Can We Still Archive? Privacy and Social Science Data Archiving After the GDPR

by

Allison Rae Bobyak Tyler

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2022

Doctoral Committee:

Professor Elizabeth Yakel, Chair
Professor Margaret C. Levenstein
Associate Professor Florian Schaub
Professor Janet A. Weiss
Dedication

This dissertation is dedicated to my family and friends who have supported me through this endeavor. And to Callie and Cas, without whose comfort and presence, the struggles would have been harder to overcome.
Acknowledgements

I want to take this opportunity to thank those without whom my successes these last six years would not have been possible. This dissertation and the work and learning that led into it owe much to the support I have received from my friends, family, colleagues, and mentors.

My deepest gratitude to my advisor, Beth Yakel. I know I was not the easiest PhD student to work with (I promise, I never intended to break as many bones as I have these last six years!) but her constant confidence in me and support of my goals has meant more than I can possibly express. As a student and a researcher, I could not have asked for a better mentor and advisor.

My deepest thanks as well to Maggie Levenstein her unending support and the opportunity to work with her and her team at ICPSR in my first year. The research experience I gained, my exposure to research data archives around the world, and the connections I made in those early years of my PhD have been beyond valuable. I could not have been luckier than I was in 2017 to join the Researcher Passport project.

I would also like to thank the other members of my dissertation committee, Florian Schaub and Janet Weiss, for their support, feedback, and guidance on this dissertation. This is a better dissertation because of them.

I am beyond grateful for the community of scholars and friends that have supported me over the years. Heeryung Choi, Jeremy York, Rasha Alahmad, Brad Iott, Elizabeth Whittaker, Allen Martel, Lia Bozarth… without our lunches, coffee breaks, writing sessions,
commiserations, practice talks, and everything in between, I would not have made it this far. Thank you.

To the staff and stakeholders of the Czech Social Science Data Archive, the Finnish Social Science Data Archive, DANS, and GESIS, thank you very much for allowing me to base my dissertation research on your organizations. Děkuji vám mnohokrát, kiitos paljon, hartelijk dank, vielen Dank! Your support and interest in my work have been invaluable, and I hope that my dissertation provides additional support for your on-going work. Thank you to Anja, Kara, and Marlene for supporting my year as a visiting researcher at GESIS.

To Kate McDonnell, Deborah Wiltshire, and Ann and Randy Yarger, for everything this last year and beyond, I cannot thank you enough.

The support of my family in so many ways and so many times have made this possible. Francis Tyler, Linda Holder, Lynette Sledge… for your unconditional love and support through everything, there are not enough thanks in the world that I can give.

The motivation for this dissertation came from an unanswered question remaining from the Alfred P. Sloan Foundation-funded Researcher Passport project (G-2016-7212).

This dissertation was funded in part by the Rackham Graduate Student Research Grant. This dissertation was funded in part by the United States State Department and the Directorate-General for Education and Culture of the European Commission’s Fulbright-Schuman Study/Research grant (2021-2022, deferred from 2020-2021). Thank you to everyone at the Fulbright Commission in Belgium and the German-American Fulbright Commission for your support and assistance over this last year.
Table of Contents

Dedication ....................................................................................................................................... ii

Acknowledgements ........................................................................................................................ iii

List of Tables .................................................................................................................................. x

List of Figures ............................................................................................................................... xii

List of Appendices ....................................................................................................................... xiv

Abstract ......................................................................................................................................... xv

Chapter 1 Introduction .................................................................................................................... 1

1.1 Background and problem statement ...................................................................................... 1

1.2 Theoretical framework .......................................................................................................... 7

1.3 Research questions .............................................................................................................. 11

1.4 Overview of research design ............................................................................................... 12

1.5 Significance ......................................................................................................................... 13

1.6 Structure .............................................................................................................................. 14

Chapter 2 Literature Review ......................................................................................................... 16

2.1 Definitions ........................................................................................................................... 17

2.2 Privacy conceptualizations .................................................................................................. 18

2.2.1 Informational privacy interpretations ........................................................................... 19

2.2.2 Privacy and technological advancement ...................................................................... 34

2.3 Data protection legislation ................................................................................................. 37

2.3.1 Historical motivations for privacy protection............................................................ 38
2.3.2 European Union goals and norms................................................................. 44
2.3.3 Pre-1995 data protection laws ................................................................. 46
2.3.4 Data Protection Directive ........................................................................... 51
2.3.5 General Data Protection Regulation .......................................................... 67
2.4 Data Archives and Organizational Change ..................................................... 99
  2.4.1 Data archiving ......................................................................................... 100
  2.4.2 GDPR and organizational change ......................................................... 104
  2.4.3 Privacy concerns for data archives ........................................................ 111
2.5 Conclusion .................................................................................................... 114
  2.5.1 Informational privacy in data archives .................................................. 117
  2.5.2 Operational impacts of new data protection legislation ....................... 118
  2.5.3 Addressing the gaps in the literature .................................................... 118
Chapter 3 Research Design and Methodology ..................................................... 120
  3.1 Rationale for a qualitative, multi-case study ............................................. 120
  3.2 Case Selection ........................................................................................... 123
    3.2.1 Czech Social Science Data Archive (ČSDA) ........................................ 129
    3.2.2 Finnish Social Science Data Archive (FSD) ......................................... 131
    3.2.3 Data Archiving and Networked Services ............................................ 133
    3.2.4 GESIS Leibniz Institute for the Social Sciences ................................. 135
  3.3 Data Sources .............................................................................................. 137
    3.3.1 Interviews ............................................................................................ 139
    3.3.2 Administrative documentation ............................................................ 142
  3.4 Data analysis .............................................................................................. 153
  3.5 Validity ....................................................................................................... 158
  3.6 Limitations ................................................................................................. 160
Chapter 4 Case Study Reports and Findings

4.1 Czech Social Science Data Archive (ČSDA)

4.1.1 Archive description

4.1.2 Data archiving in the Czech Republic

4.1.3 Privacy and the GDPR in the Czech Republic

4.1.4 Timeline of GDPR response

4.1.5 Priorities in the response

4.1.6 Challenges at ČSDA

4.1.7 Impacts on ČSDA operations

4.2 Finnish Social Science Data Archive (FSD)

4.2.1 Archive description

4.2.2 Data archiving in Finland

4.2.3 Privacy and the GDPR in Finland

4.2.4 Timeline of GDPR response

4.2.5 Priorities in the response

4.2.6 Challenges at FSD

4.2.7 Impacts on FSD operations

4.3 Data Archiving and Networked Services (DANS)

4.3.1 Archive description

4.3.2 Data archiving in the Netherlands

4.3.3 Privacy and the GDPR in the Netherlands

4.3.4 Timeline of GDPR response

4.3.5 Priorities in the response

4.3.6 Challenges at DANS

4.3.7 Impacts on DANS operations
List of Tables

Table 3.1 Case study field site selection criteria................................................................. 125
Table 3.2: Interviews by archive and respondent category............................................... 140
Table 4.1 Controller (C) or Processor (P) role adopted by each data archive for which type of personal data. .......................................................................................................................... 363
Table 4.2 Changes or new implementations at the four data archives categorized under "Regulating Information Flows."................................................................................................................ 368
Table 4.3 Changes or new implementations at the four data archives categorized under "Minimizing Harms.”................................................................................................................................. 370
Table 4.4 Changes or new implementations at the four data archives categorized under "Response to Technological Risks.”............................................................................................................ 375
Table 4.5 New roles established to manage data protection and privacy issues................. 377
Table 4.6 Challenges ČSDA, FSD, DANS, and GESIS faced while responding to the GDPR. 388
Table D.1 Qualitative code set used for interview and archive documentation analysis. ....... 444
Table E.1 List of Documentation for the Czech Social Science Data Archive ..................... 461
Table E.2 List of Documentation for the Finnish Social Science Data Archive ..................... 462
Table E.3 List of Documentation for DANS........................................................................... 464
Table E.4 List of Documentation for GESIS........................................................................... 466
Table F.1 National responses for European Union and field site locations to: QB4 (Eurobarometer 83.1): How much control do you feel you have over the information you provide online?................................................................................................................. 470
Table F.2 National responses for European Union and field site locations to: QB5 (Eurobarometer 83.1): How concerned are you about not having complete control over the information you provide online?................................................................................................................. 470
Table F.3 National responses for European Union and field site locations to: QB17 (Eurobarometer 83.1): Should your explicit approval be required before any kind of personal information is collected and processed? ........................................................................................................ 471

Table G.1 National responses for European Union and field site locations to: QB17 (Special Eurobarometer 487): Have you heard of the General Data Protection Regulation, which came into force in 2018? and to: QB18(1-6) Special Eurobarometer 487): The General Data Protection Regulation guarantees a number of rights. Have you heard of each of the following rights? ........................................................................................................................................ 472

Table G.2 National responses for European Union and field site locations to: QB9 (Special Eurobarometer 487): How much control do you feel you have over the information you provide online? ................................................................................................................................................. 474

Table G.3 National responses for European Union and field site locations to: QB10 (Special Eurobarometer 487): How concerned are you about not having complete control over the information you provide online? ........................................................................................................................................ 474
List of Figures

Figure 1.1 The relationship between the three approaches to privacy and informational privacy. 8
Figure 2.1 The relationship between the three approaches to privacy and informational privacy. 33
Figure 3.1 GESIS Data Archive and SowiDataNet|datorium data are searched for through the main GESIS webpage’s search function. 152
Figure 4.1 Screenshot of the Nesstar catalog webpage showing both the quantitative ČSDA data archive and the qualitative Medard archive access to metadata records. 166
Figure 4.2 ČSDA's End-of-Year total number of data sets in the data archive (2011-2021). 205
Figure 4.3 Total data sets published per year at ČSDA (2012-2020). 206
Figure 4.4 The growth in number of registered ČSDA users (2011-2020). 207
Figure 4.5 Annual total data sets available for reuse at FSD (2010-2021). 236
Figure 4.6 Qualitative, quantitative and total data sets published at FSD (2015-2021). 237
Figure 4.7 Qualitative, quantitative, and total data downloads from Aila (2015-2021). 239
Figure 4.8 Annual Aila data access requests (2015-2021). 240
Figure 4.9 Total registered Aila users (2015-2020). 240
Figure 4.10 DANS homepage with the links to the thematic data stations, as of December 10, 2021. 247
Figure 4.11 Question about presence of personal data in data deposit, added to the DANS deposit form in response to the GDPR. 267
Figure 4.12 Annual total data sets in DANS EASY (2010-2020). 289
Figure 4.13 Number of data sets published in DANS EASY per year (2010-2020). 289
Figure 4.14 Annual end-of-year total archaeological data sets in DANS EASY (2010-2020). 290
Figure 4.15 Annual end-of-year total data sets by discipline, excluding archaeology (2010-2020)..................................................................................................................................................... 291

Figure 4.16 Biannual Availability of DANS EASY Data sets by Access Category (2012-2018). ..................................................................................................................................................... 292

Figure 4.17 Total annual data downloads from DANS EASY (2010-2019). ......................... 293

Figure 4.18 The research project lifecycle into which GESIS orients its services. ................. 298

Figure 4.19 The mandatory data use and consent agreement in the pre-GDPR GESIS account registration. ................................................................................................................................. 334

Figure 4.20 The mandatory data use and consent agreement in the post-GDPR GESIS account registration. ................................................................................................................................. 334

Figure 4.21 Annual total data sets in the GESIS Data Archive and SowiDataNet|atorium (2010-2021). .......................................................................................................................................... 351

Figure 4.22 Annual total data sets in the GESIS Data Archive by Access Category (2010-2021). ..................................................................................................................................................... 351

Figure 4.23 Annual total data sets in SowiDataNet|atorium by access category (2011-2018). 352

Figure 4.24 Annual data set publications in the GESIS Data Archive and SowiDataNet|atorium (2010-2020). ..................................................................................................................................................... 353

Figure 4.25 Annual data downloads for GESIS Data Archive by access category (2010-2018). ..................................................................................................................................................... 354

Figure 4.26 Annual data downloads for SowiDataNet|atorium (2014-2018). .......................... 354

Figure 4.27 GESIS Secure Data Center usage by annual number of users and requested output checks. ............................................................................................................................................... 355

Figure 4.28 GESIS Secure Data Center usage by annual number of users and user days........ 356

Figure 4.29 Timeline of GDPR-related events at ČSDA (purple), FSD (maroon), DANS (orange), and GESIS (blue) (2016-2022). ................................................................................................................................. 358

Figure 5.1 Theoretical framework of informational privacy behaviors in data archives......... 422
List of Appendices

Appendix A Data Archive Staff Interview Protocol ................................................................. 433
Appendix B Legal Interview Protocol .................................................................................. 438
Appendix C Archive Stakeholder Interview Protocol ............................................................ 441
Appendix D Qualitative Code Set ....................................................................................... 444
Appendix E List of Data Archive Documentation ................................................................. 461
Appendix F Eurobarometer 83.1 Data Tables ..................................................................... 470
Appendix G Special Eurobarometer 487 Data Tables ........................................................... 472
Abstract

This dissertation investigates the impact of new privacy regulations on scientific research data infrastructures. It focuses on the experiences of and outcomes at four European Union (EU) social science data archives—the Czech Social Science Data Archive, the Finnish Social Science Data Archive, Data Archiving and Networked Services, and GESIS — Leibniz Institute for the Social Sciences—during the period after 2016 as they attempted to become compliant with the EU’s new General Data Protection Regulation (GDPR), a revised EU-wide law that brought further restrictions to the use and reuse of personal data. However, as the GDPR and the subsequent national legislation sought to increase protections for an individual’s personal data, they did not significantly consider how the new requirements would affect the research community for whom personal data are vital to creating new knowledge. This research seeks to understand how the changes required to comply with the GDPR have affected the data archives’ abilities to operate.

This study is a qualitative, multiple case comparative case study of four national social science data archives. It relies on semi-structured interviews with data archive staff, senior leadership, IT staff, legal advisors, and external stakeholders, document analysis, and data archive usage statistics.

At the start of this dissertation, I developed a framework to better understand informational privacy in the contexts of data archives, where an individual’s direct control over their personal data is instead the purview of researchers and data archive staff. This framework conceived of informational privacy as resulting from three other conceptualizations of privacy:
derived by regulating the flow of information, the minimization of harm, and a response to technological risks to privacy. My findings support a further revision of this proposed framework that, instead of modeling informational privacy itself, models the actions taken to promote and ensure informational privacy; actions that reflect the three aspects of privacy, with an overarching management component to coordinate these privacy protection efforts, and recognizing the internal and external factors that affect organizational responses.

My findings indicate that there were three main factors that determined how the data archives responded to the GDPR: whether (if at all) they adopted a processor or controller role for their personal data, national data protection legislation considerations, and the role of a parent or supervisory organization. Another important finding is that the changes that the data archives did make focused on two of the three concepts of privacy in my model: regulating information flows and minimizing harm, as well as a fourth set of changes intended to manage these efforts. I also found that the GDPR did not have a significant, measurable impact on data archive operations such as archiving new data sets or disseminating data to data users. Instead, the biggest impact of the GDPR on the organizations can be found in the data archive staff mindsets: even though all four archives have long-standing traditions of data protection, the GDPR has reinforced the importance of data protection and data subject privacy in all aspects of their work.

This research represents the first empirical study of the GDPR implementation at social science data archives, and provides guidance and lessons learned to research infrastructures in Europe and elsewhere who are adapting to new digital and data privacy legislation and need to balance data subject privacy and the goals of open science.
Chapter 1 Introduction

1.1 Background and problem statement

In the European Union, the United States, and elsewhere, the last decade has witnessed an increased emphasis on data sharing and inter-institutional and cross-border collaborative research studies (Holdren, 2013; Moedas, 2015). In addition to big data sets from social media providers, mobile technologies, and administrative operations that are often used for marketing and business analytics, data are collected at different scales through surveys, longitudinal studies, interviews, and other research methods. These data are then often used for further research aimed at answering a wide range of societal questions. In response to funder, journal publisher, government policy, and institutional requirements, data repositories and archives have increasingly made research and administrative data available for reuse and secondary analysis. These data are valuable resources for researchers conducting comparative and historical research across many disciplines, including the social sciences. Their utility is also maximized when used for additional research in new contexts.

Data repositories and archives have long been sites for the curation and preservation of quantitative and qualitative research data covering a multitude of subject areas. While many began as repositories and archives for analog data, the last several decades have seen a rise in digital access to data, either through direct download of datasets or remote access, such as through virtual private network (VPN) connections or repository-provided online data analysis platforms. In addition to individual institutions providing data storage, curation, and sharing services, there are a growing number of networks of archives that promote collaboration, both
between the archives and repositories themselves, and between researchers at different institutions, even in different countries.

In Europe, the Consortium of European Social Science Data Archives (CESSDA) is one such network for social science archives. The member archives “seek to increase the scientific excellence and efficacy of European research in the social sciences, as well as to expand easy access to data and metadata regardless of borders” (Consortium of European Social Science Data Archives, n.d.b). International data archive and research institution collaboratives also exist in other fields in support of global research, including biology (e.g., ProteomeXchange Consortium, n.d.), environmental science, (e.g., National Environment Research Council, 2020), and the marine sciences (e.g., European Commission, n.d.). Outside of North American and Europe, data repositories and archives are being established to better support researchers in other parts of the world: for example, the West African Vegetation Database (West African Vegetation Database, 2020), the Social Science Japan Data Archive (University of Tokyo, n.d.), and Brazil’s Centro de Documentação e Acervo Digital da Pesquisa (Centro de Documentação e Acervo Digital da Pesquisa, n.d.). The data which institutions such as these preserve and make available to other researchers have many uses: they allow for research finding verification, enable comparative research, reduce the burden on vulnerable populations, and promote collaborative, global research.

