and easy way to stay in touch.

Second, even when participants are presented with specific information and a consent form for a specific project, they are unlikely to have any grasp of the myriad possible avenues of future research projects that might use their data if the form also includes broad consent to future use. Broad consent respects autonomy inasmuch as some participants will be happy for their data to be used in all types of research but disrespects it in the sense that it

³ Article 4(11) of the GDPR stipulates that consent of the data subject means any: a) freely given, b) specific, c) informed and d) unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

constitutes a waiver to specific consent for future projects (Mikkelsen et al., 2019; Steinsbekk et al., 2013). Even if institutional review boards or research ethics communities review any such future studies – which is less likely in Big Data and social media research than in biomedical studies that reuse data – the use of broad consent pays lip service to traditional concepts of consent and autonomy rather than respecting them properly. Knowing about the twin dilemmas of failure to read/understand at the initial point of consent and the ethical failings of broad consent in this context, a different approach to consent might be required.

b) Consent, Sensitive Data⁴ and Temporality

Consent is critical when sensitive data is being processed. Thus, laws such as the GDPR and the FADP require explicit consent in the case of such data processing or prohibit such processing altogether. In the case of anonymised data, however, this mechanism does not apply but is rather achieved by an opt-out mechanism that was introduced at collection time by using broad consent. As long as for example in the health environment data was seen as distinct categories this approach might have been working. However, as Vayena et al (Vayena & Blasimme, 2018) pointed out, the notion of what is considered health data has considerably evolved. So-called biomedical Big Data nowadays ranges from data produced by health services, public health activities, and biomedical research to data registering exposure to environmental factors, such as sunlight and pollution, or data revealing lifestyle, socioeconomic conditions, and behavioral patterns, such as those from wellness and fitness apps, social media, and wearable devices. There is thus a paradigm shift from the notion of individual data producers and distinct categories to a more complex notion of a data ecosystem (Vayena & Blasimme, 2018).

Traditionally consent affects data that are primarily generated prospectively (ie, right now or recently) and have excluded data from the more distant past. However, research projects, such as the Time Machine project (EPFL, 2019), and the digitalisation of large amount of paper based repositories of vast amounts of past paper-based data will change this paradigm. Indeed, digitisation of past data is progressing rapidly, for example, in the digital humanities and in the digitalisation of large amounts of business and health data (Jones & Nevell, 2016). A new notion is thus introduced to the concept of consent. As

⁴ The focus of this paragraph will be on health data as in our study the focus has been mainly on health-related data. As health-related data can be seen as particularly sensitive other sensitive data would be subject to the same standards or in case of less sensitive data protection could be less rigorous.

today's consent is based on the assumption that the future use of data must be regulated, the linking of old or past data in connection with digitalisation would also require consent when an identified or identifiable person is concerned. Consent will therefore have to deal with the past, present, and future use of both past and prospective data (*chapter 3*). Current consent for data sharing is to a large extent blind owing to its broad nature. As argued in *chapter 3* consent systems actually need to look far back and far forward, as well as in close detail at the present. In essence, “consent must be capable of time travel, just as data is capable of time travel” (Schneble 2020a).

c) **New Consent Mechanisms**

Respecting the temporality of data (*chapter 3*) and acknowledging the ubiquity and variety of data scholars are voicing that traditional models of obtaining consent have reached their limits in today's highly data-driven and data-intensive research (Ioannidis, 2013; Mostert et al., 2016). The use of new technologies and automation offers great potential, especially when it comes to automating the consent process. Dynamic consent (Budin-Ljøsne et al., 2017; Williams et al., 2015) portals have already been proposed but have not yet been able to establish themselves. Such portals as well as data cooperatives (Hafen, 2015) are only a small step in the right direction. In an environment where people have to behave more and more reactively, it is about taking the burden away instead of imposing more. It will be central to automate the use of data. It would be conceivable to create individual profiles of individuals' preferences based on a questionnaire and then only release data if there is a positive match between the query and the profile. This would have advantages for the user but also for the researcher who would only have to define the danger fields and access the ethical automated data-sharing and collection fields. On the other hand, given that getting broad consent may be the only viable solution in some cases it is paramount to foster IRBs.⁵

III. The “Ethical Legal Tension”

There is often a conflict between ethical guidelines and current law. This tension is not new, as normative ethics points to possible optimal paths, and laws operate within common law and are ultimately the result of a political process (Cohn & Daar, 2017; Henriksen et al., 2017). Especially with the increased use of Big Data methods in research, this conflict

⁵ See also *chapter 5* for the empirical grounding of this claim.

is also part of the discussion about shaping Big Data research and developing ethical guidelines. In fact, this tension is also addressed by different approaches to data protection. Cate et al, who are opponents of a more liberal approach, have argued for self-regulation of data protection principles (Cate et al., 2013). In contrast, authors like McDermont et al advocate data protection as a human right, with its underpinning ethical principles of privacy, transparency, autonomy, and non-discrimination. These principles are particularly important in light of the increasing use of large amounts of data and algorithmic prediction (McDermont, 2017). This is also highlighted by Wachter et al, who called for a new right of reasonable inferences to close the accountability gap currently posed by "high-risk inferences," especially regarding predictions drawn from Big Data analytics with low verifiability and thus possible damaging effects on individuals (Wachter & Mittelstadt, 2019).

On closer inspection, this tension between law and ethics is not as strong as the following examples are intended to illustrate. First especially in the public discussion Data-ownership is often associated with better control and data sovereignty for data subjects. However, as *chapter 7* has been shown this tension is often fuelled by the different understanding of the disciplines/parties. The second example is pointing towards the dilemma of disrespecting the right to privacy to have benefit. Although the discussed case is special, as potentially many lives are at stake, it is a common narrative found as well in the ethical discussion as in different current laws that foresee exemptions if the use is for the benefit of society.

a) Data-Ownership

Interest in the question of who owns data has grown enormously significantly as part of the digital economy data are collected, stored and processed or a result from the evaluation of large data sets. They represent a tremendous economic value. As part of this process, the topic of "data ownership" has become the subject of an intensive and controversial jurisprudential discussion. Although many lawyers argue that an ownership model in the strict sense does not exist (Fröhlich-Bleuler, 2017; Thouvenin et al., 2017) as data is incompatible with property law, some others argue that, especially in the light of the commercial value of large data-sets some sort of ownership would make sense (Amstutz, 2018) or by their binary-nature could at least in the swiss-context be regarded as property (Eckert 2016). Another reason why so much attention is paid to ownership is the belief that if you own the data, you need consent to use or share it.

From an ethical point of view, the question of data ownership plays a secondary role. More critical are principles such as respect for privacy, the consent of the data subjects and the associated provision of sufficient information. An interesting aspect that emerged from the interviews was the view that although many of the participants were in line with the current legal doctrine of data being not part of property law, their personal beliefs differed. Many agreed that health data belongs to both the patient and the institution. Resulting from that standpoint they agreed that in the light of the "economic" value should benefit in whatever form. Concluding from this perspective data cooperatives such as MIDATA (Hafen, 2015) could be an increasingly important model. Data cooperatives manage the data of the data subjects and act as a single point of contact for both researchers and institutions. However, data cooperatives also have advantages for the data subjects, namely representing their interests vis-à-vis the data processors. This is precisely the point made by Amstutz (Amstutz 2018), as an advocate of data ownership. He assumes that data ownership and the assertion of rights would not be done by individuals but by a group. Data cooperatives could therefore be a model that could balance the current unequal weights between large providers of social media services and their users and give users thus more bargaining power (Blasimme et al., 2018; Hafen, 2015).

b) Data Protection vs. Life

There is often a tension between the use of data and the protection of the latter. Especially in the case of governments using or processing data. As the European guidelines have pointed out the imbalance of power between the government and citizens remains unequal and it is unlikely that governments can rely on consent. Whereas those examples have been mostly on the small scale, like municipality collecting data for informing citizens about construction work (The European Parliament, 2020) the pandemic has raised this question to a completely different level encompassing suddenly millions of millions of data being used. The increased availability of data, ranging from sensor data to geolocation data and the use of predictive algorithms, is a promising approach to fighting the pandemic and has already proven being successful on a smaller scale (WHO, 2020). Although many would argue that the collection of data for research might be reasonable and beneficial (McLennan et al., 2020), the use of such tools at the citizens level has been prone to be difficult and controversially discussed by the public, despite the fact that efficacy has been proved in some countries (Normille, 2020). From an ethical point of view, it is difficult to prohibit the introduction of such tools as, there is a great benefit to be gained from the apps

and, human lives can be saved. However, the introduction of such methods must be bound by strict guidelines (Ienca & Vayena, 2020).

First, the government should limit the timeframe during which data can be used for monitoring or contact tracing (sunset clause). Using such an invasive tracing of large groups must remain an exception and should not be used to pursue a political agenda. Therefore, a periodic reevaluation is required to guarantee the sole use of data for the specific purpose of monitoring infections. Second, using identifiable cell phone data for monitoring without consent should be limited to governments under the aforementioned premises, and any such use must be justified and proportional: The benefits resulting from tracking individuals must be significantly greater for society than the potential loss of privacy. In addition, any such use of data must not contravene any national laws. Third, apps need to implement state-of-the-art consent mechanisms or have to be democratically endorsed before implementation. Uptake and efficiency will highly depend on transparency and user confidence. Transparency is a key for success; without it, app penetration will not reach sufficient numbers to defeat the virus. Decentralised data excluding geo-localisation should be used as long as this permits efficient contact tracing and people comply with quarantine measures on their own (Thüsing et al, 2020). Lastly, apps should be justified by strong scientific arguments. If any of these conditions are not fulfilled, use of the app should be terminated. Ethical digital contact tracing is possible, but certain standards must be met. (Schneble et al., 2020a)

IV. Solution to the Big Data Consent Dilemma

a) The Different Levels of Ethical Principles

Current frameworks often lack an end-to-end pipeline; either address the principles at the meta-level or present solutions for specific problems or address specific algorithmic issues at programming level. This complicates a comprehensive overview, as each level is addressed independently of the other. To address this gap it is necessary that data scientist have a basic understanding of ethical concepts (Davis, 2020; Saltz et al., 2018) and ethicists a broader understanding of computer science respectively of the underlying discipline (Hedgecoe, 2004). The following paragraph illustrates the complexity and the interplay at the different levels.