The challenge inherent in social science data and the repositories that curate them in particular is their reliance on data about people. Survey, interview, and observation data, as well as other administrative data and documentation, often contain identifiable information about the data subject even after direct identifiers have been removed or anonymized. Depending on the population, even anonymized data may still be re-identifiable (e.g., small populations with
distinct attributes or features, or through linking one or more data sets together for analysis).

Susha, Grönlund, and Van Tulder (2019) identified thirty-five challenges to data reuse, the most cited of which were privacy issues. Consent issues—particularly the consent to have data collected for one purpose and used for another, and by researchers other than the original collectors—are closely associated with data subject privacy concerns and concerns over data ownership (Clark et al., 2018; El Emam & Moher, 2013; Parry & Mauthner, 2004). These concerns have led to legislative efforts around the world to ensure data protection and data subject privacy, one of the most prominent and recent of which is the European Union’s General Data Protection Regulation (GDPR).¹

In April 2016, the Council of the European Union and the European Parliament adopted the GDPR (fully implemented as of May 25, 2018). Applicable to all data about individuals within the European Economic Area,² the GDPR standardizes the legal bases for processing of personal data, consent to data collection, and the rights of data subjects over the data collected about them. These include data which are identified as “special” by Article 9 of the GDPR—data about racial and ethnic origin, political views, religious and philosophical beliefs, membership in trade unions, genetics and biometrics, health, and an individual’s sexual relationships and orientation—and which are desirable for researchers in many social science disciplines and interdisciplinary collaborations. Article 6(4) GDPR also sets out what data controllers³ must consider when evaluating whether a dataset should be made available to new researchers for

² The European Economic Area consists of all European Union Member States, as well as non-European Union members Iceland, Liechtenstein, and Norway. After withdrawing from the European Union, the United Kingdom currently remains a member of the European Economic Area, though its future status is unknown.
³ Article 4(7) GDPR defines “controller” as “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data”
reuse. Universities, research institutions, and data archives were among the organizations obligated to adjust their policies and practices about data collection, documentation, and dissemination. The effects the GDPR has had on the sharing and reuse of data by researchers and data archives in the social sciences within the Europe Union (EU) have not yet been studied. Understanding these effects is vitally important, given that many of the recent new or updated data protection laws outside of the EU have been inspired by or developed in response to GDPR. The EU’s regulation is expected to continue influencing global data protection laws, due to the extraterritorial scope of its requirements (Art. 3(1)).

In 2017, I conducted interviews with United States, Australian, and German data repository staff as part of a larger study of how data reusers were authorized to access data which are sensitive and restricted for secondary analysis and reuse (Levenstein et al., 2018; Tyler, 2020). This research, conducted with the Inter-university Consortium for Political and Social Research, took place during the period between adoption and entry into force of the new GDPR. Many of the repositories and data archives I investigated both within and outside of Europe contain data that fall under GDPR requirements, due to the extended territorial scope of the Regulation (Article 3(2))5. These include: multinational collaborative and comparative studies, national census results, and datasets specifically and strategically stored outside the country of origin which are available for reuse. At the time of those 2017 interviews, no Europe-based respondents were certain about how GDPR would affect either their archive practices or

---

4 These laws include: the 2020 amendments to Japan’s 個人情報の保護に関する法律等の一部を改正する法律案 [Act on the Protection of Personal Information], Law No. 57 of May 30, 2003, as last amended Act No. 44 of Jun. 17, 2020 and Brazil’s newly entered into force data protection law, Lei Geral de Proteção de Dados Pessoais (General Personal Data Protection Law) 13.709, de 14.8.2018. Both the Japanese amendments and the Brazilian law were heavily influenced by the GDPR; as well, Japan has been determined to be “adequate” for EU data transfers under Art. 45 GDPR.

5 A significant contribution of the GDPR is the extension of GDPR requirements and penalties to EU-based data processors and controllers who process data about EU data subjects but who conduct that processing outside of the EU (Article 3(1)). It also applies to non-EU-based data processors and controllers who process data about EU data subjects for the purposes of offering goods or services or who are monitoring the behavior of data subjects within the EU (Article 3(2)).
researchers’ continued willingness and ability to share and reuse data; many were also waiting for their national legislative implementations of the GDPR principles to be determined. That uncertainty motivated my interest in the impact of new privacy regulations on data reuse, and the balance between data subject privacy protection and promotion of data reuse that data archives must maintain.

Despite long-standing social science data archiving practices in parts of the EU, data sharing and data reuse are still relatively new practices across the continent. In a special 2011 issue of the IASSIST Quarterly that presented a baseline understanding of social science data sharing at the European national level, digital curation researchers throughout Europe commented on their countries’ perspectives on and research infrastructure supporting both qualitative and quantitative social science data sharing and archiving (Binder & Filipkowski, 2011; Bishop & Neale, 2011; Butviliénė & Butvilas, 2011; Čižek, 2011; Gárdos & Ivacs, 2011; Gray & O’Carroll, 2011; Henriksen et al., 2011; Kleiner et al., 2011; Kuula, 2011; Medjedović & Witzel, 2011; Neale & Bishop, 2011; Smioski, 2011; Stebe et al., 2011; Titarenko & Tereschenko, 2011). There were significant differences in both research practices, from Poland where “data sharing is not a part of research culture” (Binder & Filipkowski, 2011, p. 71) to Switzerland, where qualitative researchers in particular “are not familiar with the possibilities of secondary data analysis” in their own work (Kleiner et al., 2011, p. 78), and infrastructures, from the lack of archiving practice and official or professional policies in Belarus (Titarenko & Tereschenko, 2011) to the well-developed policies and organizations in support of data archiving in the United Kingdom and Germany (Bishop & Neale, 2011; Medjedović & Witzel, 2011). Despite these disparities in practices, there were similar concerns about the ethics of sharing and reusing these data, particularly related to privacy risks on behalf of the data subjects. These
concerns align with privacy concerns raised by researchers, data subjects, and policy-makers in other disciplines about data sharing and reuse (e.g., Andersson & Sørvik, 2013; Bishop, 2009; Broom et al., 2009; Corti et al., 2000; Curty, 2016; Dallmeier-Tiessen et al., 2014; Law, 2006; Mayer-Schönberger & Cukier, 2014) which will be discussed further in Chapter 2.

Even with concerns over data subject privacy and the ethical reuse of data, there has been a growth of data archives and research collaborations within Europe over the last decade (Bender et al., 2014; Hazeleger et al., 2018). According to the re3data.org registry of data archives, repositories and project-specific databases, there are at least 879 within the European Union Member States (re3data.org, 2022). This registry is not inclusive of every data archive or service in the European Union or elsewhere (for example, it does not include the Malta Data Portal 6 a government data archive, but it does include the Research Data Centre of the German Federal Employment Agency 7), but it is indicative of the growing prevalence of these services in the world. While many of the entries in the registry are project-specific databases, the rest are, for the most part, disciplinary-focused archives. They include open-access archives as well as archives with access restrictions and requirements, supported by research institutions and government bodies. Along with this growth in the availability of data archives and repositories has been the movement towards data protection legislation designed to preserve and promote data subject control over information collected about them, culminating most recently with the GDPR. How these European social science data repositories and archives have responded to, and

---

6 The Malta Data Portal is a government-supported platform provides data access through an open data portal. A fairly new platform, it came online in a Beta form in 2019, and currently contains 205 datasets and 43 registers in the following areas: culture and leisure; police, justice and defence; education, science and technology; health and community care; environment, energy, agriculture and fisheries; work and employment services; transport and communication; inclusion, equality and social welfare; identity, citizenship and immigration; tax and finance; economy, business and trade; and other. For more information, see https://open.data.gov.mt/.

7 The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) is the German government access point for employment micro data specifically for use in empirical research. Data access is provided remotely as well as on-site at one of several locations. For more information, see https://fdz.iab.de/.
are affected by, GDPR has not yet been investigated. Since 2016, commentaries on and evaluations of GDPR’s impact on data reuse for research purposes have focused on biomedical and genetics research (e.g., Custers et al., 2019; Demotes-Mainard et al., 2019; Jacobs & Popma, 2019; Marelli & Testa, 2018; Stockdale et al., 2019; van Veen, 2018).

1.2 Theoretical framework

In this dissertation, I examine the impact of a large-scale attempt to regulate personal data protection in a specific context: four social science data archives within the European Union. Many of the discussions of the need for stronger government protections of data subjects have been grounded in the idea that the more data which are available about people, the greater the risk of privacy violations that potentially cause social, reputational, emotional, or physical harm (C. J. Bennett, 1992). At the same time, however, there has been a growing call to make these data, whether in identifiable, pseudonymized, or anonymized form, more widely accessible through data archiving and data sharing in the interest of transparency and scientific integrity.

The data archives of interest in this study are archives which support the social sciences, collecting and disseminating data about people from a number of disciplines, including economics, sociology, public health, and criminal justice. These data are everything from national census records to surveys and longitudinal studies, including those that are of special populations or about specific social issues like poverty, public health, or social justice. These are all data which are valuable to researchers who want to conduct comparative studies and to reuse these data outside of the original context in which they were collected. However, at the same time, especially for relatively small populations, the risk of re-identification and the harms that can come from that re-identification can be substantial.
The concept of “informational privacy” forms the basis of this dissertation. However, interpretations of privacy and informational privacy frequently center the individual and are not sufficient for an analysis of privacy as resulting from institutional and organizational protection efforts. Therefore, I have reconceptualized informational privacy in research data archives as something achieved and maintained through organizational structures and processes that address three constructs of privacy: the regulation of information flows, minimization of privacy harms, and responses to technological risks to privacy (Fig. 1.1). These three conceptualizations of privacy provide the dimensions through which the effects of the General Data Protection Regulation on data archives can be evaluated. I will briefly describe each dimension below.

Figure 1.1 The relationship between the three approaches to privacy and informational privacy.

Nissenbaum’s theory of contextual integrity posits that there are “context-appropriate” information flows, and that privacy is violated when those flows occur in inappropriate contexts (Barocas & Nissenbaum, 2014; Nissenbaum, 2010). Data archives do not just provide long-term storage of data but also disseminate them. Especially when the data include identifiable or potentially identifiable information, data archives must determine both whether the data provider
has the legal basis or bases to share their research data with the archive, and whether the potential data reuser has met requirements to use the archived data. This theory provides a lens for examining just what the data archives’ interpretations of the GDPR’s requirements mean for those information flows in the context of a third-party repository.

Solove’s (2006, 2008) taxonomy of privacy identifies specific behaviors and activities which lead to privacy harms. Data archive policies, which implement national laws and norms, regulate those behaviors in an attempt to minimize the risk to data subjects. Solove’s taxonomy of potentially harmful behaviors allows for an understanding of how the data archives have identified what and where these may occur, and to identify the data archives’ policies and practices intended to minimize them.

Privacy per Warren and Brandeis (1890), the “right to be let alone,” is promulgated through the laws, regulations, and policies which attempt to maintain it in the face of privacy risks resulting from technological advancement. The GDPR, or at least, a revision of its predecessor, the 1995 Data Protection Directive, was considered to be necessary in part because the Directive was not considered adequate for the speed and volume of data transfers possible in the new millennium. This conceptualization supports an evaluation of whether and how the data archives’ changes in response to the GDPR are recognizing and addressing privacy risks that result from technological advancements and new data processing practices, such as the enhanced ability to aggregate disparate data to identify individuals.

The goal of GPDR was not only to clarify when personal data processing is authorized, it was also designed to give data subjects greater control over the data collected about them. Therefore, in this dissertation, I drew not just on the implementations of legal requirements embedded in GDPR, but also the three conceptualizations of privacy as important lenses through
which to examine this problem. What Nissenbaum proposes in her theory about the contextual nature of the data transmission (i.e., the data sharing and reuse) highlights the challenge of privacy, particularly that privacy is not just this singular monolithic concept. Just how privacy is conceived of is different depending on the context it is experienced in, and that includes the context in which the original data were gathered, and the context in which they were aggregated, shared, and reanalyzed. This also means that the privacy harms that may come from the harmful behaviors Solove (2006, 2008) discusses are also not monolithic. Citron and Solove (2022) have lately presented a typology of privacy harms—physical, reputational, economic, relationship, discrimination, psychological, and autonomy (which additions subtypes of the final two)—which allows for further recognition and understanding of the importance of data protection and privacy.

The harms that can come from privacy violations, particularly in terms of the reuse of social science data, are not universally applicable. Some uses and transmissions of these data are more potentially harmful than others, and when looking at data reuse within data archives, we must be concerned about the changing harms and the changing contexts in which harms can happen through technological advancements. For example, a data subject who responds to a survey about her sexual orientation on the assumption that those responses will be anonymized during their specific processing by the original research team, to be used only by them, may feel that her privacy has been violated if those data, in identifiable form or not, are shared through a secure access data archive. The secure remote access is a new technology in a new context than what her original expectation of privacy was based on. Depending on what restrictions may be in place about anonymization, data linking, and other processing tools, and how well the staff working with the data have been trained, the actual risk of privacy harm varies, but her
perception of a privacy risk is increased because she was not aware that her data would be used elsewhere. The three conceptualizations of privacy thus are applicable to studying the GDPR and allow me to trace how information flows from data collection to the repositories to the reusers, and the constraints that are placed on them, affect informational privacy protection at social science data archives.

While I used this framework to structure my data analysis, I also tested the effectiveness of this model of informational privacy in reflecting how informational privacy is conceived of and enacted at my four research sites. As I will discuss in Section 5.3, during the data collection and data analysis stages, I determined that this model is more appropriately a model of informational privacy behaviors at the four data archives, not a model of informational privacy itself. In the revised model, the three dimensions—data archive behaviors intended to regulate information flows, minimize harms, and respond to technological risks to privacy—together do support the maintenance of data subject informational privacy. However, there is an additional fourth aspect intended to manage these behaviors, represented by the creation of new staff positions with personal data protection management and coordination as a primary focus. The revised model also recognizes that there are internal and external factors that may influence those behaviors.

1.3 Research questions

To address what impacts and outcomes the GDPR and the GDPR transition process had on my four data archives, I designed my study around two broad research questions:

How do the ways in which General Data Protection Regulation has been enacted increase our understanding about approaches to privacy in data archives?
How has the full adoption of the General Data Protection Regulation affected data sharing and subsequent data reuse at four European Union data archives?

To further address these guiding questions, I have six subsidiary research questions:

1. What factors determined the four social science data archives’ responses to the General Data Protection Regulation?
2. What do the different archive implementations of the General Data Protection Regulation indicate about what aspects of privacy are prioritized?
3. What are the factors motivating the archives’ data protection policy and practice adaptations to the General Data Protection Regulation?
4. What are the mechanisms used at the data archives to address their stated goals of General Data Protection Regulation implementation?
5. In what ways have the different data archives’ policies and practices converged and diverged as a result of the General Data Protection Regulation?
6. How have the data archives conceptualized “contextually appropriate” information flows in response to the General Data Protection Regulation requirements?

1.4 Overview of research design

By answering these research questions, my dissertation intends to address the gaps in current research identified in Chapter 2—the insufficiency of current conceptualizations of informational privacy in the context of data archives, and a lack of research into the operational impacts of the GDPR on social science data archives—by investigating the actions taken, challenges faced, and outcomes experienced during the GDPR transition period of 2016-2020. To do so, I conducted a qualitative, multi-case comparative case study of four European Union-based social science data archives: the Czech Social Science Data Archive (ČSDA), the Finnish Social Science Data Archive (FSD), Data Archiving and Networked Services (DANS), and GESIS — Leibniz Institute for the Social Sciences (GESIS). I had three sources of data. First, I conducted in-depth semi-structured interviews with data archive leadership, IT staff, legal advisors, data archive staff, and external archive stakeholders (Table 3.2). I also reviewed data
archive policy documents, legal agreements, data management guidance, data processing
guidance, and other documentation, and where available, compared the post-GDPR
documentation with their pre-GDPR counterparts, to identify how changes discussed by the
interviewees were implemented. Finally, I analyzed data archive usage statistics—records of
annual data set deposits or publications, data download and/or usage statistics, annual data
archive user registrations, and requests for access to restricted or secure data—as a third method
to measure the operational impacts of the GDPR and the changes at the data archives.

1.5 Significance

Over the last decade and more, there has been a growing push from research funders,
journal publishers, research institutions, and governments for researchers to make their research
data available for reuse. There is also an increased emphasis on collaboration on a multi-national
scale, especially in the wake of the COVID-19 pandemic (e.g., Kim & Cho, 2021; Maher & Van
Noorden, 2021; National Science Foundation, 2020) which has seen the “normal” modes of
research collaboration disrupted by travel restrictions, work-from-home policies, and the impact
of the pandemic on family life. It is a global research community that, in the years before the
pandemic began, had already had to respond to and understand another disruption: the
implementation of the new European Union-wide GDPR. As will be discussed in Chapters 2 and
4, there have been many discussions about possible implications of the GDPR on research and
research infrastructures across many disciplines, including the social sciences; however, in the
years since the GDPR entered into force in 2018, no formal study has been made of just what
happened to and within data archives for the social sciences.