Chapter 4 introduced a way of mapping the four principles (Respect for autonomy, beneficence, non-maleficence and justice) to mechanisms and concepts and those again to concrete measures and implementations thus presenting an end-to-end pipeline. Respect for autonomy, for example, would entail the concept of voluntariness, meaning that each participant in the case of research or each user in the case of the commercial social-media sphere can decide whether or not he wants be part of the research or whether or not he wants to sign up. In order to safeguard those principles, mechanisms need to be implemented⁶. In terms of the aforementioned aspect of voluntariness, one of the underlying mechanisms would be for example, consent. This mechanism in turn, must seek a concrete implementation. Speaking for consent, these would mean, for example, to set up sophisticated age-verification algorithms preventing children from sign-up. The flipside would also be to foster information for younger children by using, for example, pictorial consent or use short video sequences. Using for example, sophisticated mechanism, however, could have adverse effects and foster digital divide⁷, which in turns would have an impact on the principle of justice. This example shows the complexity and the difficulty of a one size-fits-all approach.

b) From Consent to Governance and IRB-Oversight

Not all exponents of the study have been giving consent such high importance but have instead opted that use of data should be regulated (chapter 6). This concept has also been known as Big Data exceptionalism (Nissenbaum, 2017; Rothstein, 2015); the rationale for this standpoint is that it becomes increasingly difficult to oversee data collection. Focusing on the use of data could therefore mitigate risks stemming from specific harming implementations of those technologies. However, such a paradigm change puts an essential burden on IRB's which are currently facing challenges to assess Big Data projects (Favaretto, De Clercq, Briel, et al., 2020; Ferretti et al., 2020; Ienca et al., 2018). Current advances in developing guidelines (Franzke et al., 2020), the setup of simple rules that researchers (Zimmer, 2018) can follow and the fact that more and more universities are implementing IRB's will mitigate those concerns on the long run. There is an increasing gap between public research and research with commercial data. This gap was also increasingly voiced

⁶ In the following we refer to *chapter 4* where the scope was to design ethical social-media apps that respect the rights of children and vulnerable persons.

⁷ For example, such an algorithm could be designed to use a credit card to prove that parental consent has been sought. Parental consent for example is mandatory in the GDPR if a child wants to use a service before the minimum age.

in the context of the NRP 75 research project. In addition, several scandals in commercial research have shown that it is urgently necessary that research projects should also be subject to increased ethics review. Possible ways out of this impasse would be the introduction of IRBS in larger companies. Chapter 9 and 10 delineate this necessity (Schneble et al., 2018, 2020). Beside that, IRBS should govern research; it is paramount to establish guidelines that researcher or users of those technologies can follow on the everyday base. Those guidelines need to be anchored in the institution and ultimately need to be able to be implemented, as the example of the use of social messaging in hospitals has shown (*chapter8*). It should also be borne in mind that such guidelines are prone to the tension of ethics and law, whereby the normative aspect of the law have to be met.

V. Towards a Procedural Approach

The previous chapters of the thesis have shown that there is a great interdependence between the legal and ethical and not least social aspects. This problem is not new, and this conflict also exists in clinical ethics and is part of good reflective ethical practice (Cohn & Daar, 2017; Henriksen et al., 2017). Therefore, mitigating risks and harms would entail a procedural approach where the aforementioned end-to-end pipeline would be subject to iterative cycles. For example, the introduction of the GDPR has affected the concrete implementations of the mechanism. The discovery of a new de-identification algorithm would affect anonymisation. Those two examples are only two of a myriad of examples that affect the implementation, but they show the necessity of revising and updating guidelines, procedures and code of conducts. From which the partnership of researchers/participants and companies /users will only benefit.

VI. Limitations

This thesis has several limitations. This section addresses the limitations of the interview study before declaring the overall limitation of this dissertation. For this study interviews have been conducted amongst swiss cantonal data protection officers and hospital lawyers and the comparative group in the US consisting of lawyers and ethical/legal scholars. While the swiss sample covers a majority of the DPO in cantons with research facilities, universities, and university hospitals, the US sample's findings are only limitedly generalisable for several reasons. The US field of privacy and data protection law is heterogeneous. First, it is sectorial, and it is hard to oversee all different sectors in such a study, the

study thus addressed general aspects and focused on the issue introduced by big tech companies and the distinct aspects of health data. Second there is no direct privacy law at the federal level, and many states have therefore enacted special laws.

Big Data research is such a broad field, and every discipline faces different challenges when it comes to Big Data research (Favaretto et al., 2020); results and elaborations therefore focus mainly on the health-related and Internet aspects. Domain-specific aspects for other disciplines would consequently need further investigation. Also, it has been difficult to delineate the research aspects from the "commercial aspects" as sometimes the scope of different laws such as the FDPA and the GDPR covers both the use of data in research and in the commercial and private sphere. On the flipside sector-specific laws such as the HRA also influence some parts of research that focus use of health-related personal data⁸.

VII. Conclusion and Further Research

As already addressed in the Limitation section, this PhD thesis does not provide comprehensive overview of all aspects of ethical Big Data research but rather focused on the aspects surrounding consent. Thus, it touched on the main elements of privacy and procedural mitigations and governance being key for the use of Big Data both in research and the commercial sphere. Therefore, further research would need to broaden up the scope and include aspects like the accountability of algorithms and fairness.

Although consent has probably been one of the most investigated aspects in research ethics, it is paramount that research investigates on topics like e-consent, although there has been substantial research on aspects of dynamic consent there are seldom platforms currently being productive. Further research would also include on how consent could be automated, meaning that a platform acts as proxy on behalf of users' preferences that he has defined earlier on. Another aspect that should be given more attention in the future is the development of data cooperatives that bundle rights to the data, which would certainly make the question of data ownership less important in the longer term.