This dissertation furthers the research community’s knowledge about the impact and
outcomes of data protection legislation on research infrastructures in three ways: 1) it provides a
new theoretical framework to guide organizational changes aimed at protecting and promoting data subject informational privacy; 2) it contributes to the literature about the impact of the GDPR on research, and specifically responds to the lack of social science data archive-focused literature; and 3) it offers lessons learned and guidance to data archives and repositories in general considering similar changes or addressing similar laws (Section 5.4).

1.6 Structure

This dissertation is organized into five chapters. In Chapter 1, I have introduced the background and motivation for my study. I have presented the theoretical framework, research questions and methods, and discussed the significance of this work.

In Chapter 2, I review the literature underpinning the relevant aspects of my study: conceptualizations of informational privacy, European Union data protection legislation, data archiving and organizational change, and present my theoretical framework. Through this chapter, I identify gaps in the current literature about data protection and privacy in the context of research data. In particular, the need for research into how data protection legislation like the GDPR and its national implementations have led to changes in social science data archiving practices, and how those changes have affected the use of social science data archives by data providers and data users.

In Chapter 3, I describe my research methods, including the data collection and data analysis methods, sample selection, and limitations.

In Chapter 4, I present my findings in the form of four case reports, one for each of my field sites, and end with a comparison of the experiences of the four sites. I present the three significant findings of this dissertation. First, my findings also indicate that there were three factors that were most influential in how the data archives experienced the GDPR transition
process: what processor or controller role, if any, the data archives adopted for which personal data, the national or subnational legislation, and the presence of a parent or supervisory organization. Second, that the practices of the four social science data archives are converging in line with the GDPR’s goal for data protection harmonization, and that the changes made at each archive primarily address two of the privacy dimensions of my theoretical framework: privacy as regulating information flows, and privacy as minimizing harm. Lastly, the GDPR did not result in any measurable changes to use of the data archives by data providers or data users; instead, the most significant impacts were in the data archive staff mindsets and certain policy and process changes.

In Chapter 5, I discuss the implications of my findings and evaluate the efficacy of the model of informational privacy in the context of the work of social science data archives. I summarize the contributions and significance of this research and conclude with a discussion of possible future work.
Chapter 2 Literature Review

In this literature review, I bring together scholarship from diverse research areas to provide the foundation for a study of the impact of legislated privacy and data protection on the sharing and reuse of data. I start by defining key terms that will be used in this dissertation. In section 2.2, I present the challenge of defining privacy, and review the components and types of privacy which have been influential in the development and spread of data protection efforts. In particular, I identify the three conceptualizations of privacy that structure my evaluation of the General Data Protection Regulation (GDPR): privacy by regulating information flows, privacy as minimizing harm, and privacy in response to technological risks. These three conceptualizations provide three dimensions along which GDPR can be evaluated, as GDPR’s goals are a response to different types of privacy violations related to technology, protection of data subjects, and controlling the flow of information. Differing values toward and interpretations of privacy have characterized how different nations have undertaken data protection efforts, leading to disharmony that has restricted the ability to share data. I review these perspectives to better understand the importance and relevance of the General Data Protection Regulation as a response to the potential for privacy violations.

In section 2.3, I discuss the GDPR and its intended goals. To properly situate the GDPR, I also describe the historical motivations for data protection in Europe, as well as prior efforts to regulate privacy and data protection. Finally, in section 2.4 I discuss the current state of data archiving, research related to the GDPR and organizational change, and the privacy risks at data
archives. I conclude this chapter with the theoretical framework and a discussion of the gaps in the literature.

2.1 Definitions

In this section, I define several key terms and phrases used in this dissertation.

**Repository or archive:** The “system and set of services designed as an archive for digital data with context, fixity, and persistence” (Kowalczyk & Shankar, 2011, p. 261).

**Secondary data:** Data which have been collected for one purpose that are then used to fulfill another purpose (Faniel & Jacobsen, 2010).

**Data reuse:** The processing of data which have been collected for one purpose to answer new questions, replicate or validate previous results, in a new or same context (Law, 2006; N. Moore, 2007; Sun & Khoo, 2017; Zimmerman, 2008).

**Data sharing:** The “conscious decision to make a new [data] resource available” for reuse by other people, through physical or electronic data object transfer or deposit (Sumner & Dawe, 2001, p. 418).

**Data acquisition:** The “process of adding to the holdings of a records center or archives by transfer under an established and legally based procedure, by deposit, purchase, gift, or bequest” (International Council on Archives, 2004).

**Data set publication:** In this dissertation, data set publication is considered the process of adding a curated data set and associated metadata to a data archive’s publicly viewable and accessible data catalog.

**Data processing:** As defined in GDPR Art. 4(2). Any operation performed on data, including the “collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.”

**Processor:** As defined in GDPR Art. 4(8). A “natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.”

**Controller:** As defined in GDPR Art. 4(7). The “natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.”
**Personal data:** As defined in GDPR Art. 4(1). Any “information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.”

**Special categories of personal data:** As defined in GDPR Art. 9(1). Any “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.”

### 2.2 Privacy conceptualizations

As will be discussed in section 2.3.5, the GDPR is intended to protect personal data, and an expected outcome of that protection would be that privacy is better protected. As Dove reminds us,

> data protection and privacy are related but nevertheless distinct concepts. They should be seen as working together, especially with data protection seen as helping to ensure our freedoms and dignity. … The aims of data protection, then, are broader than simply privacy protection, but at the same time, data protection is a crucial tool to ensure our privacy (Dove, 2018, p. 1014).

Despite this, scholarly discussions and criticisms of the GDPR (Kadlecová & Weiss, 2021), as well as public perceptions of it (including the research participants in this dissertation), frequently frame their observations and comments around the concept of privacy protection. Therefore, we must develop an understanding of privacy itself to understand the ways in which data archives, and archive staff, have approached the GDPR requirements. However, just what “privacy” *is* is a complex concept to which “it is not possible to give a single, unitary definition that covers the many diverse privacy interests” (DeCew, 1997, p. 4). The challenge of understanding what privacy *is* is not just a matter of developing a proper definition; privacy, and what actions and information are considered ‘private’ are themselves culturally defined. This
has led to distinctly different approaches to privacy and data protection regulations between the United States and the European Union since the 1960s, when technological advances began to cause greater concerns about the risk of privacy violations. The approaches taken in the United States, according to Whitman, focus on liberty and freedom from intrusion, while in Europe, the “privacy protections are, at their core, a form of protection of a right to respect and personal dignity” (Whitman, 2004, p. 1161). While this dissertation focuses on a European Union created and implemented regulation, the extra-territorial application of the GDPR has and will continue to influence further development of data protection and privacy legislation outside the European Union (even where European views of privacy (broadly speaking) may not apply). For this reason, it is useful to discuss the different interpretations of privacy, how they have developed out of technological advances, and how they have led to different legislative and judicial implementations.

2.2.1 Informational privacy interpretations

In political and institutional policy debates over what privacy is, one constant is that it is hard to define, or, as Bennett describes it, “notoriously vague” (1992, p. 12). What behaviors and information are considered private, and therefore worthy of protection, are often culturally constructed and differ from between countries or groups, in everything from architectural design (Abu-Gazzeh, 1995; Ozaki, 2002), disclosure of private information to social networks, both in-person and online (Krasnova et al., 2012; Vasalou et al., 2010), or even just what information is considered to be private (Van Der Horst & Messing, 2006; Whitman, 2004). Attempts to define privacy in western literature tend to fall into one or more broad categories of privacy: as based on the context of the information flow (Barocas & Nissenbaum, 2014; Nissenbaum, 2010; Tavani, 2008), as based on the desire to be left alone and to control access (Gavison, 1992;
Reiman, 1995; Warren & Brandeis, 1890), as based on specific actions that cause harm (Solove, 2006, 2008), and as states of being (Westin, 1970). Many privacy scholars, instead of attempting to define a singular overarching ‘privacy’, have instead conceived of different types of privacy that should be protected. These types of privacy include genetic (Haeusermann et al., 2018; Kaye, 2012), physical (Allen, 2011; Floridi, 2013, 2014; Gavison, 1980; Tavani, 2008), decisional (Floridi, 2013, 2014; Tavani, 2008), locational (Allen, 2011), and racial (Allen, 2011).

Two types of privacy that are particularly important for this dissertation are information(al) privacy (Allen, 2011; Clarke, 2016; DeCew, 1997; Hartlev, 2004; Junge, 2018; Mai, 2016; Regan, 1995; Stockdale et al., 2019; Tavani, 2008; Westin, 1970) and the right to dignity, a view of privacy that underlies many European, especially German, views of privacy (Espindola, 2011; Gavison, 1992; van Deursen & Kummeling, 2019; Warren & Brandeis, 1890; Whitman, 2004). Some scholars, such as Floridi (2013, 2014), view privacy in terms of ‘freedom from’ interference, while others conceive of types of privacy in terms of the risk and harms of their violation (Citron & Solove, 2022; Solove, 2008; Warren & Brandeis, 1890), ideas that interweave through informational privacy and the right to dignity.

Floridi (2013) identifies four types of privacy that are grounded in and linked by the concept of privacy as the personal, individual freedom from interference in various contexts of existing. One of these is information privacy, which he defines as the “freedom from epistemic interference… a restriction on facts about [someone] that are unknown or unknowable” (p. 230). Within information privacy, he further defines an information gap as “a function of the degree of accessibility of personal data” (p. 231). This gap is what is bridged when information is accessed, and the ability of individuals (whether mediated by technology or other protections) to control that access is often the focus of regulatory attempts to enforce privacy. Drawing on this
view of information privacy, Floridi (2013) thus sees the violation of that privacy as “a violation of the nature of the informational self” (p. 260), not a violation based on ownership, rules, or even personal rights.

This perception contrasts to a degree with other conceptions of privacy which are grounded in the harms that come from privacy violations. Warren and Brandeis (1890) addressed the new harms and new potential sources for harms that came with technological advancement, and saw their right to be let alone as the way in which individuals restrict physical intrusions into the private sphere. In the United States, this private sphere generally means the home, where Americans have “private sovereignty within [their] own walls” (Whitman, 2004, p. 1162); in continental Europe, the private sphere reflects “a form of protection of a right to respect and personal dignity … the right to control the sorts of information about oneself” [emphasis in original] (Whitman, 2004, p. 1161). Solove’s (2008) taxonomy of privacy, discussed further in Section 2.2.1.2, orients privacy in terms of the types of activities that cause “problems that affect a private matter or activity” (p. 102). Ohm (2009) argues that privacy within different sectors should be approached in terms of the risks inherent in those sectors, though there should be a baseline of protection that all sectors should adhere to.

While there are a number of related definitions, a common definition of informational privacy is provided by Clarke (2016): “the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves.” It closely corresponds with communications privacy, where the data subject’s feeling of having lost control over both the use and sharing of information about themselves is closely associated with the technological ability to share data without the subject’s knowledge (Regan, 1995), and with Westin’s definition of privacy itself: “the claim of individuals, groups, or institutions to determine for themselves when,
how, and to what extent information about them is communicated to others” (Westin, 1970, p. 7). This was already a concern in the 1970s, as seen in the introduction of data protection laws beginning in 1970 with the Hessisches Datenschutz (see Section 2.3.3), and it continues to pose a privacy infringement risk, indeed an increased risk, today (Junge, 2018; Westin, 1970). Hartlev (2004) claims that the European Union’s 1995 Data Protection Directive was intended to address and protect specifically an individual’s informational privacy.

This idea of information privacy as a person’s ability to control access to their personal information is shared by others. For example, Tavani (2008), Mai (2016), and others discuss two influential and interrelated conceptual approaches to informational privacy: restricting access theory and control theory. The first proposes that an individual has informational privacy “when she is able to limit or restrict others from access to [personal information]” (Tavani, 2008, p. 141). Gavison (1980) similarly described privacy along these lines, as the “limitation of others’ access to an individual” (p. 428). Moor (1990) calls restricted access the “most defensible” conceptualization of privacy, and further argues that “an individual or group has privacy in a situation if and only if in that situation the individual or group or information related to the individual or group is protected from intrusion, observation, and surveillance by others” (p. 76). Here, the “situation” is used in place of zones or contexts intentionally, to indicate the range of affairs, events, and behaviors one would expect to be considered private: e.g., a particular location such as the home, specific relationships in which privacy is expected (spousal privilege, doctor-patient, lawyer-clients, etc.), or the storage of information in protected or restricted access databases. The focus of the restricted access theory is on the idea that there exists a zone, context, or situation wherein privacy is expected, established through isolation, barriers, or
agreement not to transmit information, that also forms some basis of protection within those zones.

These zones could be natural, where the individuals are “because of the circumstances of the situation, are naturally protected from intrusion or information-gathering,” or normative, where “the protection may be natural [based on the situation] but is essentially legal or moral” (Moor, 1990, p. 77). It is these two types of zones that account for different cultural ideas of what behaviors, acts, and information are considered private, and accounts for the fact that not every situation in which information is gathered about individuals is a violation of privacy. In France, for example, there is a protected “sphere of privacy,” within which are categories of information that an individual has the right to expect respect towards: they include “family life, sexual activity and orientation, illness and death, and even private repose and leisure” (Hauch, 1993, p. 1246). In the Netherlands, the window of a home provides an unblocked view into the private sphere of the home, though there are social norms around invading that private sphere (Van Der Horst & Messing, 2006; Vera, 1989).

This view of privacy is compatible with other types of privacy, such as physical privacy, but also is important for conceptualizing privacy from and through information technology. Restricted access networks, password protection, encrypted data transfers, and other tools are viewed as privacy enhancing because they can create or support those zones of privacy. In so doing, they minimize the risk of unauthorized access to information that ought to remain private. However, this view of privacy has been criticized for ignoring or underestimating the role of choice and control in restricting access to information about the individual (Junge, 2018; Tavani, 2008). Every instance of information dissemination is thus considered to be a privacy violation, even though as individuals we do see a difference between choosing to tell a trusted friend
private information and having that information captured through surveillance and broadcast to the world. It also does not distinguish effectively between what zones and contexts are private or public, even though this differentiation is what the theory relies on (Elgesem, 1999).

The second conceptualization is similarly interested in the role of access in informational privacy, but in terms of the choices the data subject has in determining whether others are able to access the information. Where restricted access theory has been criticized for ignoring individual choice over who and when to give access to information, control theory “characterizes privacy as a form of control over personal information” (Junge, 2018), where the individual actively chooses whether access to themselves (and access to information about themselves) is granted to another (Mai, 2016; Tavani, 2008). This is a frequent conceptualization of privacy, and the implications of it will be discussed below. A common response to the question “what is privacy?” is, as McCreary (2008) describes, “partly a form of self-possession—custody of the facts of one’s life, from strings of digits to tastes and preferences. Matters of personal health and finance… are in most instances nobody’s business but our own—unless we decide otherwise.” This is the understanding of privacy described by the data subjects in a number of studies investigating the willingness to share data. Hondius (1980) found that public anxiety over processing of their data was a result of their exclusion from data handling decisions, while Damschroder et al. (2007) found similar opinions held by Veterans Affairs hospital patients. In the latter example, research participants were willing to share their medical records with researchers as long as the research had societal impact, they could be assured of their confidentiality, and they had some measure of control over how their records were used.

Informational privacy is, according to Allen (2011), what is disturbed when personal information which subjects represented in the data wanted to keep secret or unidentifiable are
shared. In the modern era, informational privacy is increasingly infringed upon by the technological advancements that allow a large volume of data to be collected, processed, and shared quickly. Tavani (2008) and Junge (2018) both identified that volume and speed, as well as the duration at which data are stored by public and private entities and the type of data which can be collected and processed, as the sources of exacerbated informational privacy risks. Therefore, protection of informational privacy “shields individuals from intrusions as well as fear of threats of intrusions, and it also affords individuals control in deciding who has access…and for what purpose” (DeCew, 1997, p. 75).

Informational privacy is interwoven into the conceptualizations of privacy that form the dimensions along which I will study data archives after the GDPR. One of the fundamental bases for the importance of consent in research is the respect for human dignity and autonomy (Perez Vallejos et al., 2019). Hartlev (2004) sees a relationship between informational privacy and the rights of dignity, integrity, and autonomy; specifically, the individual ability to defend their dignity while also retaining the autonomy to decide when to give up their privacy by sharing their information. In this view of privacy as a matter of concern about dignity, personal information is considered to be sensitive in a given context because of the potential harms to an individual’s welfare or dignity (Allen, 2011). Dignity-based privacy is what Solove also sees as at risk due to exposure: that exposure “[strips] people of their dignity” (2006, p. 537).

This respect for dignity is integral for many of the European national prioritizations of privacy. Violations of human dignity in Germany and the former Czechoslovakia are the foundations of the inclusions of specific rights to privacy in their national constitutions. The right to dignity is the foundation of the Grundgesetz für die Bundesrepublik Deutschland (the
German constitution, also known as the “Basic Law”\textsuperscript{8}), embodied in Article 1(1).\textsuperscript{9} The Charter of Fundamental Rights and Basic Freedoms (Listina základních práv a svobod) (LZPS)\textsuperscript{10} of the Czech Republic similarly declares that “all people are free and equal in their dignity and rights” (Art. 1), with the specific guarantee of privacy given in Article 7.\textsuperscript{11} This exemplifies a key difference in how privacy is conceived of in the United States and in parts of continental Europe. To return again to Whitman (2004), “[c]ontinental [i.e., European] privacy protections are, at their core, a form of protection of a right to respect and personal dignity” [emphasis in original] (p. 1161), while in the United States, privacy is instead concerned with liberty and freedom from intrusion. The ideals of this “continental” view of privacy are further formalized through the GDPR, where “damage to reputation” is identified as a potential result of data processing, because of the damage to the rights and freedoms of natural persons (Preamble para. 75), and personal data breaches (Preamble para. 85).

2.2.1.1 Theory of contextual integrity

The theory of contextual integrity developed by Nissenbaum understands privacy as “the requirement that information about people (‘personal information’) flows appropriately, where appropriateness means in accordance with informational norms” (Barocas & Nissenbaum, 2014,


\textsuperscript{9} Grundgesetz für die Bundesrepublik Deutschland, Artikel 1 Abs. 1: „Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.“ Translation from the Bundesministerium der Justiz und für Verbraucherschutz: “Human dignity shall be inviolable. To respect and protect it shall be the duty of all state authority.”


\textsuperscript{11} Article 7(1), LZPS: “The inviolability of the person and of her privacy is guaranteed. They may be limited only in cases provided for by law.”
Unlike other approaches to privacy which rely on privacy protection by means of access limitation or maintaining control over the information in question, the theory of contextual integrity situates the personal control of information as only one of the possible constraints on information flow, all of which are context dependent (Nissenbaum, 2010; Tavani, 2008). This theory thus frames privacy not as an issue of the fear of violation, but as representing a transgression of contextually-relative informational norms (Nissenbaum, 2010). When violations occur, they occur because the information sharing was in violation of specific social norms of intimacy and disclosure (Perez Vallejos et al., 2019).

This is an important interpretation of privacy when considering the GDPR. There is a two-part goal of the GDPR: “to ensure a consistent and high level of protection of natural persons and to remove the obstacles to flows of personal data within the Union” (Preamble (10)), which can be seen as a paradox, particularly for research data archives. How do you ensure a person’s privacy and protection if/while also sharing the data about them? Nissenbaum (2010) builds on the idea that, within societies, there are culturally and socially constructed norms of what is considered to be private (also discussed by Westin (1970) when considering the existence of norms of privacy that do not align with American cultural norms). Therefore, by considering the “right to privacy” to be a “right to context-appropriate [information] flows” (Nissenbaum, 2010, p. 187), it allows for an approach to privacy and data protection like the GDPR, with its paradoxical approach to data protection and data sharing, and its moments of flexibility enabling Member States to adapt their requirements to fit their own national needs. This conceptualization of privacy enables me to understand the data archive’s privacy protection behaviors and actions assuring that the data flows into and out of the archive are contextually appropriate, and to identify what those contexts are.
2.2.1.2 Privacy as protection from harm

The GDPR’s requirements are applicable to personal data and to what it refers to as “special categories of data” in accordance with Article 9(1) (see definitions in Section 2.1). The types of data herein are explicitly identified by GDPR’s Article 9(1) and other privacy proponents due to historical misuse of these data (e.g., the Holocaust, the Rwandan genocide, the World War II internment of Japanese-Americans) as well as more modern data privacy violations (e.g., the Target pregnancy scandal, Soviet and Communist surveillance programs, anonymization failures at AOL and Netflix). What connects these examples is the risk of harm to the individual when these data are accessed inappropriately, immorally, or illegally, and then misused: harm that can be reputational, emotional, and in the most serious instances, fatal.

Solove (2006, 2008) classified four basic groups of harmful activities that can lead to privacy harms: collection (the act of data collection, surveillance, interrogation); processing (aggregation, identification, insecurity, secondary use, exclusion); dissemination (confidentiality breach, disclosure, exposure, increased access, blackmail, appropriation, distortion); and, invasion (intrusion, decisional interference). While there are clear connections between the collection, processing, and dissemination harms, as Solove defined it, the “invasion” harms are different though related because information is not always involved. Intrusion causes harm because it “interrupts one’s activities through the unwanted presence or activities of another person” (Solove, 2006, p. 553). Decisional interference, on the other hand, refers specifically to “governmental interference with people’s decisions regarding certain matters of their lives” (Solove, 2006, p. 558). Solove saw decisional interference in particular as being linked to many of the other activities that could harm informational privacy. This list of data-related activities is
effectively all inclusive, covering the processing activities that GDPR is interested in regulating, as delineated in Article 4(2) (see definition in Section 2.1).

Citron and Solove (2022) have addressed the difficulty in identifying specific privacy harms (which can result from many different harmful activities, alone or in concert) by presenting a new typology of privacy harms, many of which reflect the damages identified in the GDPR:

Typology of privacy harms: physical harms, economic harms, reputational harms, psychological harms (including the subtypes of emotional distress and disturbance), autonomy harms (including the subtypes of coercion, manipulation, failure to inform, thwarted expectations, lack of control, and chilling effects), discrimination harms, and relationship harms (Citron & Solove, 2022).

By understanding privacy in terms of the consequences of its violations, Solove (2006, 2008) and Citron and Solove (2022) take a pluralistic approach to privacy. In doing so, not only can we better identify potential sources of privacy violations, we are forced to recognize that there are multiple types of privacy that are thus endangered, such as those discussed above, and potential harms that may result. The GDPR focuses on informational privacy violations, but it recognizes
that there are a multitude of harms that come from them, and that violating informational privacy can lead to the violation of other privacies, and thus other harms. This conceptualization of privacy enables me to understand and identify data archive behaviors intended to minimize or prevent just such privacy harms.

2.2.1.3 Privacy in response to technological risks

In the discussion of privacy law, the foundational piece is Warren and Brandeis’ (1890) article “The right to privacy” in the *Harvard Law Review*. This paper is often cited as the first explicit American declaration that a “right to privacy” actually exists. In their discussion, they define privacy ultimately as “the right of the individual to be let alone” (Warren & Brandeis, 1890, p. 205) or, as it has been conceptualized, the right to restrict physical intrusion into the private sphere. Their interest in the question of whether privacy rights exist stem from, as the narrative goes, Warren’s anger at the publication of “intimate” or “private” information about his family’s home life in newspapers (Glancy, 1979). The use of the media, which can be preserved and replicated over time, to share these intimate details, paired with technologies that captured instantaneous records in the form of photographs, raised questions about privacy and just what rights existed in the United States at the time to protect it in the new technological age (Warren & Brandeis, 1890).

The original debate focused on the ability to capture information and share it at a greater scale than ever before. This may seem familiar to modern readers from the early 21st century European debates over the “right to be forgotten” (e.g., Kodde, 2016; Kostov & Scheckner, 2017; Padova, 2019; Post, 2018; Powles, 2015), particularly as these modern concerns refer not just to the sharing of intimate information but the ability to preserve it seemingly in perpetuity. However, these two conceptions of privacy are distinct; in fact, American organizations and
companies have criticized the implications that such a right to be forgotten would interfere with their business needs (S. C. Bennett, 2012) because of its focus on where information is stored, and not on the source of information. What the two perspectives share, however, is that both views of privacy have been offered in response to technological advancement (the specific privacy concerns from technology will be discussed in detail in Section 2.2.2). Others, such as Tavani (2008), have criticized their definition for conflating “privacy” and “being let alone,” which they argue should instead be considered two concepts, where “being let alone” is a specific type of privacy: physical.

The events that led to the Warren and Brandeis article resulted from the invention of instantaneous photography, which no longer required subjects to remain still for extended periods of time, and a burgeoning journalistic practice that called for intrusions into private communications and relationships and the sharing of such private and intimate information thus acquired. Their argument was that a right to be let alone exists even in the new technological environment, though they acknowledged that there were practical limitations to that right of privacy. More than a century later, Gavison (1992) argued that Warren and Brandeis’ analysis is still valid, and applicable even beyond the United States. For example, just as Warren and Brandeis recognized 19th century French laws on privacy as a basis on which to develop their argument, French law in the second half of the 20th century built again on their idea of “being let alone.” There, the right to respect of private lives and personal privacy is tied tightly to autonomy and freedom from intrusion as much as it is to secrecy of information.

The laws about privacy and data protection that have been implemented in the United States and in Europe over the last century, including the GDPR, have primarily been reactive. Technologies are developed and used, from computing to surveillance to communications
technologies, only for users and regulators to later realize that they were potentially or actually damaging to the privacy and reputation of data subjects. The original laws passed, as will be discussed in Section 2.3.3, responded first to the collection and processing of personal data by local and national governments. Over time, those protections would be extended to other sectors through omnibus privacy laws in the European Union. Data archives were among the institutions responding to the new legal requirements, implementing policies which prioritized data subject privacy and data protection within the archive. However, data archives also preserve data for further dissemination and processing; thus, they must balance data protection obligations with that mission in ways that other organizations do not. The conceptualization of privacy as the right to be let alone, which I refer to here as “privacy in response to technological risks,” in the contexts of GDPR and social science data archives enables me to understand how privacy is both protected by and at risk through technology, and how the data archives have identified and addressed these risks through their internal policies and practices.

2.2.1.4 Theoretical framework

The General Data Protection Regulation is grounded in the idea that it is possible to respect and protect a data subject’s informational privacy while also accommodating and facilitating the use of those data in the wider society. This section has reviewed the complex inter-weavings and challenges of the multitude of approaches to the concept of “privacy” leading to the three conceptual approaches I will draw on in this dissertation. One conceptualization of privacy is not sufficient to evaluate the outcome of data privacy laws, because the laws and organizational responses do not address only one aspect of privacy, nor do they do so in isolation. Nor can a single conceptualization centered entirely on the data subject adequately frame informational privacy in a context where that data sharing further and further from the data
subject is expected. Data subjects find themselves in a similar position with data archives: there are data collected about them, stored in physical and/or digital form in distant data files, over which they have limited control when it comes to who, how, when, and why their data is accessed and processed. Data archive staff instead establish policies and practices designed to maintain and promote data subject informational privacy.

![Figure 2.1](image)

**Figure 2.1** The relationship between the three approaches to privacy and informational privacy.

Therefore, conceptualizations of informational privacy centered on the data subject’s own control over their personal information are not sufficient for a study of informational privacy in the context of data archives. Understanding how a law protects data subjects from harmful data activities requires not just looking at the potential harmful outcomes, but also at what and how technologies have enabled both the activities and the harm, and how the information flows in one context are harmful but are not in another context. The new framework of informational privacy (Fig. 2.1) considers informational privacy as resulting from the data archive’s efforts to regulate the flow of data into and out of the data archive, to minimize harm to the data subjects, and in response to technological risks to privacy. It also allows for potential relationships between each
dimension; for example, when a process or policy is in place that addresses more than one dimension. I used these conceptualizations to structure my analysis of the three data archives in Chapter 4. In so doing, I also tested the utility of this framework to understand informational privacy; this will be discussed further in Section 5.3.

2.2.2 Privacy and technological advancement

As previously discussed, advances in technology led to the first American legal discussion over the existence of a right to privacy. However, in many ways, this discussion remained primarily within legal circles and in response to the so-called “yellow journalism.” Privacy entered wide-spread public consciousness in the 1960’s with the development of technologies that enabled greater surveillance capabilities and data processing, particularly by government authorities (Regan, 1995; Schneier, 2015; Westin, 1970). The computing technology then was nowhere near as powerful or advanced as today, but it was a continuation of technology developed before World War II which was used, in addition to administrative and business data analyses, to carry out census tabulations and population statistics; technology that would later be used to carry out the European-wide Holocaust (Black, 2001; Ketelaar, 2020).

By the 1980’s, after a number of regional and national data protection laws were implemented, there was growing disagreement over the focus of those laws. One of the leading motivations for those laws, particularly in Europe, was to ensure that the Nazi abuse of personal data could not be repeated elsewhere. But in the 1980’s, there was growing concern that the focus of the laws was too heavily on the technology and not enough on the idea of privacy as an independent concept that needed protection (Flaherty, 1984) (a concern and criticism that remains even today (e.g., Dove, 2018; Koops, 2014; Wachter, 2019; Waldman, 2021)). The problem with focusing on the technology is that it reinforces the idea that privacy is only
important in the way technology violates it and ignores that privacy violations occur in other contexts as well. It is a limited perspective, and it also presumes that if the technological issues are solved, then privacy will automatically be protected, ignoring the human component. The laws were often reactive, responding to privacy risks and harms only after the harm had occurred, instead of proactively identifying the potential risks while the technologies were developed or shared. Despite these concerns, however, the privacy legislation passed around the world in the following decades focused primarily on information privacy as experienced through the development of technologies intended for sharing and use of data.

Westin (1970) argued that, in addition to the development of surveillance technology (described in his book in a manner reminiscent of a James Bond film), the development of recordkeeping, data processing capacity, and the computerization of business and government created new privacy risks while exacerbating old ones. Thirty years later, the growing influence and accessibility of e-commerce and electronic communications raised serious concerns not just from shady figures listening at the door or digging through trash, but in the digital realm (Stratford & Stratford, 1999). It was not, however, simply a matter of online business and banking platforms capturing individuals’ financial information, or hackable email platforms that increased the risk of accidentally sharing sensitive information.

It is the advanced capabilities to gather, store, and access information (Duke & Porter, 2013; Floridi, 2014; Junge, 2018; Lachaud, 2018; Nissenbaum, 2010; Quinn, 2018; Westin, 1970) as much as it is the capabilities for linking and processing at scale (C. J. Bennett, 2018; DeCew, 1997; Lachaud, 2018; Lievens & Verdoodt, 2018; Mayer-Schönberger & Padova, 2016; Nissenbaum, 2010; Riccardi, 1983; Vestoso, 2018; Westin, 1970). The advancements in the last thirty years occurred far more quickly and to a much greater degree than had been predicted by
the drafters of 1970s-era legislation, much less the authors of the 1995 Data Protection Directive. The Directive, written and implemented before the era of Google, Facebook, and smart devices, had significant shortcomings by the early 21st century (Junge, 2018; Lachaud, 2018; Lindqvist, 2018; Mayer-Schönberger & Padova, 2016). New types of personal data are also available now for processing that were not reflected in the original text of the Directive—e.g., social media data, genetics data, biometrics data—though later implementations of the Directive from new EU Member States in the 2000s included them. Gavison’s (1980) argument that, as technology advances, it is increasingly difficult to protect privacy to the same extent, still holds true today. Each new technological advancement brings new privacy risks for the misuse and abuse or personal data.

However, it is not just the technology that has changed and given rise to new privacy risks. The data used with these technologies have also advanced even beyond the detailed population registers and personal information collected by governments and businesses (Black, 2001; Ketelaar, 2020). There are several different, interrelated data concerns. One is that, because of advances in technology, both as producers of data (e.g., passive tracking on smart devices) and data collecting tools (e.g., programs to scrape online social media data), a not insignificant amount (in terms of both volume and scope) of data is now being collected about individuals without specific purposes for their collection (Floridi, 2014; Lane et al., 2014). These data, collected ostensibly through data subject agreement to terms and conditions or acceptance of cookies (Degeling et al., 2019), can be further linked, aggregated, cross-referenced, and shared in what Susha, Grönlund, and van Tulder (2019) refer to as the “data revolution.” The combination of technologies and “bigger” data disrupt those constraints on information flows that previously gave some measure of protection to individual privacy.
(Barocas & Nissenbaum, 2014), from data collecting and analysis (Lievens & Verdoort, 2018; Schneier, 2015; Vestoso, 2018) to the global transborder data flows for research and commercial use (Farago, 2014; Hondius, 1980; Kuner et al., 2017; Quinn, 2018). Recent years have seen increased efforts to develop processes to link social media data with formally collected administrative or survey data, to enhance research data with geospatial information, or to link between distinct data sets; all of these efforts increase to potential for data subject reidentification even as they increase the utility of the original data sets (e.g., Beuthner et al., 2021; Breuer et al., 2020, 2021; Jones et al., 2014; Jutte et al., 2011; Stier et al., 2020).

2.3 Data protection legislation

The General Data Protection Regulation is not the first attempt to legislate data protection and privacy, though the history of data protection is not as long-standing as one might imagine. Prior to the 1970 adoption in the Federal Republic of Germany’s state of Hesse of the Hessisches Datenschutz, the first formalized data protection law in the world, privacy was regulated primarily through legal rulings and broad declarations in, among other documents, the 1948 Universal Declaration of Human Rights (United Nations, 1948). Since that 1970 law, however, data protection laws have been implemented at an increasing rate worldwide, and as of March 2022, 137 countries now have data protection laws of some form (United Nations Conference on Trade and Development, n.d.). In order to take part in global research and the global economy, many non-European Union nations have implemented their own data protection laws after the GDPR, as seen with new laws adopted in 2021 in Zambia, Rwanda, Uganda,

---

12 Datenschutz, 7.1.1970, GVBl. 1 1970, p. 625-627. This state law still remains in effect, and has been amended several times, most recently in 2018 to be compatible with GDPR. The most recent iteration of the Hessian data protection law is the Hessisches Datenschutz- und Informationsfreiheitsgesetz, 3.5.2018, GVBl 2018, p. 82.
13 Specifically, Article 12: “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.”
Zimbabwe, China, Andorra, Ecuador, Panama, Oman, and the United Arab Emirates (OneTrust DataGuidance, 2022). These laws are all in various stages of implementation. The GDPR is expected to continue influencing global data protection laws in the future, because data controllers (those who determine the purposes and means of data processing) and processors (those who process personal data on the controller’s behalf) in countries which want to process data about European Union residents, for whatever purpose, would need to abide by the GDPR anyway to gain access to the data in accordance with the GDPR’s Article 3(1).