When it comes to vulnerable groups and especially in the field of social-media further research should cover aspects of making the terms and conditions more tangible and on building principles of trust between users and large tech companies.

⁸ In Art 3 the HRA defines health-related personal data as: Information about an identified or identifiable individual that relates to his or her health or disease, including his or her genetic data.

VIII. References

- Amstutz, M. (2018). Dateneigentum. *Archiv Für Die Civilistische Praxis*, 218(2–4).
- Beauchamp, T. L. (2011). Informed consent: Its history, meaning, and present challenges. In *Cambridge Quarterly of Healthcare Ethics* (Vol. 20, Issue 4). <https://doi.org/10.1017/S0963180111000259>
- Blasimme, A., Vayena, E., & Hafen, E. (2018). Democratizing Health Research Through Data Cooperatives. *Philosophy & Technology*, 31(3), 473–479. <https://doi.org/10.1007/s13347-018-0320-8>
- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67–73. <https://doi.org/10.1093/idpl/ipt005>
- Cohn, F., & Daar, J. (2017). Ethics and Law: The Many Tensions. In *American Journal of Bioethics* (Vol. 17, Issue 7). <https://doi.org/10.1080/15265161.2017.1314052>
- Davis, K. C. (2020). Ethics in data science education. ASEE Annual Conference and Exposition, Conference Proceedings, 2020-June. <https://doi.org/10.18260/1-2--34589>
- EPFL. (2019). Time Machine: Big Data of the Past for the Future of Europe. <https://www.timemachine.eu/>
- Favaretto, M., De Clercq, E., Briel, M., & Elger, B. S. (2020). Working Through Ethics Review of Big Data Research Projects: An Investigation into the Experiences of Swiss and American Researchers. *Journal of Empirical Research on Human Research Ethics*, 155626462093522. <https://doi.org/10.1177/1556264620935223>
- Favaretto, M., De Clercq, E., Gaab, J., & Elger, B. S. (2020). First do no harm: An exploration of researchers' ethics of conduct in Big Data behavioral studies. *PLoS ONE*, 15(11 November). <https://doi.org/10.1371/journal.pone.0241865>
- Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0228987>
- Ferretti, A., Ienca, M., Hurst, S., & Vayena, E. (2020). Big Data, Biomedical Research, and Ethics Review: New Challenges for IRBs. *Ethics and Human Research*, 42(5). <https://doi.org/10.1002/eahr.500065>
- Franzke, aline shakti, Bechmann, A., Zimmer, M., & Ess, C. M. (2020). Internet Research: Ethical Guidelines 3.0. <https://aoir.org/reports/ethics3.pdf>
- Fröhlich-Bleuler, G. (2017). Eigentum an Daten? Jusletter.
- Hafen, E. (2015). MIDATA cooperatives - citizen-controlled use of health data is a prerequisite for big data analysis, economic success and a democratization of the personal data economy. *TROPICAL MEDICINE & INTERNATIONAL HEALTH*.
- Hedgecoe, A. M. (2004). Critical bioethics: Beyond the social science critique of applied ethics. In *Bioethics* (Vol. 18, Issue 2). <https://doi.org/10.1111/j.1467-8519.2004.00385.x>
- Henriksen, J. M., Remtema, M. S., & Whitford, K. (2017). Law Is Not Enough: The Importance of Ethics Consultation in Complex Cases. In *American Journal of*

- Bioethics (Vol. 17, Issue 7). <https://doi.org/10.1080/15265161.2017.1314055>
- Hunter, D., & Evans, N. (2016). Facebook emotional contagion experiment controversy. *Research Ethics*, 12(1), 2–3. <https://doi.org/10.1177/1747016115626341>
- Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PLOS ONE*, 13(10), e0204937. <https://doi.org/10.1371/journal.pone.0204937>
- Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. In *Nature Medicine* (Vol. 26, Issue 4). <https://doi.org/10.1038/s41591-020-0832-5>
- Ioannidis, J. P. A. (2013). Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research. *The American Journal of Bioethics*, 13(4), 40–42. <https://doi.org/10.1080/15265161.2013.768864>
- Jones, L., & Nevell, R. (2016). Plagued by doubt and viral misinformation: the need for evidence-based use of historical disease images. *The Lancet. Infectious Diseases*, 16(10), e235–e240. [https://doi.org/10.1016/S1473-3099\(16\)30119-0](https://doi.org/10.1016/S1473-3099(16)30119-0)
- Kaplan, F. (2015). A Map for Big Data Research in Digital Humanities. *Frontiers in Digital Humanities*, 2. <https://doi.org/10.3389/fdigh.2015.00001>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics : EJHG*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 205395171452848. <https://doi.org/10.1177/2053951714528481>
- McDermott, Y. (2017). Conceptualising the right to data protection in an era of Big Data. In *Big Data and Society*. <https://doi.org/10.1177/2053951716686994>
- McLennan, S., Celi, L. A., & Buyx, A. (2020). COVID-19: Putting the general data protection regulation to the test. In *JMIR Public Health and Surveillance* (Vol. 6, Issue 2). <https://doi.org/10.2196/19279>
- Mikkelsen, R. B., Gjerris, M., Waldemar, G., & Sandøe, P. (2019). Broad consent for biobanks is best-provided it is also deep. *BMC Medical Ethics*, 20(1). <https://doi.org/10.1186/s12910-019-0414-6>
- Mostert, M., Bredenoord, A. L., Biesart, M. C. I. H., & van Delden, J. J. M. (2016). Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics*, 24(7), 956–960. <https://doi.org/10.1038/ejhg.2015.239>
- Nissenbaum, H. (2017). Deregulating Collection: Must Privacy Give Way to Use Regulation? *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282
- Normille, D. (2020). Coronavirus cases have dropped sharply in South Korea. What's the secret to its success? *Science*. <https://doi.org/10.1126/science.abb7566>
- Rothstein, M. A. (2015). Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics. *The Journal of Law, Medicine & Ethics*, 43(2), 425–429. <https://doi.org/10.1111/jlme.12258>