This section will discuss data protection legislation in order to identify why its, and GDPR’s in particular, impact on data use is an important subject for study. First, Section 2.3.1 reviews the historical motivations for data protection and privacy legislation. Then, in Section 2.3.2, I will introduce the goals of the European Union, leading to the section 2.3.3 discussion of pre-2016 data protection legislation laws and their limitations, and Section 2.3.4’s discussion of the 1995 Data Protection Directive and how it was implemented in the different Member States. Then, Section 2.3.5 will describe the GDPR, beginning with why it was necessary and what it involves, its implementation (focusing primarily on the four countries selected for this dissertation) and ending with a discussion of the limitations of the GDPR with respect to research data sharing.

### 2.3.1 Historical motivations for privacy protection

While considerations of private affairs and private information date back earlier than Warren and Brandeis’ seminal work, events of the 20th century were the driving force for modern data protection laws. Warren and Brandeis (1890), and many of the contemporary American and European legal cases involving claims of privacy violations, were focused on those privacy violations that come from explicit acts of observation, theft, and unscrupulous publishing.
practices (Warren & Brandeis, 1890; Whitman, 2004). They did not speak directly to the risks inherent in the collecting of personal information in national censuses, identification cards, or other government or organizational records. The data of interest to this research are these personal or ‘sensitive’ data about people; this section, therefore, introduces historical data uses and abuses that have motivated calls for greater data protection.

Population censuses are not a modern invention; the earliest record of a census came from Herodotus (ca. 430 B.C.E./1909), describing censuses in the late Middle Kingdom and New Kingdom of Egypt. In Europe, population censuses were carried out in the Middle Ages for taxation purposes, such as the Domesday Book in England in 1086 and the L'état des paroisses et des feux in France in 1328. In the territories that would become the Netherlands, official registration of marriages, baptisms, deaths, and so forth began with the Catholic Church in 1214, and religious registration continued even after the Protestant Reformation. Civic registration, which served as the foundation for the Dutch persoonskaarten, began on a wide scale under French occupation in the late 1700’s (Ketelaar, 2020). These persoonskaarten would play an important role during World War II. In 1853, the First International Statistical Congress

---

14 For example, in New York, Marion Manola v. Stevens & Myers, Unreported, (N.Y. 1890), a case that went to the New York Supreme Court, seeking damages for the un-consented photographing and distribution of photographs of Broadway actress Manola. In the United Kingdom, Gee v. Pritchard, 2 Swanst. 402, 413 (1818), a case in the former Chancery Court that sought to prevent publication of private correspondence. Both of these cases, as with many others, relied not specifically on “privacy” but on property rights. In France, Dumas c. Liébert, CA Paris, May 25, 1867, 13 A.P.I.A.L. 247 (1867), a case where French author Alexander Dumas sued a photographer; Dumas had not established an agreement preventing the photographer from publishing, who then had sought to copyright and then publish them. Unlike the American and British cases, the French court established that, though Dumas did not have a property right to the photographs, “privacy… must sometimes be allowed to trump property, at least where lascivious images were involved: One’s privacy, like other aspects of one’s honor, was not a market commodity that could simply be definitively sold” (as referenced in Whitman, 2004, p. 1176).

15 For example, from Herodotus (as translated from ancient Greek by George Rawlinson), The history of Herodotus, line 177: “… while the number of inhabited cities was not less than twenty thousand. It was this king Amasis who established the law that every Egyptian should appear once a year before the governor of his canton, and show his means of living… Solon the Athenian borrowed this law from the Egyptians, and imposed it on his countrymen, who have observed it ever since. It is indeed an excellent custom.”

16 The Domesday Book was compiled in England in 1085-86 on the orders of King William I (William the Conqueror) as a record of land ownership. While not a full population census, it provided a comprehensive survey of landholdings in support of taxation. It is currently in the National Archives in London, and a digital edition is available at https://opendomesday.org/. L'état des paroisses et des feux de 1328 was a survey of “tax fires” in Paris in 1328, a form of taxation based on the idea of the “fire” being the family home, representing a taxable person.
recommended that every country implement similar population registers; at a national level, Belgium, Sweden, Finland, Hungary, the Netherlands, and Italy had such registers by the end of the 1860’s (Gutmann & van de Walle, 1978). In most European countries, decennial or near decennial censuses, population indexes, and population registers are usually carried out, and have been since the early 1900’s, if not earlier (Czech Statistical Offices, 2020). These data themselves were not considered “private” or “risky” on their own; it is how these data were later used that changed how they are viewed. While in some countries like the Netherlands, which faced protests against the censuses in the 1970’s, and pre- and post-reunification Germany, limited recent data collection has been carried out (Borchers, 2007; Ketelaar, 2020), that does not mean that similar data are not being collected by the government for other purposes, such as the social welfare system or by researchers in many disciplines.

As the Nazis invaded Europe, they took advantage of these national population registers. They were aided during the war years by the widespread use of IBM’s Hollerith punch cards and tabulating machines, used since the 1890’s in the United States and various European countries for the censuses (Black, 2001). The Nazis either carried out their own population censuses or converted extant national registers into Hollerith punch cards with detailed personal data, including religious identifiers, in nearly every country they invaded. These were then used to identify and sort the population and enable population transfers to concentration camps and ghettos in Eastern Europe (Bhaimia, 2018; Black, 2001; Ketelaar, 2020; Seltzer & Anderson, 2001). After the war ended, governments recognized the role that personal data, especially those identifiers that had been used for persecution (religion, ethnic background, etc.), played in the atrocities during the war. It is not that being Jewish or Roma, for example, was secret so much as how being able to identify people through these structured, detailed data including religious
affiliation and ethnic origin was used at scale to intentionally harm them. Over the next several decades, the United Nations, the European Community, and individual nations began promoting data handling principles, limitations on data collection, and limitations on data sharing and transfer (Aldhouse, 2018; Bhaimia, 2018; Kaye, 2012; Mayer-Schönberger & Cukier, 2014; Seltzer & Anderson, 2001; United Nations, 1948; Council of Europe, 1981). The intention of these actions was to prevent the misuse of personal data for purposes of persecution from ever happening again (Aldhouse, 2018; C. J. Bennett, 1992; Bhaimia, 2018; Flaherty, 1984).

The Holocaust, however, was not the last time records containing personal information were used in this way in the 20th century, nor was it the only use during World War II. Within two days of the Imperial Japanese Navy Air Service attack on Pearl Harbor in 1941, the United States Census Bureau, using the same IBM technology as in Nazi-occupied territories in Europe along with the 1940 decennial census data, produced the first report identifying all Japanese-Americans in extreme detail. These reports (several more would follow the first) were used to displace American citizens of Japanese descent to internment camps for the rest of the war (Black, 2001; Mayer-Schönberger & Cukier, 2014; Seltzer & Anderson, 2001). In another example, by 1994, when the Rwandan genocide began, a population registry system implemented by the Belgian colonial government which included an ethnic identifier had been in place since 1933. This identification system of classifying and disclosing ethnicity in Rwanda was used by the Hutu government and militias during those three months to identify Tutsi, Twa, and moderate Hutus during the conflict (Fussell, 2001; Nardone, 2010; Seltzer & Anderson, 2001). As a result of these abuses of personal information, post-World War II and more recent efforts to prioritize privacy and data protection specifically emphasize the importance of protecting identifiable religious and ethnicity information.
Solove’s (2008) taxonomy of privacy argues that a “privacy violation occurs when a certain activity [such as collecting, processing, or disseminating data] causes problems that affect a private matter or activity” (p. 102), whether the information were disclosed as a part of civic life or as part of research studies. Information about religion, ethnic background, sexuality, and so forth are not always private in every context, and the use itself is not always misuse or abuse. The privacy violations may come instead when that information is used in other contexts and without consent, even if the goal itself of that use is not to harm the data subject. As well, the use of personal information for genocidal purposes is not the only data use that result in privacy harms, though it is the most extreme. Data collected through other means than population registries, such as through surveillance, have also been used against individuals.

The post-war period coincided with the advancement in computing and data processing capabilities discussed in Section 2.2.2. It also coincided with a reinforced association of privacy and data protection with the right to dignity. An interesting outcome of the Communist-era in Europe is the conflict that arose, for example, in post-reunification Germany when Stasi (Staatssicherheitsdienst) surveillance files were made available. This conflict was over whether the public identification of Stasi informants who contributed to the creation of those files was justified, even though it was itself a further privacy violation (Borneman, 1997; Espindola, 2011). It raised questions in Germany, where concern over the right to dignity and reputation otherwise underscore German approaches to privacy, of whether the disclosure of such information that caused damage to the former informants’ dignity should be allowed. Justifiably, the subjects of the Stasi files were concerned about justice for how the Stasi surveillance disrupted and endangered their lives, but from the moment the first informants were publicly identified, arguments for the protection of their privacy arose. As Espindola (2011) found, the
“public identification of former collaborators in order to collectively shame them [was] regarded… as unduly disrespecting them by stigmatizing and holding them in contempt” (p. 34).

Taking the nature of the revelation out of the discussion for a moment, this concern over disrespect, stigmatization, and public humiliation is an otherwise widely-held motivation behind both the hesitation to share data and the need to protect those data. Experiences under communist rule also affected Czech parliamentary willingness to grant any exemptions to the reuse of personal data, even for scientific purposes, in the Data Protection Directive implementation law in 2000, something which would only change with the adoption of the GDPR.

Just as not all data uses result in governmental genocidal machinations, not all data and information disclosures are the product of justice seeking. In the last twenty years, high-profile personal data disclosures, and specifically the re-identification of individuals through data linking, algorithmic predictions, and hacking, have also come from data used by companies for commercial or administrative purposes. An interesting aspect of many of these violations and re-identifications is that they occurred despite efforts taken to anonymize data sets before they were made available for public use. The resulting re-identifications resulted from the technological ability to link across multiple anonymized data sources and other systems, such as public social media accounts, in order to triangulate down to an individual person. Ohm (2009) argues that the re-identification of individuals in anonymized AOL, Netflix, and Massachusetts’ Group Insurance Commission (see also Sweeney, 2002) data was due to a failure of and over-reliance on anonymization as a data protection mechanism (see also Mayer-Schönberger & Cukier, 2014; Solove, 2008). Other recent examples of data privacy violations are the well-known Target pregnancy data scandal deriving from big data mining and algorithmic processing (Barocas &
Nissenbaum, 2014; Junge, 2018); Facebook’s newsfeed experiment, which was also concerning for its poor research ethics and lack of informed consent (Clark et al., 2018); and the cyberattack on Equifax (Andriotis & Rapoport, 2018; Bernard et al., 2017; Ronaldson, 2019). It is clear just from these examples from the last century that information privacy, especially as it applies to data about people, is something that must be protected, and that it is still an on-going process to do so.

2.3.2 European Union goals and norms

While privacy has been a concern throughout history, and surreptitious surveillance and the disclosure of intimate details to the public did not begin with the advent of instantaneous photography, the twentieth and early twenty-first centuries have witnessed, and perhaps have been defined by, advances in technology that have drastically increased both the scale of data collection and the speed and impact of data processing. It is understandable, therefore, that legislative interventions in the use of both the technology and the data have taken a harm prevention approach. In the United States, legislation has taken a piece-meal approach, legislating data use and disclosure sector by sector, what Citron and Solove refer to as “a sprawling patchwork of various types of law, from contract and tort to statutes and other bodies of law” (Citron & Solove, 2022, p. 796). In the European Union, the laws have been broader, applicable to all data which meet criteria of being “personal data,” however those data are defined, and aimed at creating a standardized approach across the EU. This section will situate pre-GDPR attempts within Europe to regulate data sharing and use.

Since the end of World War II, a succession of international bodies were created that encouraged the centralization of authority of independent European countries: the Council of
Europe (1948), the European Coal and Steel Community (1951), the European Economic Community (1957), and the European Union (1992). In that period, the community of nations has grown from the original six countries to 27 (the 28th, the United Kingdom, formally left the Union on 31 January 2020; its future as a member of the European Economic Area is at this time uncertain). That growth, Shore (2000) argues, “has been accompanied by a shift in emphasis from integration, perceived as a rational by-product of economic prosperity and legal harmonisation, to more recent concerns with integration as a cultural process” (p. 1). In his study of European Union politicians, Shore found that the goal of the European Union was the construction of this idea of “Europeanness,” one that transcended national ideologies (see also Calligaro, 2013). Laïdi (2008) and Manners both underscore the European Union’s exercising of what Manners (2002) called “normative power”: “the ability to define what passes for ‘normal’ in world politics” (p. 253).

In a follow-up work, Manners (2008) acknowledged that the European Union was struggling to meet its goals of setting global norms, due to several factors: that not enough time had passed to evaluate the adoption of European norms elsewhere; the global political climate (specifically, the early 2000’s War on Terror); sovereignty (Russia, China, and the United States resist sharing and derogating their sovereignty to international law); and, universality (the need to convince others of the universality of European norms). Of the features that Laïdi (2008)
identified as defining European power, one important feature is that norms are “not only Europe’s best shield but also its finest banner… [because] the collective capacity to resort to force will never constitute a political goal in itself for Europe” (p. 139). The idea of “European norms” is not simply the product of the post-war period; pre-1945 Europe (as a collection of individual sovereign States) was still active in setting global norms, though they were often spread by force (i.e., through European colonialism) (Postel-Vinay, 2008). Coman (2018) describes how the European Union “governs through and by the rule of law when it seeks to regulate behavior and constrain action” [emphasis in original] (p. 88), thereby creating a “community of values” shared, and thus normalized, across the various Member States. The GDPR’s wide-applicability and requirement of non-European Union Member States to be deemed “adequate” in order to receive data (Article 45) can be seen as a further attempt to spread these values.

2.3.3 Pre-1995 data protection laws

Almost as long as European nation states have been building their post-war community, laws and declarations about the right to privacy have become prominent. Ethical standards for human subjects research have been established and refined (e.g., The Nuremberg Code (Nuremberg Military Tribunals, 1949, pp. 181–182), the Declaration of Helsinki (World Medical Association, 1964/2013), and the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979)). Article 1 of the Universal Declaration of Human Rights (1948) states that “All human beings are born free and equal in dignity and rights.” This document itself is not legally binding, but it has provided a shared conception of “fundamental freedoms” and “human rights,” including “life, liberty and security of person” (Art. 3), “equal protection of the law” (Art. 7), and that “no one shall be subjected to
arbitrary interference with his privacy… nor to attacks upon his honour or reputation” (Art. 12). In so doing, by the late 1970’s, it represented “the consensus of the international community on the human rights which each of its members must respect, promote and observe” (Sohn, 1977, p. 133). The ideals of the non-binding Declaration can be traced through to later binding legislation, reiterating the right to privacy and rights to data protection. In the United States, an additional ethical rule was implemented in 1981: the Common Rule. It followed the first revision of the Declaration of Helsinki in 1975, which added the requirement for a “specially appointed independent committee” to comment on and provide guidance for experimental research with human subjects (Art. 1(2)); in the United States this has been implemented as the Institutional Review Boards (IRB) requirement in the Common Rule. This applied initially to medical experimentation, but it has more recently evolved to cover social science research as well. In Europe, ethics boards like the IRB are becoming more common but are still not a requirement.

The 1970’s saw a series of data protection laws come into effect across Europe. The very first formal data protection law was the 1970 Hessisches Datenschutz. Under this law, data protection covered all documentation prepared for the purposes of mechanical data processing as well as all stored data and the results of their processing under the purview of the German state of Hesse. It also established an independent data protection commissioner to oversee and enforce the requirements. In 1974 in nearby Rhineland-Palatinate, a similar data protection law was implemented that only applied to data processed for state government purposes (as

---

21 Protection of Human Subjects, 45 C.F.R. § 46 (1981). Retrieved from https://www.govinfo.gov/content/pkg/CFR-2016-title45-vol1/pdf/CFR-2016-title45-vol1-part46.pdf. This was revised extensively in 2017; the most recent version can be found here: https://www.ecfr.gov/cgi-bin/text-idx?SID=300df04ebf09c7b23735d902a3f645a&mc=true&tpl=/ecfrbrowse/Title45/45cfr46_main_02.tpl.

22 §1 para 1, Datenschutz: “Der Datenschutz erfaßt alle für Zwecke der maschinellen Datenverarbeitung erstellten Unterlagen sowie alle gespeicherten Daten und die Ergebnisse ihrer Verarbeitung.”
discussed in C. J. Bennett, 1992). By 1977, West Germany passed the first iteration of Bundesdatenschutzgesetz (BDSG)\textsuperscript{23} to address the growing dangers and concerns over the advancing technologies for collection, processing, and disseminating data (Riccardi, 1983). This federal law was grounded in the constitutional Articles 1(1) and 2(1) of the Basic Law.\textsuperscript{24} In addition to Article 1’s right to dignity, Article 2(1) established the right to personality; both rights from the Basic Law form the basis for privacy in Germany as discussed above. The goal of the BDSG was to protect “the privacy of the individual by protecting his data from misuse” (Riccardi, 1983, p. 248).

The second European data protection law was the first national data protection law: the Swedish Data Act of 1973.\textsuperscript{25} The law was applicable only to processing personal data in computerized registers (personregister). It did not directly regulate when and how those registers were used, but unlike the laws in West Germany, it established a licensing system where no personal register could be created without the permission of the Data Inspection Board (Datainspektionen) (C. J. Bennett, 1992; Kirby, 1980; Datalag, 1973). As Bennett (1992) describes, the Swedish perspective was that regulation of these data ought to be at the point of collection (i.e., when data were added to the personregister) and not only after data misuse occurs, a different approach to the otherwise very reactive approaches taken in later laws.

The other major law passed in the 1970’s was the United States Privacy Act of 1974.\textsuperscript{26} In the United States, rights to privacy and data protection have generally been established

\begin{flushright}
\textsuperscript{24} Grundgesetz der Bundesrepublik Deutschland (1949). Retrieved from http://www.documentarchiv.de/brd/1949/grundgesetz.html. Translated by the author. The first post-World War II constitution in the Federal Republic of Germany came into effect on 23 May 1949. The text of Article 1(1) and 2(1) have remained unchanged since the first iteration of the Basic Law.
\textsuperscript{25} Datalag (1973:289), which was updated in 1992 and then repealed in 1998. Current data protection in Sweden, in compliance with GDPR, is the Lag (2018:218) med kompletterande bestämmelser till EU:s dataskyddsförordning.
\textsuperscript{26} Privacy Act of 1974, 5 U.S.C. § 552a (2014). Within the U.S.C., this section is entitled “Records maintained on individuals.”
\end{flushright}
through the court system applying combinations of the First, Third, Fourth, Fifth, Ninth, and Fourteenth Amendments (e.g., *Olmstead v. United States*, *Loving v. Virginia*, *Stanley v. Georgia*, *Eisenstadt v. Baird*, and *Roe v. Wade*). These court cases specifically focused on personal data and information about the intimate sphere, including wiretapping personal communications, marriage, personal use of pornography, contraception, and reproductive choice. The Privacy Act itself was only applicable to the personally identifiable information collected and maintained in records by federal agencies.

The Privacy Act arose out a growing public concern in the 1970s about the idea of a common government databank containing extensive information about individuals, but also about the growing prominence of data collected and shared through credit reporting, surveillance, criminal records, psychological tests for employment, etc. (C. J. Bennett, 1992). However, because the application of the Privacy Act is limited to data collected and maintained by federal agencies, and therefore is not applicable data to data collected by other organizations—a frequent criticism of the United States approach to privacy is the lack of an omnibus legal approach (see C. J. Bennett, 1992; Citron & Solove, 2022; Cockcroft & Rekker, 2016; Greenleaf, 2012; Schneier, 2015; Stratford & Stratford, 1999; Whitman, 2004)—privacy and data protection are elsewhere enforced through case law and individual industry protection regulations (e.g., HIPAA, FERPA).

At about the same time as the Privacy Act was implemented, the United States Department of Health, Education, and Welfare proposed and named the first Code of Fair Information Practices in its report on the potential consequences of computer-based record keeping and automated processing systems. There were three outcomes of these technologies that they saw as potentially dangerous: that data processing capacity was significantly enlarged,
that they facilitated access to personal data both within and across organizational boundaries, and the establishment of “record keepers” who were focused on the technologies (such as an IT department) and thus were far removed from both the original data producers and the future data users (Secretary’s Advisory Committee on Automated Personal Data Systems, 1973). These principles recommended that

- There must be no personal data record-keeping systems whose very existence is secret.
- There must be a way for an individual to find out what information about him is in a record and how it is used.
- There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data (Secretary’s Advisory Committee on Automated Personal Data Systems, 1973, pp. xx–xxi).

The 1980’s witnessed further passage of laws and attempts at establishing universal principles for data protection and privacy. The Council of Europe in particular was concerned about transborder data flows. Kirby argued that privacy was being jeopardized by “the expansion of automated data processing and the resulting proliferation of transborder data flows” (1980, p. 27). What the Council of Europe was worried about, in particular, was the potential that data protection could be circumvented by transferring data to other countries with lesser levels of protection (Aldhouse, 2018; Hondius, 1980; Kirby, 1980). Prior to the adoption of the Convention 108,27 some control over these international data transfers was written into some of

---

the extant data protection laws: e.g., the Swedish Data Act, and the Private Registers Act of 1978 and the Public Registers Act of 1978 of Denmark, all of which regulated these transfers through licensing systems. Convention 108, adopted in 1981, was intended to

secure in the territory of each Party for every individual, whatever his nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him (‘data protection’) (Art. 1).

Article 5 established binding principles for automated data processing which would re-appear in both the Directive and the GDPR, while Article 12 set down provisions for transborder data flows. By the 1990s, thus, data protection legislation, though different in terms of applicability, enforcement, and level of protection, were becoming more common, as was the recognition that harmonization of principles and protections was needed to prevent data protection circumvention.

2.3.4 Data Protection Directive

As a mechanism for sharing European norms and values, the 1995 Data Protection Directive had a wider global impact than had been anticipated (Birnhack, 2008). Laïdi’s (2008) discussion of European normative powers may be grounded in the economic and trade adaptations that Europe’s trade partners have had to make, but the EU has carried out a parallel process in its approach data protection. The Directive’s principles themselves have drawn on the Organisation for Economic Co-operation and Development’s (OECD) 1980 non-binding


guidelines,\textsuperscript{30} which outlined the fair information practice principles, and the Council of Europe’s 1981 Convention 108, the first \textit{binding} international treaty aimed at protecting privacy with regard to the collection and processing of personal data. Compared to the OECD guidelines and the Convention 108, the Directive had stronger principles as well as stronger requirements for enforcement of those principles (Greenleaf, 2012; Stratford & Stratford, 1999). The objective of the Directive, laid down in Article 1, was two-fold:

1. In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.

2. Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1.

It is important to note that while the Directive did not want to interrupt data sharing, the priority here was the protection of the right to privacy, in recognition that there were specific types of data that, while not on their own risky, could be and had been used to identify and cause harm to the data subjects. The Directive was intended also to create a uniform level of data protection across the Member States (Allen, 2011; Poullet, 2006). In so doing, it was “a completely new regime in human rights protection, covering all systematic processing” (Rouillé-Mirza & Wright, 2004, p. 127).

There were a number of important principles in the Directive. Article 3 established the scope of the processing, both automatic and by other means of a structured data set, and of processing within the scope of European Community law. Processing of personal data, according to the Directive, was

any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction (Art. 2(b)).

Article 5 gave Member States the flexibility to “determine more precisely the conditions under which the processing of personal data is lawful.” Article 6 identified the five principles of data quality the States must provide for: that personal data were processed fairly and lawfully; that they were collected and processed for specific purposes; that they were adequate, relevant, and not excessive; that they were accurate and up to date; and not retained for longer than necessary. As Nys (2004) highlighted, Article 6(1)(b) allowed for data processing, including possible reuse, for research purposes. However, Article 7 restricted the legitimacy of personal data processing. The criteria for determining legitimacy were that: there had been unambiguous consent; processing was necessary to meet contractual obligations; processing was necessary for compliance with legal obligations; or to carry out tasks that were in the public interest.

Article 8(1) of the Directive prohibited processing of “special categories” of data unless one of five exception criteria are met (Art. 8(2)-8(5)). These special categories were those which have been used to identify specific groups for discrimination, and which the European Union and the United Nations have been invested in protecting: “personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life” (Art. 8(1)). The transfer of data to third countries (those outside of the European Union or the European Economic Area, though the inclusion of the latter differed across the Member States) was governed by Article 25, which required that “the country in question ensures an adequate level of protection” (Art. 25(1)), and Article 26, which identified the conditions under which the third countries could receive data when they do not meet those criteria.
Intended to respond to advancements in data processing capabilities through the mid-1990s, the scope of the Directive applied to both automated and non-automated processing of personal data, except where data were processed for security, defense, criminal law, and European Union administrative processing operations (Art. 3(1) and 3(2)). Its focus was not solely on the protection of the data; it also recognized a right to and a need for personal data processing in certain specified contexts, and also that privacy did not exist in a vacuum. It must be, according to Hartlev (2004) and Townend (2004), reconcilable with broader societal needs and interests.

2.3.4.1 Implementations of the Data Protection Directive

In the 21 years following the adoption of the Directive, the Member States transposed it into their own national legal regimes, as required by Article 4. These differences, paired with technological advances, would, in the following 20 years, lead the various Members States and the European Union more broadly to recognize that there was a need for a revision, if not a wholesale replacement, of the law. For example, Austria, Finland, Sweden, and the United Kingdom jointly proposed a series of amendments to the Directive (Freedom of Information and Data Protection Division, 2002) which included clarifications about the special categories of data, whose vital interests took precedence in data processing without consent, consent withdrawal, information to be provided to data subjects, and data subject rights of access and rectification. They also proposed clarifications about third party data transfers and what “adequacy” actually meant.

At the time of the first European Commission (2003b) report on Directive implementations, the Commission did not agree that revisions were necessary, but within a decade that opinion would change. The legislation described here reflects those Directive
implementations in place by 2012; both unofficial and official translations (sources of the translations are identified in the footnotes) of these laws into English from the relevant Member State authorities were reviewed, as were several European Commission studies of the Member State implementations (European Commission, 2003a, 2003b; Korff, 2002). In this section, I will identify areas of the Directive where the different transposition, over the intervening twenty years, resulted in increasing disharmony and divergence across the European Union data protection regime.

The first component where these differences were both apparent and trigger further deviations are the transposed definitions from the Directive’s Article 2. Some terms, like “personal data,” “processing,” “processor,”31 “third party,” and “data subject’s consent,” were either verbatim or very closely defined in line with the Directives definitions by at least fifteen of the then-twenty-seven EU Member States in 2012 when the GDPR was first proposed. Where there were differences, they were usually expansions on the Directive’s terms, though not always.

Some definitions, such as for “personal data” and “processing,” were expanded upon, with additional types of personal data (e.g., sound and imagery data or genetic data) or processing (e.g., collecting biometric data). Many of these additions would end up included in the GDPR in 2016. Others, such as “controller,”32 saw significant deviation, with only 12 of the 27 national laws using the Directive’s definition verbatim. The different iterations were more than just a variation on wording; in several national laws, the responsibilities of the “controller”

31 Article 2(e) Directive defines the “processor” as: “a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller.”
32 Article 2(d) Directive defines the “controller” as: “the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law.”
were altered. For example, Germany, the Czech Republic, and Bulgaria incorporated the “controller” role into the “processor” role, while in Austria, the controller only determined the purposes of processing and not the means. The biggest challenge in how the “controller” was defined by the different Member States, according to Korff (2002) was that the Directive’s definition included “alone or jointly with others” (Art. 2(d)), which was transposed directly into many of the implementations, indicating that there could be more than one controller for a given data set. However, many later provisions in the Directive assumed that there would only be one controller, which Korff argued could raise compliance challenges in situations where there were multiple controllers.

In its first report on the implementations, the European Commission recognized that there were “certain provisions which leave little or no margin to the Member States and where divergences have never the less occurred” (European Commission, 2003b, p. 11). The authors specifically identified the definitions as a component of the Directive which should not have been deviated from and with which the States did not comply. Expanding beyond these differences in how the national implementations defined relevant terms, these definitions influenced how the remaining principles and requirements of the Directive were implemented.

Following Article 3’s explication on the scope of the Directive, Article 4 established the territorial scope of the Directive. In particular, in Article 4(1)(c) it addressed the application to processing when the controller was not a part of the European Community:

Each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where: … (c) the controller is not established on Community territory and, for purposes of processing personal data makes use of equipment, automated or otherwise, situated on the territory of the said Member State, unless such equipment is used only for purposes of transit through the territory of the Community (Art. 4(1)).

Countries outside of the European Community were known as “third countries,” a label that continued after the Treaty of Lisbon to mean countries outside of the European Union. There was no mention within the Directive about the European Economic Area; three countries (Norway, Iceland, and Liechtenstein) were and still are members of the EEA but are not in the European Union. Questions were raised as to whether these three countries ought to be considered a part of the European Community for the purposes of the Directive. Requirements among the States for transferring personal data to non-European Union and non-European Economic Area countries were overall very consistent with later Directive articles about that transfer. However, when it came to the EEA countries, the States were split between treating those three countries as if they were European Union States, and thus having a supposedly more streamlined approach to data transfer because they met the Directive’s requirements, or as “third countries,” where there was a greater burden of proof necessary to show that the countries met the Directive’s adequacy requirements.

As presented in the previous section, the Directive’s Article 6 identified five principles about data quality. In general, these principles were transposed into national laws in accordance with the Directive in the same structure that they were presented in the Directive: as a list of principles and obligations for the controller. Among these laws, differences were primarily in
the form of different-but-related word choices in the translation, or expansion on obligations (this expansion was especially visible in the Finnish law\(^\text{37}\)), but the overall interpretation of the principles aligned with the intentions of the original principles. The first principle required that the personal data “must be processed fairly and lawfully” (Art. 6(1)(a)).

The second principle required that personal data be

<table>
<thead>
<tr>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards (Art. 6(1)(b)).</td>
</tr>
</tbody>
</table>

The first part of this principle was the most consistently transposed of the five principles into the national laws. However, the Czech, Cypriot,\(^\text{38}\) Estonian,\(^\text{39}\) and Hungarian\(^\text{40}\) laws did not include the clarification about historical, statistical, or scientific processing, and a related clarification about storage of data for the same purposes (Art. 6(1)(e)) was missing entirely from the Danish,\(^\text{41}\) Czech, Estonian, and Hungarian laws. As discussed in Section 4.1, this omission from the Czech law continues to affect Czech social science data archiving to this day: when ČSDA opened in 1998, the only data that were allowed to be archived for reuse were anonymized data. There were no significant differences in the transposition of the third principle, the requirement that the data be “adequate, relevant, and not excessive” (Art. 6(1)(c)).

---


The fourth principle, requiring that the data be “accurate and, where necessary, kept up to date” (Art. 6(1)(d)), however, was different. While many Member States transposed this principle essentially the same, a few laws excluded the final clause requiring the erasure or rectification of inaccurate or incomplete data (Cyprus, Croatia,42 and Poland43), though not all changes were exclusions. With regard to the fifth principle, in which the Directive limited the retention of identifiable data, it was overall transposed in accordance with the Directive. The most significant differences were those discussed above, where the Czech, Danish, and Hungarian laws did not include the clarification about extended storage for historical, statistical, or scientific usages, and Estonia did not include this principle in any form.

The implementations may be similar in wording to the Directive or to each other regarding the five principles, but the Directive (and thus those laws) was vague in how the principles were to be applied. It was left up to the Member States themselves to interpret them, even in terms of just what were to be considered “appropriate safeguards” for personal data (as described in the Directive’s Article 6(1)(c)). Some laws went into significant detail about what the principles meant in practice, while others remained, like the Directive, as broad principles open to interpretation with little formal application.

Article 7 of the Directive presented the six criteria for making data processing legitimate; The six criteria were essentially transposed verbatim into all national laws, though with some expansions or attempts to clarify. There are: the first, that “personal data may be processed only if the data subject has unambiguously given his consent” (Art. 7(a)); the second, that personal data may be processed “if necessary for the performance of a contract” (Art. 7(b)); and, the third,

“if processing is necessary for compliance with a legal obligation” (Art. 7(c)). For the fourth criterion, which stated that “personal data may be processed only if processing is necessary in order to protect the vital interests of the data subject” (Art. 7(d)), nearly all Member States expanded on the wording. The Directive did not define “vital interests,” leaving it a very broad term that could be interpreted as anything from personal health to finances to civil rights. Austrian, Finnish, Swedish, and British concern about the meaning of “vital interests” and what claims of it as a legitimate reason to process data led to their proposal that the proviso be deleted from this Article (Freedom of Information and Data Protection Division, 2002). Several Member States which joined the EU in 2004 explicitly clarified this by replacing “vital interests” with language more explicitly in line with other EU declarations of rights (e.g., life, health, etc.), including the Treaty on European Union and the EU Charter of Human Rights (the latter ratified in 2000).

There was some variation in how the States transposed the fifth criteria for processing “in the public interest” (Art. 7(e)), but most transpositions were verbatim or in accordance with the Directive. Finland and Italy\(^4\) limited processing for public interest to processing provided for by law and only to the extent that is necessary for that legal requirement; if controllers wanted data processed for public interest that did not meet this limitation, the Finnish law required a permit from the data protection authority. The final criterion (Art. 7(f)) consisted of what Korff (2002) referred to as the “balance test”:

\[
\text{processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental}
\]

rights and freedoms of the data subject which require protection under Article 1(1).

As with the fourth criterion, most Member States transposed it either verbatim or in accordance with the Directive. In addition to these principles, several Member States added additional criteria for legitimate personal data processing. For example, the Czech law allowed for processing personal data about “publicly active persons… that reveals information on their public or administrative activity, their functional or working position” (§ 5(2)(f)). However, as referenced above, this did not apply to scientific research and data collection.