-
- Saltz, J. S., Dewar, N. I., & Heckman, R. (2018). Key concepts for a data science ethics curriculum. SIGCSE 2018 - Proceedings of the 49th ACM Technical Symposium on Computer Science Education, 2018-January. <https://doi.org/10.1145/3159450.3159483>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020a). All our data will be health data one day: the need for universal data protection and comprehensive consent (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020b). Google's Project Nightingale highlights the necessity of data science ethics review. *EMBO Molecular Medicine*. <https://doi.org/10.15252/emmm.202012053>
- Steinsbekk, K. S., Kare Myskja, B., & Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem. *European Journal of Human Genetics*, 21(9). <https://doi.org/10.1038/ejhg.2012.282>
- The European Parliament. (2020). Guidelines 05/2020 on consent under Regulation 2016/679. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf
- General Data Protection Regulation, Official Journal of the European Union (2016). https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf
- Thouvenin, F., Früh, A., & Lomard, A. (2017). Eigentum an Sachdaten: Eine Standortbestimmung. *Schweizerische Zeitschrift Für Wirtschafts- Und Finanzmarktrecht*, 25–34.
- Thüsing, G., Kugelmann, D., & Schwartmann, R. (2020, April 9). Datenschutz-Experten beurteilen Corona-App. *Frankfurter Allgemeine*.
- Vayena, E., & Blasimme, A. (2018). Health Research with Big Data: Time for Systemic Oversight. *The Journal of Law, Medicine & Ethics*, 46(1), 119–129. <https://doi.org/10.1177/1073110518766026>
- Vayena, E., Mastroianni, A., & Kahn, J. (2013). Caught in the web: informed consent for online health research. *Science Translational Medicine*, 5(173), 173fs6. <https://doi.org/10.1126/scitranslmed.3004798>
- Wachter, S., & Mittelstadt, B. D. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review*, 2, 130. <https://ssrn.com/abstract=3248829>
- We Are Social. (2018). Most famous social network sites worldwide as of October 2018, ranked by number of active users (in millions). Statista - The Statistics Portal. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- WHO. (2020). Go.Data: Managing complex data in outbreaks. <https://www.who.int/godata>
- Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An

Application of Contextual Integrity. *Social Media and Society*, 4(2).
<https://doi.org/10.1177/2056305118768300>

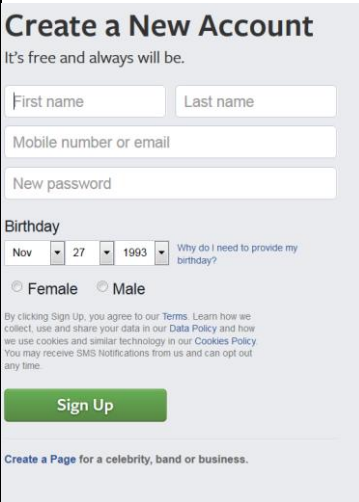
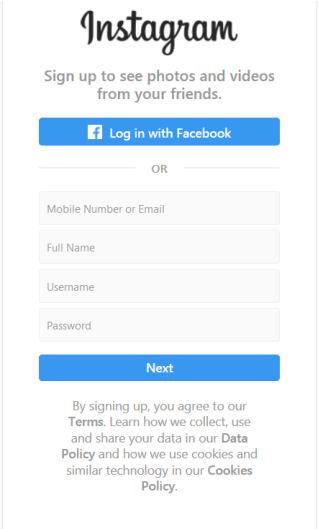
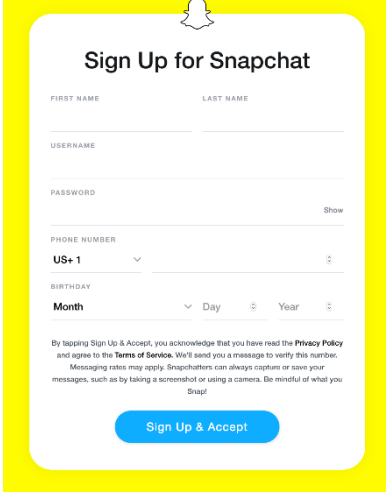
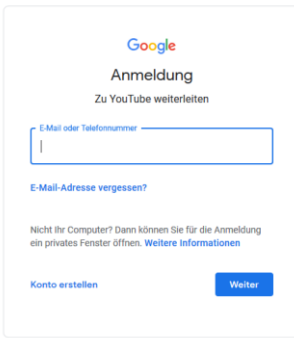

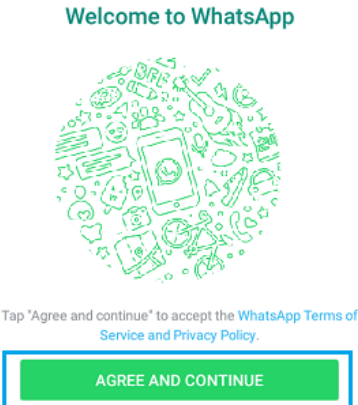
Appendix

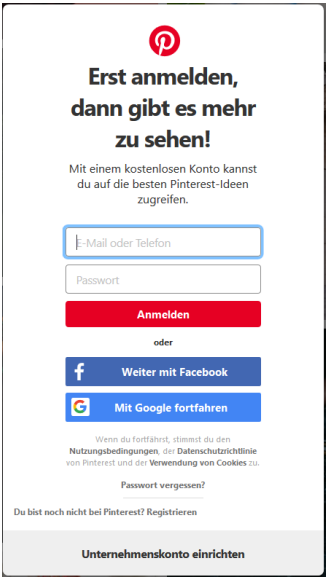

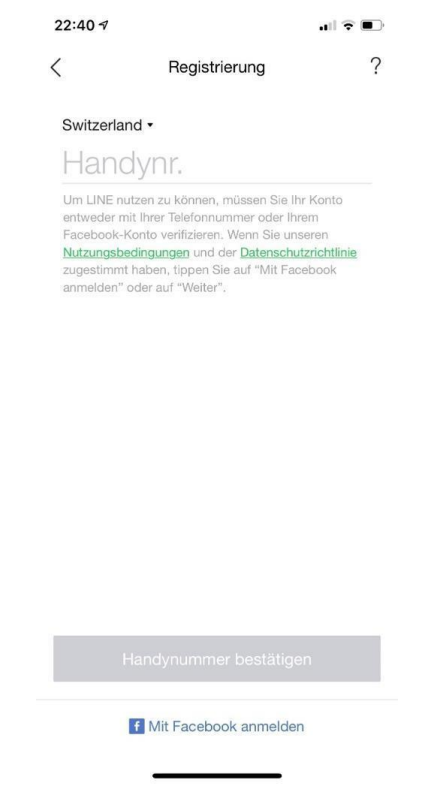
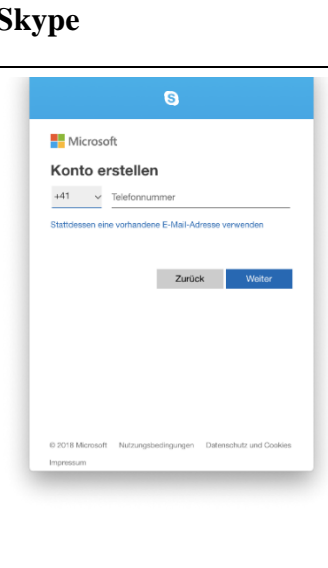
I. Tables

Appendix Table 1. Link to the Terms and Conditions

Platform/App	Link to the Terms and Conditions
Social Media	
Facebook	https://web.facebook.com/legal/terms/update
YouTube	https://www.youtube.com/t/terms
WhatsApp	https://www.whatsapp.com/legal/?eea=1#terms-of-service
Instagram	https://help.instagram.com/581066165581870
Tik Tok	https://www.tiktok.com/i18n/terms/
Twitter	https://twitter.com/en/tos
Skype	https://www.microsoft.com/en/servicesagreement/
Snapchat	https://www.snap.com/en-US/terms/
Pinterest	https://policy.pinterest.com/en/terms-of-service
LINE	https://terms.line.me/line_terms/
Ecosystems	
IOS (Apple ID)	https://www.apple.com/legal/internet-services/itunes/us/terms.html
Android Play Store	https://play.google.com/about/play-terms/index.html

Appendix Table 2 How are the Terms and Conditions presented?