When it came to processing special categories of personal data (also referred to as sensitive data in many definitions) as defined in Article 8(1):

Member States shall prohibit the processing of personal data revealing racial or ethnic original, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life, the Member States all identified the categories of data given in the Directive, at a minimum. The changes were all inclusions of additional categories of data (e.g., genetic data, criminal conviction data, sexual orientation, etc.), which would later be added to the GDPR’s revised definitions. In terms of the limitations on processing special categories of data, the Member States generally transposed both the prohibition on processing and the exceptions to that prohibition in line with the Directive. Every law allowed for data subject consent to exempt processing from the prohibition (in accordance with Art. 8(2)(a)). Other exceptions allowed for processing to “protect vital interests of the data subject” if consent is not possible (Art. 8(2)(c)), processing for legitimate activities of foundations or non-profit organizations (Art. 8(2)(d)), and processing of personal data already made public (Art. 8(2)(e)). Article 8(4) gave Member States the authority to, in matters of “substantial public interest,” implement additional exemptions to the prohibition on sensitive data processing.
The previously discussed provisions in the Directive focused primarily on protecting and preserving the privacy of data subjects and understanding when that protection could be overruled. This was in line with the first goal of the Directive. The second goal of the Directive was to “neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with [that protection]” (Art. 1(2)). The two goals together thus provided, in principle, a “free zone” within which data could be transferred within the European Community as long as they were within the scope of European Community law (as dictated in Article 3(2)). However, most Member States did not explicitly limit such data transfers to that scope, which meant that, effectively, no restrictions actually existed.

Only Austria’s law expressed this limitation fully in accordance with the Directive, recognizing a lack of restriction on data transfer to other States in the European Union, but not applying that freedom to “data exchange between public sector controllers in fields that are not subject to the law of the European Union” (Art. 2 Part 2 § 12(1)). The other main exception to this free flow of data, even when it was within the scope of European Community law, was when the data were to be shared with “representations of foreign governments or intergovernmental institutions in Austria,” which required Commission permission (Art. 2 Part 2 § 13(6)). Otherwise, the requirements for licensing and permits for Austrian data were only applicable to third countries which did not meet an adequate level of protection. Similarly, the Greek, Estonian and Portuguese laws explicitly stated that there was no restriction on intra-Member State data transfer, though they did so without limiting that freedom in line with the Directive. The German law was situated somewhere between the Austrian and the Greek, Estonian, and Portuguese approaches, distinguishing between transfers within and beyond the scope, but also not stipulating the “free flow” of data (Korff, 2002). Otherwise, the laws which did not mention
intra-Member State transfers implicitly promoted the data transfer both within and beyond the scope of European Community law, and thus, not in accordance with the Directive. The problem inherent in this conflict in applicability was that, even with the Directive’s expectations, combined with other variations in implementations, Member States which already did not have the same level of data protection (and potentially also not even “adequate” data protection, but were grandfathered in by virtue of being in the European Union) could still authorize unrestricted data transfers to those other States.

Beyond the European Union, the Directive’s Articles 25 and 26 provided the principles and derogations of that data transfer to third countries. The Directive did allow for the transfer of personal data to third countries, but only if “the third country in question ensures an adequate level of protection” (Art. 25(1)). Just what “adequacy” meant was not clearly defined. As discussed above, there were already significant differences in whether the non-European Union countries which are yet in the European Economic Area (Iceland, Liechtenstein, and Norway) were considered to be “third countries” for the purposes of these data transfers. Within the “circumstances” which the Directive said must be considered when making the adequacy decision, there were not clear thresholds provided in the Directive that the third countries must meet. The Directive also allowed the Member States to make these determinations, not just the European Commission, and the Member States took different approaches to do so. About half of them required that either the European Commission itself or their own national data protection authority make that determination. The other countries allowed the controller to make that determination, consulting with the European Commission only if they themselves were in doubt.

The challenge inherent in different definitions of key legal terms, as insignificant as some of them may seem, was in how they trickled down into the later transpositions of other Directive
provisions discussed in this section. Just as the European Commission (2003b) report identified non-compliance with Community law in terms of the definition transpositions, in the same sentence, the authors also identified the “closed list” articles (7, 8(1), and 26 as discussed in this section, as well as 10\textsuperscript{45} and 13\textsuperscript{46}) and other places where the Member States failed to comply with Community laws. Inconsistent definitions resulted in inconsistent applications, raising concerns of such potential scenarios as this one: researchers in one country conducting multi-national comparisons where the data included deceased persons, whose country did not consider these data to be personal data and thus subject to protections under the Directive, run into difficulties accessing similar data from another country that did consider those data to be personal. The two countries’ institutions might come to an agreement about how those data were to be handled, but the fact that it would need to be negotiated at all between two European Union Member States ostensibly beholden to the same requirements was indicative of the problems faced in trying to achieve the free flow of information the Directive called for in Article 1(2).

The use of different criteria for processing, different levels of authority determining whether other organizations or states implement “adequate” data security, and so forth, between

\textsuperscript{45} Article 10, Directive 95/45/EC, gave the requirements for data subject notification, stating that “Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information, except where he already has it: (a) the identity of the controller and of his representative, if any; (b) the purposes of the processing for which the data are intended; (c) any further information such as - the recipients or categories of recipients of the data, - whether replies to the questions are obligatory or voluntary, as well as the possible consequences of failure to reply, - the existence of the right of access to and the right to rectify the data concerning him, in so far as such further information is necessary, having regard to the specific circumstances in which the data are collected, to guarantee fair processing in respect of the data subject.”

\textsuperscript{46} Article 13, Directive 95/45/EC, provided the exemptions that allowed Member States to restrict the rights of data subjects “provided for in Articles 6 (1), 10, 11 (1), 12 and 21 when such a restriction constitutes a necessary measures to safeguard: (a) national security; (b) defence; (c) public security; (d) the prevention, investigation, detection and prosecution of criminal offences, or of breaches of ethics for regulated professions; (e) an important economic or financial interest of a Member State or of the European Union, including monetary, budgetary and taxation matters; (f) a monitoring, inspection or regulatory function connected, even occasionally, with the exercise of official authority in cases referred to in (c), (d) and (e); (g) the protection of the data subject or of the rights and freedoms of others. 2. Subject to adequate legal safeguards, in particular that the data are not used for taking measures or decisions regarding any particular individual, Member States may, where there is clearly no risk of breaching the privacy of the data subject, restrict by a legislative measure the rights provided for in Article 12 when data are processed solely for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics.”
27 Member States\textsuperscript{47} by 2012 left a legal quagmire which data controllers and date reusers had to navigate, especially those interested in data from multiple countries. As this section has described, even where States approached one aspect in line with the Directive and with other States, other aspects were frequently applied differently. The implementations were inconsistent, confusing, and occasionally contradictory, and as information technology capabilities advanced, especially in terms of the volume and speed of data transfers, the flaws in the Directive became more apparent.

2.3.4.2 Limitations of the Data Protection Directive

By focusing on automated as well as manual methods of data collection and processing, the Directive was responding to the growing concerns about the ability of technological advancements to infringe on privacy rights. Computing technologies and processing power advance rapidly, as predicted by Moore’s Law (G. E. Moore, 1965), driving down the cost of the technologies while increasing the speed and ability to complete desired tasks. By the 1990s, the automated data processing capabilities were concerning enough for the European Union to regulate it, resulting in the Directive. Their identification of specific types of data as potentially risky and harmful recognized that harms to individuals and groups were possible through the processing and identification of those data subjects, by opening them up to persecution through their data. Therefore, by identifying the data as well as the data processing which required particular care and oversight, the Directive aimed to protect the individual data subjects from those harms by regulating who, how, and why personal data were processed.

Despite the Directive’s objectives, however, it had significant limitations that ultimately led the European Commission in 2012 to propose a new framework for personal data protection.

\textsuperscript{47} Croatia joined the EU in 2013 after the GDPR was first proposed.
The most important limitation could be found in its name: that it was a Directive, not a Regulation. Where European Union regulations are legally binding both in terms of aims and the methods of meeting those aims, directives are only binding in terms of the aims (Beyleveld, 2004b). The Directive was to be transposed onto all Member State’s national legislation, but because Article 5 allowed significant variability in how the Member States transposed the Directive, as discussed in detail above, it did not result in the desired harmonization, something which became increasingly apparent as the technology for data collection, processing, and sharing became more advanced (Custers et al., 2018; Poullet, 2006; Tikkinen-Piri et al., 2018).

Custers et al. (2018) further argue that this lack of harmonization is due to cultural as well as legal differences between the Member States that resulted in varied legal transpositions of the Directive within those national laws. They base their argument on the different practical and legal implementations as well, drawing on their comparison of data protection laws in Germany, Sweden, the United Kingdom, Ireland, France, the Netherlands, Romania, and Italy prior to GDPR going into effect. In their analysis of the laws in the post-Directive period, leading up to the full implementation of GDPR, they found that differences existed in: whether international and national law were separate legal regimes within Member States; whether a right to privacy was provided for in a constitution, other legislation, or through judicial precedent; whether privacy and data protection were matters of public debate; and, whether the national transposition of the Directive had indeed been more strict than the broad principles outlined in 1995. The Member States also regulated data sharing and processing of personal and special categories of data in different ways. Some used registration as the means to regulate and

supervise personal data processing, while others used notifying the data protection authority (with or without further registration).

These differences, whether they were large or small, raised concerns about how different transpositions of the Directive and the methods for ensuring their enforcement would restrict the willingness and ability to transfer personal data if one State’s data protection practices were seen as “less” than another State’s (Casabona, 2004; Dove, 2018; Nurmi et al., 2018). While the focus was on commercial and governmental data processing, research institutions were also concerned. Beyleveld (2004b) criticized the Directive’s lack of specificity about just what practices must be put into place to safeguard data subjects. Casabona (2004) specifically challenged anonymization as a primary data protection method. Per the Preamble of the Directive, the “principles of data protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable” (para. 26). However, Casabona argued that “anonymisation is a relative concept since its irreversibility is not always definite” (2004, p. 48), not to mention that technical capability to link and process big data sets have advanced beyond what was conceived of in 1995, leading to increased risk of re-identification (Koops, 2014; Lachaud, 2018; Lindqvist, 2018; Mayer-Schönberger & Padova, 2016; Sweeney, 2002; Wong & Henderson, 2019). Ohm (2009) recognized the growing ease of re-identification, especially as related to the broad principles instead of specific requirements, and recommended the need specifically for a comprehensive regulation to set a specific baseline for data protection.

2.3.5 General Data Protection Regulation

The limitations of the Directive were increasingly apparent by the end of the 2010s. Confusion over requirements, the “patchwork of national approaches” (van Deursen & Kummeling, 2019, p. 4), and the ineffectiveness of the Directive in the internet age meant that, if
the European Union was truly concerned about enabling the free flow of data while maintaining individual privacy, the Directive would need to be updated, if not completely overhauled (C. J. Bennett, 2018; European Commission, 2012; Tikkinen-Piri et al., 2018). The 2012 proposal for the GDPR explicitly identified the following in its argument that a new legal framework was necessary:

Rapid technological developments have brought new challenges for the protection of personal data. The scale of data sharing and collecting has increased dramatically. Technology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Individuals increasingly make personal information available publicly and globally. Technology has transformed both the economy and social life (European Commission, 2012, p. 1).

After the proposal, the Article 29 Working Party, in its role as an advisory board for data protection in the European Union, provided opinions and recommendations about the proposed regulation. The draft GDPR was accepted by the European Parliament in 2014, and after further discussion and agreement on the implementation plan, the full GDPR was adopted by the Parliament on April 14, 2016. There was a two-year period for the Member States to bring their national laws into congruence with the GDPR, and for entities which process personal data to bring their operations into compliance, and it entered into force on May 25, 2018. Concerns over the advancement of technologies for data processing, storing, and sharing, the associated risks of harm to data subjects, and the loss of data subject control over how, when, and where their data were being used, interweave throughout the Regulation, with different Articles addressing one or more of the concerns.

2.3.5.1 Overview of the GDPR

The intention of GDPR is to harmonize data protection within the European Union, so that data can flow freely across borders. The limitations of the Directive ultimately lay in the
fact that it was a directive. As Beyleveld (2004a) described, directives are binding on Member States only with regard to what the Directive aims to achieve. They are not binding on the method by which that aim is met. The flexibility the Directive offered Member States ultimately resulted in 28 different legal regimes by 2016, which meant that any efforts to enable the transfer of data across national borders, especially for multinational institutions which had to comply with the Directive under multiple legislative standards (Poullet, 2006), became more complex.

A European Union regulation, on the other hand, is directly enforceable, not requiring national legislation to implement its requirements as the Member States deemed necessary to meet requirements (Cornock, 2018). Regulations are an instrument for imposing European Union norms (Laïdi, 2008); the GDPR is designed to establish norms through legal coherency and consistency, in place of those individual approaches (Albrecht, 2016; Dove, 2018). The regulation is thus a much stronger legal instrument, because the Member State cannot freely transpose the requirements into their own laws to fit their own legal and cultural practices (Blume, 2015). De Hert and Papakonstantinou (2016) see this to be the most important contribution of the GDPR. In practice, the States have updated or drafted new data protection legislation to reflect the GDPR. The harmonization is only possible because the states are limited in their ability to adapt to the requirements as much as they could under the Directive. Mondschein and Monda (2019), however, raise the concern that what allowances GDPR Article 89(2)\footnote{Article 89(2), GDPR, states that “Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.”} does provide may still lead to continued fragmentation that, especially for researchers, could continue to prove challenging for data users and data sharers within the European Union and beyond it (see Section 2.3.5.3 for a description of how the Czech Republic, Finland, the
Netherlands, and Germany have taken advantage of these opportunities). This article specifically addresses personal data processing for scientific, historical, or statistical research purposes. Whether their concerns are valid remains to be seen.

As with the Directive, GDPR takes an omnibus approach to data protection; its requirements and restrictions are applicable to all personal data, not just personal data over specific subjects. Much of the text of the GDPR is an expansion on the Directive, and a comparison of many articles indicate that many of the changes in the GDPR are intended as expansions or clarifications reflecting new data processing capabilities and recognition of the risk inherent in other types of revelatory data not previously considered. Several of these, as well as what articles are entirely new, will be discussed below.

In recognition of the technological and data processing advancements, the GDPR expanded what data activities would be considered “personal data processing” and extended the data types which are considered “special” and of which processing is prohibited unless specific conditions are met. As mentioned in the previous section, many of these additions were already in later Directive implementations and were brought into the revised text of the GDPR. As discussed in Section 2.3.1, the 20th Century witnessed a multitude of events and actions wherein particular identifying information, whether about religion or ethnic identity among many others, have been used to discriminate, sometimes fatally, against certain populations. By classifying data related to these topics as deserving of extra protection, both the Directive and the GDPR have recognized their potential for harmful misuse. Restricting the use of the data as they do aims to minimize their use against individuals.

50 Article 4(2), GDPR, defines “processing” as “operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.”
The significant change in the new GDPR restriction on the special categories of personal data processing (Art. 9(1)) specifies “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,” where the Directive identifies “the processing of data concerning health or sex life” (Art. 8(1)). The inclusion of genetic and biometric data, and not just general “health” data reflects technological advancements that have enabled easier identification, especially regarding the exchange and use of biometric data for administrative and governmental purposes. The privacy risks that come from this type of data have been heavily discussed in both academic and popular literature (e.g., Kindt, 2013; Natgunanathan et al., 2016; Prescott, 2020; Sharvit, 2009).

It is important to recognize that the GDPR, like the Directive, does not outright prohibit the processing of personal data as long as the processing principles are followed and one of the legal bases apply. Processing of the special categories of data identified in Article 9(1) are prohibited, unless one of ten processing exemptions to the prohibitions listed in Article 9(2) apply. Among these exemptions are processing that has been explicitly consented to, processing to protect the vital interests of those who are unable to consent, processing of special data that the subject has made public, processing for “reasons of substantial public interest” (Art. 9(2)(g)), processing for medical and public health purposes, and processing for “archiving purposes in the public interest, scientific or historical research purposes or statistical purposes” (Art. 9(2)(j)).

These exemptions align with the criteria for lawful processing provided in Article 6(1). The provisions in this article are almost identical to the Directive’s Article 7, except for the

---

51 Article 9(1), GDPR, states that “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.”
distinction between the GDPR’s identification that “[p]rocessing should be lawful” (Art. 6(1)), while the Directive was concerned with the “legitimacy” of the processing. This is an important distinction, as Art. 5 of the Directive allowed and required Member States to identify for themselves what was lawful and legitimate processing. Article 6(2) GDPR does allow Member States to identify additional requirements pertaining to compliance with Member State legal obligations or legitimate data controller interests (Art. 6(1)(c) and (e)), but not to the degree the Directive previously called for.

These lawful bases for personal data apply not just to the processing (including use) of data by the initial data collectors, but also to the reuse of data. Data reuse by the same users, for the same purpose in the same context as to what was originally consented, is fairly straightforward, but reuse for different purposes in the same or in a different context is less so (Custers et al., 2019), because of the challenges brought on by the GDPR’s approach to consent. On the one hand, both Article 6(1)(a) and Article 9(2)(a) require consent for processing to be lawful, but on the other hand, neither article defines just what that means. The definition of consent in Article 4(12) does add the requirement that it be unambiguous, incorporating a common Member State transposition of the Directive’s definition not originally included, but it is not clear how consent should be demonstrated. The rights and requirements for consent are instead delineated in Article 7: that consent to processing must be demonstrated; if it is given in a written declaration, it must be separate from anything else in the document or contract; that consent may be withdrawn easily at any time; and, determining whether contract fulfillment or service use requires the data subject to consent to data processing not actually required for that contract or service.
De Hert and Papakonstantinou (2016) state that consent is the “most important legal ground for personal data processing” (p. 187), while Jacobs and Popma (2019) identify informed consent as the only basis for processing and sharing data for scientific medical purposes. Companies collecting data about users must obtain consent to use the data which they collect, including data gathered online (Degeling et al., 2019; Rüeger et al., 2018). Not only is consent a mechanism for ensuring ethical research conduct, it also provides data subjects with a sense of control over how the data about them will be used. They know, based on what they consented to, how and in what context their data will be processed, shared, and for what purposes. Consent ing at the point of data collection to all possible potential future processing is nigh impossible because knowing in advance what possible uses of data may be put to is not possible (Wallis, 2014), and GDPR makes it both difficult to obtain broad consent in the first place (Mayer-Schönberger & Padova, 2016), while also limiting the collection of “extra” data that might be useful for said future processing (Art. 5(1)(c)). However, the granting of consent for processing does not necessarily derogate other limitations that GDPR imposes on processing (Hoofnagle et al., 2019).