Facebook	Instagram	Snapchat
 <p>Create a New Account It's free and always will be.</p> <p>First name <input type="text"/> Last name <input type="text"/></p> <p>Mobile number or email <input type="text"/></p> <p>New password <input type="password"/></p> <p>Birthday Nov <input type="text"/> 27 <input type="text"/> 1993 <input type="text"/> Why do I need to provide my birthday?</p> <p><input type="radio"/> Female <input type="radio"/> Male</p> <p>By clicking Sign Up, you agree to our Terms. Learn how we collect, use and share your data in our Data Policy and how we use cookies and similar technology in our Cookies Policy. You may receive SMS Notifications from us and can opt out any time.</p> <p>Sign Up</p> <p>Create a Page for a celebrity, band or business.</p>	 <p>Instagram</p> <p>Sign up to see photos and videos from your friends.</p> <p>f Log in with Facebook</p> <p>OR</p> <p>Mobile Number or Email <input type="text"/></p> <p>Full Name <input type="text"/></p> <p>Username <input type="text"/></p> <p>Password <input type="password"/></p> <p>Next</p> <p>By signing up, you agree to our Terms. Learn how we collect, use and share your data in our Data Policy and how we use cookies and similar technology in our Cookies Policy.</p>	 <p>Sign Up for Snapchat</p> <p>FIRST NAME <input type="text"/> LAST NAME <input type="text"/></p> <p>USERNAME <input type="text"/></p> <p>PASSWORD <input type="password"/> Show</p> <p>PHONE NUMBER <input type="text"/></p> <p>US+ 1 <input type="text"/></p> <p>BIRTHDAY Month <input type="text"/> Day <input type="text"/> Year <input type="text"/></p> <p>By tapping Sign Up & Accept, you acknowledge that you have read the Privacy Policy and agree to the Terms of Service. We'll send you a message to verify this number. Messaging rates may apply. Snapchatters can always capture or save your messages, such as by taking a screenshot or using a camera. Be mindful of what you Snap!</p> <p>Sign Up & Accept</p>
YouTube	Tik Tok	WhatsApp
 <p>Google Anmeldung Zu YouTube weiterleiten</p> <p>E-Mail oder Telefonnummer <input type="text"/></p> <p>E-Mail-Adresse vergessen?</p> <p>Nicht Ihr Computer? Dann können Sie für die Anmeldung ein privates Fenster öffnen. Weitere Informationen</p> <p>Konto erstellen Weiter</p> <p>Deutsch <input type="text"/> Hilfe Datenschutz Nutzungsbedingungen</p>	 <p>22:01</p> <p>Folge ich Für dich</p> <p>Du brauchst ein TikTok-Konto zum Fortfahren</p> <p>Mit Telefonnummer oder E-Mail-Adresse regi...</p> <p>ODER</p> <p>f G t</p> <p>Durch die Registrierung bestätigst du, dass du unsere Nutzungsbedingungen akzeptierst. Wir informieren dich über die Verarbeitung deiner personenbezogenen Daten und deiner Rechte. Mehr zu unserer Datenschutzerklärung.</p> <p>Du hast bereits ein Konto? Anmelden.</p>	 <p>Welcome to WhatsApp</p> <p>Tap "Agree and continue" to accept the WhatsApp Terms of Service and Privacy Policy.</p> <p>AGREE AND CONTINUE</p>

Pinterest	Twitter	Line
 <p>Erst anmelden, dann gibt es mehr zu sehen!</p> <p>Mit einem kostenlosen Konto kannst du auf die besten Pinterest-Ideen zugreifen.</p> <p>E-Mail oder Telefon</p> <p>Passwort</p> <p>Anmelden</p> <p>oder</p> <p>Weiter mit Facebook</p> <p>Mit Google fortfahren</p> <p><small>Wenn du fortfährst, stimmst du den Nutzungsbedingungen, der Datenschutzerklärung von Pinterest und der Verwendung von Cookies zu.</small></p> <p><small>Passwort vergessen?</small></p> <p><small>Du bist noch nicht bei Pinterest? Registrieren</small></p> <p>Unternehmenskonto einrichten</p>	 <p>Konto erstellen</p> <p>Name 0/50</p> <p>Telefon</p> <p>E-Mail verwenden</p>	 <p>22:40</p> <p>Registrierung</p> <p>Switzerland</p> <p>Handynr.</p> <p>Um LINE nutzen zu können, müssen Sie Ihr Konto entweder mit Ihrer Telefonnummer oder Ihrem Facebook-Konto verifizieren. Wenn Sie unseren Nutzungsbedingungen und der Datenschutzerklärung zugestimmt haben, tippen Sie auf "Mit Facebook anmelden" oder auf "Weiter".</p> <p>Handynummer bestätigen</p> <p>Mit Facebook anmelden</p>
 <p>Konto erstellen</p> <p>+41 Telefonnummer</p> <p>Stattdessen eine vorhandene E-Mail-Adresse verwenden</p> <p>Zurück Weiter</p> <p><small>© 2018 Microsoft Nutzungsbedingungen Datenschutz und Cookies Impressum</small></p>		

II. Interview Guide Swiss and US Expert Interviews¹

For comparative reason, the interview guides had the same structure. However, some questions were only subject to the Swiss participants (denoted with CH Experts-only)

- 1) Can we begin by exploring the role of data in your current professional context?
- 2) How would you describe both personal and factual data?
 - a. How valid is this distinction nowadays?
 - b. In the light of re-identification techniques: Should the distinction between personal and factual data be reconsidered?
- 3) Revision of Swiss Data Protection Act (FDPA) (CH Experts-only)

The Swiss Data Protection Act (FDPA) regulates personal data and is currently being revised. Have you been confronted with the revision? What are the changes accounting for?

- 4) Data Ownership - Considering the complexity involved who is accountable for the data/health data?
 - a. In your opinion, whose data is it?
 - b. In the case of research would you answer different?
 - c. If the data is anonymized would you answer different?
- 5) Biobanks: Assess legal status for data in biobanks, university hospitals.
 - a. Are there legal uncertainties?
 - b. Are there Ownership issues?
- 6) Human subject research - From your perspective, how does research with personal data compare to research with persons, (e.g. in a clinical trial)?
 - a. Should research with personal data be regarded as human subject research?
- 7) The federal act on electronic patient records entered into force on 15 April 2017. (CH Experts)

¹ Note: The interview guide was adapted during the course of the study according to findings.

- a. How does it influence the use of personal health data?
 - b. Are there legal dispositions you consider as critical?
- 8) Categorization of data used for research in “biological sample/genetic data” and “non-genetic-health-related data” and their regulation (Art. 32 and 33) (CH Experts-only)
- a. How balanced is the HRA, when we look at Art. 32 and 33, regarding the protection of human beings on the one hand and guaranteeing research interests on the other?
- 9) The general consent aims to regulate how subjects can consent to a general use of their data...
- a. How might the general consent influence both the research process and the protection of study participants?
- 10) International influence
- a. To what extent do international legislations (e.g. EU directives or regulations, Council of Europe) have an influence on Swiss /US Law?
 - b. To what extent do you think international legislations SHOULD have an influence on Swiss /US Law?
- 11) National legal challenges
- a. So there have been some influences from outside...When we now focus on Switzerland / US, what are the legal challenges that need to be considered here?
 - b. Do you think that the federal system in Switzerland / US poses a hard challenge to Big Data research?
- 12) There are many players in the health sector, such as physicians, hospitals, biobanks and pharmaceutical companies.
- 13) Let’s say some legal person from another country would ask you for advice on revising data regulation in his/her country: What would you tell him are the most important concepts he/she should consider in this process?

III. Curriculum Vitae

Personal Information

Address: Christophe Olivier Schneble, Rothusmatt 23, 6300 Zug

Date of Birth: 16.10.1977 **ORC-ID:** 0000-0003-1967-7129

Professional career - Short version

Phd-Candidate 2017 - 2021	Institut für Bio- und Medizinethik Universität Basel SNF Projekt Big Data Research Ethics Supervisors: Prof. Bernice Elger, Dr. David Shaw, Prof. Heiko Schuldt / External Supervisor Mira Burri
Secretary 2014 - 2017	Personalkommission, ETH Zürich Sekretär, Personalkommission der ETH Zürich
Managing Director 2010 - 2013	Department of Earth Sciences, ETH Zürich Managing Director Department of Earth Sciences
Scientist FH 2009 - 2010	Fachhochschule Rapperswil Research Scientist
Software Developer 2006 - 2009	the i-engineers AG Software Developer
IT Admin 2001 - 2006	Kantonsschule Zürich Birch Systemadministrator
Research Assistant 2000 - 2001	Fachstelle Bodenschutz, Kanton Zürich Research Assistant Fachstelle Bodenschutz
Founder 1998 - 2002	Candas Internet Services Gründer der eigenen Firma Candas Internet Services

Education (University onwards)

MSc Nov 2005	Master in Geography/Minor Geology University of Zurich
CAS Apr. 2010	CAS in Computer Science ETH Zürich

Languages

German: Mother tongue

English: Business fluent

French: mother tongue

Spanish: School knowledge

Teaching activities
2021

- Lecture Geo- and Environmental Ethics
- Tutor Lecture bioethics on the topic of primate research
- "Environmental ethics" as part of the lecture on bioethics

2020

- Lecture Geo- and Environmental Ethics
- Tutor Lecture bioethics on the topic of primate research
- "Environmental ethics" as part of the lecture on bioethics

2019

- Lecture Geo- and Environmental Ethics
- Summerschool Ethics, the Internet and the Society
- "Environmental ethics" as part of the lecture on bioethics
- Guest-Lecturer ELSI Training Day - Data protection and Big Data research ethics
- Tutor Lecture bioethics on the topic of primate research

2018

- Lecture Geo- and Environmental Ethics
- Input block Big Data in the context of the lecture Nanoethics
- Tutor Lecture bioethics on the topic of primate research

2017

- Seminar: Contemporary Debates in Bioethics: Thema Big Data Research Ethic

Invited Talks

2020 Open Science Summer School, University of Zurich

2019 Guest-Lecturer ELSI Training Day - Data protection and Big Data research ethics

Grants

2021 Impuls – Environmental Teaching fund University of Basel (2kCHF)

2018 Get on Track University of Basel (10% Research Assistant for 1 Semester)

Publications
Published papers

- Widmer, M. †, Schneble, C. O. †, Egli, P., & Elger, B. S. (2021). Eigentum an Patientendaten - Wer meint damit eigentlich was? *Pflegerecht - Pflege in Politik, Wissenschaft Und Ökonomie*, 2. (†Equal Author contributionship)
- Schneble, C. O., Favaretto, M., Elger, B. S., & Shaw, D. M. (2021). Social Media Terms and Conditions and Informed Consent From Children: Ethical Analysis. *JMIR Pediatr Parent*, 4(2), e22281. <https://doi.org/10.2196/22281>
- Schneble, C. O., Martani, A., Shaw, D. M., & Elger, B. S. (2020). Social-Media und Social-Messaging im Spital. *Jusletter*.
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2020). Data protection during the coronavirus crisis. *EMBO Reports*. <https://doi.org/10.15252/embr.202051362>
- Schneble, C. O., Elger, B. S., & Shaw, D. M. (2019). All our data will be health data one day: the need for universal data protection and comprehensive consent. *JMIR Preprints*. <https://doi.org/10.2196/16879>
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579. <https://doi.org/10.15252/embr.201846579>
- Schneble, C., Seidmann, T., & Huser, H. (2009). A Distributed Shared Memory Architecture for Occasionally Connected Mobile Environments BT - *Advanced Parallel Processing Technologies* (Y. Dou, R. Gruber, & J. M. Joller (eds.); pp. 288–301). Springer Berlin Heidelberg.
- Ienca, M., Schneble, C. O., Kressig, R., & Wangmo, T. (n.d.). Research Article Digital Health Interventions for Healthy Aging: A Qualitative User Evaluation and Ethical Assessment. *BMC Geriatrics*. <https://doi.org/10.21203/rs.3.rs-121871/v1> (Accepted for Publication)
- Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0228987>