2.3.5.2 Updates to the Data Protection Directive

The Directive was an expansive piece of legislation in terms of the data it applied to and the behaviors it sought to legislate. The GDPR in many ways takes it much further; as de Hert and Papakonstantinou (2016) made clear before GDPR’s adoption, “[t]here is very little personal data processing that will remain unaffected by the combined effort of the Regulation and the Directive” (p. 180). As discussed above, Article 9(1) refined the type of data that the GDPR applied special restrictions to, though all data containing personal information that can be used to re-identify data subjects is covered (Hoofnagle et al., 2019; Art. 4(1) GDPR). Importantly for
companies and research organizations using data from the European Union, the GDPR applies to all data about European Union residents, whether the data are in the European Union, and whether the controller or processor is located in the European Union.

The direct result of this is the extension of the territorial scope of GDPR, as per Article 3. All organizations established within the European Union, whether they are collecting, processing, or sharing personal data within the Union, as well as processing of data about European Union data subjects, must meet the data security, protection, and lawful processing requirements delineated in the GDPR. But it also applies to controllers and processors established outside of the European Union if the data they are processing relate to data subjects inside the Union. Many internet users will recall the immediate aftermath of GDPR going into effect in May 2018, when websites began, en masse, requiring users in and outside of the European Union to accept or modify new cookie settings that collected data about their internet usage, and many organizations also emailed users their new GDPR-compliant privacy policies on or after that date (e.g., Degeling et al., 2019; Murphy, 2018; Sommerlad, 2018).

Whether intentional or not, one outcome of GDPR will be the spreading of European Union norms of privacy and data protection beyond the European Union, as in order to collect or receive data, organizations within the GDPR’s “third countries” must also meet the same criteria as the Member States (Greenleaf, 2012; Kuner et al., 2017; Manners, 2002; Mayer-Schönberger & Padova, 2016; Mondschein & Monda, 2019). This, however, does create beyond the borders of the European Union the scenario which GDPR is intended to correct within them. The

---

52 Article 3, GDPR states that “(1) This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not. (2) This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to: (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or (b) the monitoring of their behaviour as far as their behaviour takes place within the Union.”
challenge inherent in attempting to achieve a status of “adequate” per Article 45(1)\textsuperscript{53} across different legal regimes (van Deursen & Kummeling, 2019) cannot be underestimated. As Bhaimia (2018) argues, the case of post-Brexit United Kingdom exemplifies this; the easiest way for the United Kingdom to continue exporting data to the European Union would be to maintain the legislative level of protection they had been required to implement by 2018 (see also Kuner et al., 2016). However, plans for the new Data Reform Bill\textsuperscript{54} call for changes to the GDPR-compliant law which will likely change that status. By the time GDPR went into full effect, only a limited number of countries had been deemed “adequate” or would keep their pre-GDPR status (C. J. Bennett, 2018; Dove, 2018; Greenleaf, 2012). According to the European Commission (2021), as of the end of 2021, the only third countries who have been deemed “adequate” are Andorra, Argentina, Israel, Japan, New Zealand, South Korea, Switzerland, Uruguay, the UK and its dependencies (Guernsey, Jersey, Isle of Man), and the Faroe Islands (a Danish constituency).

In addition to the extended territorial scope of the GDPR, there are other new or extended requirements provided in the Regulation. The only entirely new principle is the Accountability Principle, which states that “the controller shall be responsible for, and be able to demonstrate compliance with” the principles of personal data processing (Art. 5(2)). It draws on the principles from the Directive, including rights of judicial remedy, claims of data processor liability, and sanctions in response to rights infringements. However, the Directive discussed

\textsuperscript{53} Article 45(1), GDPR states that “A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection. Such a transfer shall not require any specific authorisation.”

\textsuperscript{54} The replacement data protection law was first announced during the 2022 Queen’s Speech to open the British Parliament by Queen Elizabeth II (Queen Elizabeth II, 2022). The UK government published its response to the public consultation period (September 10-November 19, 2021) on 17 June 2022 ([https://www.gov.uk/government/consultations/data-a-new-direction/outcome/data-a-new-direction-government-response-to-consultation](https://www.gov.uk/government/consultations/data-a-new-direction/outcome/data-a-new-direction-government-response-to-consultation)). The responses from the consultations (public as well as targeted consultations with academic, industry, and consumer advocate stakeholders) will be considered during the revisions to the upcoming new law.
these in relation to what Member States governments “shall” implement, but there were no
details or explicit requirements and sanctions (Art. 22, 23, & 24). Here, the responsibility is
explicitly on the data controller to not just ensure but be able to prove that they and their
institution were abiding by the principles. This is much stricter, and imposes more explicit
responsibilities, than the Directive. The GDPR’s Accountability Principle imposes not just a
vague, moral obligation to respect and protect data subject rights on data controllers, but
significant monetary sanctions (Linden et al., 2019; Art. 83(4) & 83(5) GDPR). GDPR also
explicitly puts that responsibility on the data controllers, not just on the data processors. It is the
controllers who must demonstrate compliance, even when third-party data processors do the
processing, unless explicitly included in any data processing agreement (Chassang, 2017;
Lindqvist, 2018). This, Lindqvist (2018) argues, is bolstered by Article 28(2) GDPR, which
requires documentation of processing and data control.

The six principles of personal data processing are laid out in Article 5(1). They are:
lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage
limitation; and integrity and confidentiality, which are related to but an expansion on the
Directive’s Article 6(1). The principles that have received special emphasis in the GDPR are the
purpose limitation (Art. 5(1)(b)), data minimization (Art. 5(1)(c)), and storage limitation (Art.
5(1)(e)) principles. In so doing, they address the potential for harm resulting from both the
advancement and general use of technologies in data processing in several points in the data life
cycle. The purpose limitation principle55 directly effects data reuse in the European Union.
Custers et al. (2019) argue that this principle only allows data reuse when the reuse purpose is

---

55 Article 5(1)(b), GDPR states that personal data shall be “collected for specified, explicit and legitimate purposes and not
further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public
interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered
to be incompatible with the initial purposes (‘purpose limitation’)” (Art. 5(1)(b)).
compatible with the purpose for which the data were originally collected. This counters the excessive, all-encompassing approach to data collection that many businesses, apps, administrative, and research data collection methods take. That collection itself is governed by the data minimization principle,\(^{56}\) which explicitly calls for the minimum data necessary “only if (and as long as) the purposes cannot be fulfilled without personal data” (Tikkinen-Piri et al., 2018, p. 139).

Boeckhout et al. (2018) argue that data minimization still gives greater leeway for scientific and medical research as long as protection and limitation measures are put into place. However, as Wachter (2019) argues, even though the GDPR is intended to protect privacy and minimize harms, it does not regulate how the data are assessed to be the “minimum data necessary.” Once those data are collected, the storage limitation principle\(^ {57}\) allows for data stored for research, public interest, or statistical purposes to be stored for longer than is authorized as long as appropriate measures are taken. What “no longer than is necessary” means is left to the original data collectors, organizational obligations, and legal mandates for data retention to determine.

Article 9(1) prohibits the processing of the special categories of personal data. However, as with the Directive’s Article 8(2), Article 9(2) provides exemptions to that prohibition, recognizing that there are contexts in which the processing of these data would be appropriate. In general, these exemptions have been carried over into the GDPR directly from the Directive.

\(^{56}\) Article 5(1)(c), GDPR requires that personal data shall be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’).”

\(^{57}\) Article 5(1)(e), GDPR requires that personal data shall be “kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject (‘storage limitation’).”
There are some minor differences, such as the addition of “for one or more specified purposes” to the exemption based on explicit consent (Art. 9(2)(a)), where the Directive’s exemption simply stated “explicit consent to the processing of those data” (Art. 8(2)(a)). Also, where the Directive allowed for exemptions for substantial public interest (Art. 8(4)), the GDPR allows for non-public health processing for public interest only for “processing… necessary for archiving purposes in the public interest, scientific or historical research purposes of statistical purposes in accordance with Article 89(1)” (Art. 9(2)(j)). As will be discussed throughout Chapter 4, however, confusion remains about whether and how data archives may make use of “public interest” as well as other bases for data archiving, processing, and disseminating, and different legal systems allow for different applications of the “public interest” allowance. Finally, Article 9(4) allows that “Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health,” which many took advantage of.

Related to the obligation to have data subject consent are two further articles laying out data subject rights to the data about them. Article 15 gives the data subject the right to information from the data controller about why their data are being processed, the categories of data being processed, who has received the data, and the existence of a right to correct or erase their data. The right of the data subject to receive their data and to then transfer it to a different data controller (as might happen with citizen science projects, for example) is found in Article 20. This right to data portability in the GDPR is the first of its kind in the world, and as Wong and Henderson (2019) argue, all the more necessary due to the global nature of data sharing (see also Quinn, 2018). De Hert et al. (2018) call this right “one of the most important novelties” of the GDPR (p. 193). Hummel et al.’s (2018) evaluation of this right’s importance places the data...
subject at the center as a data sovereign, with the potential to control the contexts and use of data about them. As will be discussed in Section 4.3, the right of portability is also seen as a possible barrier for data archiving and research in general.

A key feature of the GDPR is Article 25, which calls for “data protection by design and by default.” This article explicitly puts the responsibility of implementing the data protection principles through technical and organizational safeguards on the data controller, who, per Article 5(2), must be able to demonstrate compliance. According to Article 25(1), this shall occur “both at the time of the determination of the means for processing and at the time of the processing itself”–data protection by design. Chassang (2017) sees this principle as carried over from scientific research settings, where sustainable data protection systems are established early in the data life cycle. In accordance with Article 25(2), “such measures shall ensure that by default personal data are not made accessible without the individual’s intervention to an indefinite number of natural persons”–data protection by design. Custers et al. (2019) recommend that, in addition to the new GDPR requirement for privacy impact assessments, organizations in the data economy conduct a data reuse impact assessment which would form the basis of the data protection by design. Another way that organizations demonstrate their adherence to Article 25 is through the creation of codes of conduct as recommended in Article 40.58 These principles, in addressing the role of technology in processing and in ensuring privacy through restricted access via that technology, seek to minimize the potential for harm through inappropriate data access and creating an environment where data protection is not just a one-off event but is the culture of data processing.

58 Article 40(1), GDPR states that “The Member States, the supervisory authorities, the Board and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized enterprises.” The following paragraph lays out possible features that could be included in these Codes of Conduct, which roughly follow the principles throughout the GDPR.
Article 32 describes the requirements for ensuring the security of the data. Acknowledging that institutions vary in their ability to implement new data protection and processing technologies as well as the variation in nature, scope, context, and purpose for data processing, the Article recommends, inter alia, the following four security measures: pseudonymization and encryption; the ability to ensure confidentiality, integrity, availability, and resilience of processing; the ability to restore availability and access of data if there are technical problems; and, a process to regularly test the effectiveness of technical and organizational security measures (Art. 32(1)(a-d)). This list is not all inclusive, nor is it the minimum required. It recognizes that, even as collecting, processing, and storage capabilities will evolve, that security and data protection measures will also evolve.

The Article 35 requirement for data protection impact assessments (DPIAs) explicitly addresses the privacy concerns stemming from technological advancements, both in terms of technological innovation and the use of extant technologies which are new to the organization. As Kloza et al. (2017) describe, the aim of impact assessments is “to address uncertainty and risk” (p. 1). Prior to the GDPR, the European Union had implemented voluntary privacy impact assessment policies—for radio-frequency identification applications in 2009—but the GDPR’s requirements for DPIAs are mandatory for all processing.

59 Article 35(1), GDPR, states: “Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.”

60 Commission Recommendation (2009/387/EC) of 12 May 2009 on the implementation of privacy and data protection principles in applications supported by radio-frequency identification [2009] OJ L 122/47. Preamble para. 6 recognizes the privacy concerns of this technology: “Because of its potential to be both ubiquitous and practically invisible, particular attention to privacy and data protection issues is required in the deployment of RFID. Consequently, privacy and information security features should be built into RFID applications before their widespread use (principle of “security and privacy-by-design”).”

61 Commission Recommendation (2012/148/EU) of 9 March 2012 on preparations for the roll-out of smart metering systems [2012] OJ L 73/9. Preamble para. 6 recognizes that the data processing by smart grid suppliers and operations would “evolve from a broad view of energy behaviour to detailed information on the energy behaviour of individual end-consumers.” Therefore, per Preamble para. 9, “data protection impact assessments should make it possible to identify from the start data protection risks in smart grid developments.”
which are likely to have significant privacy risks. These impact assessments do not exist in a vacuum; per Article 35(1), where the DPIA indicates a high risk without mitigation measures, the data controller is required to consult with the Member State’s data protection supervisory authority.

The GDPR also introduced a requirement for personal data processing organizations to appoint a data protection officer (DPO) (Art. 37 GDPR). The DPO advises controllers and processors about personal data processing, monitors compliance, advises about DPIAs, and communicates with the national data protection authorities (Art. 39(1) GDPR). Prior to the GDPR, many organizations did have staff who fulfilled this role though not as a formally appointed DPO (e.g., the Finnish Social Science Data Archive), or formally hired or contracted with DPOs (e.g., the Royal Netherlands Academy of Arts and Sciences and GESIS, respectively), but it was not a uniform requirement. While the GDPR does state that the DPO “shall be designated on the basis of professional qualities and, in particular, expert knowledge of data protection law and practices” (Art. 37(5) GDPR), the DPO is not required to be a lawyer. In practice, the DPO often is a lawyer or has a legal background, but they may also instead have experience on the data management or IT and data security aspect of GDPR compliance. There is no current requirement at the EU level that a DPO have a specific certification or qualification, though professional associations like Ireland’s Association of Data Protection Officers62 or the International Association of Privacy Professionals63 have started offering training courses and workshops targeting DPOs.

62 For example, the ADPO offers a one day course about the GDPR, as well as certifications for “Certified Data Protection Practitioner” and the “European Certified Data Protection Officer Programme” (see https://www.dpo.ie/training/adptrainingessentialsclassroom for access to other training offerings).
63 The IAPP offers ANSI/ISO-accredited CIPP/E certifications and a “GDPR Ready curriculum” targeting European data protection officers and others working with personal data (see https://iapp.org/train/cippe-training/ for more information on the IAPP training).
Data processing for secondary research purposes is not prohibited by the GDPR, even though there are significant challenges in ensuring that data subjects have given their consent to further data processing after the fact. Article 89⁶⁴ delineates the safeguards and derogations for public interest archiving purposes, as well as scientific and historical research and statistical purposes, drawing on the purpose limitation principle (Art. 5(1)(b)) which states that processing for research and statistical purposes shall not “be considered to be incompatible with the initial purpose” of data collection. Personal data processing for scientific and historical research and statistical purposes is identified in Article 89(2) as one situation where the national data protection laws can provide for derogations to data subject rights of access, rectification, processing restriction, and objection. Personal data processing for archiving in the public interest is identified in Article 89(3) as another situation where the national laws can derogate those same rights, as well as the right to data portability.

Article 9(2)(j) identifies these uses of data as an appropriate exemption to the processing of special categories of data prohibition in Article 9(1), while Article 17(3)(d)⁶⁵ lists these uses as appropriate exemptions to the right to erasure.⁶⁶ These exemptions, however, do require that

---

⁶⁴ Article 89(1), GDPR, allows for processing “for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.”

⁶⁵ Article 17(3)(d), GDPR, states that Article 17(1) and 17(2) shall not apply when processing is necessary “for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing.”

⁶⁶ Article 17(1), GDPR, states that: “The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies: (a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed; (b) the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and there are no other legal ground for the processing; (c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing; or the data subject objects to the processing pursuant to Article 21(2); (d) the personal data have been unlawfully processed; (e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject; (f) the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).”
appropriate safeguards and technical measures have been implemented to minimize the risks to
the data subjects. Additionally, Article 21(6) gives data subjects the “right to object to
processing of personal data concerning him or her, unless the processing is necessary for the
performance of a task carried out for reasons of public interest” for research and statistical reuse
of data, a right which raised concern at many research institutions (to be discussed further in
Section 4.3).

It remains to be seen just how effective the GDPR is at balancing personal data protection
and ensuring that data can freely move between Member States. The Regulation’s components
certainly are stricter, and it is longer and more detailed than the Directive. And, as a Regulation,
it is directly enforceable in a way that the Directive was not. It is important to recognize that the
GDPR, like the Directive before it, is approaching data protection and privacy from several
perspectives. There are articles focused on adapting to technological changes which often
concurrently focus on minimizing the privacy harms that may come with them (e.g., Articles 5,
6, 9, 25, 35). There are other articles which address the contextually appropriate use of the
personal data, both through restrictions on data processing and the empowerment of data subjects
to have some measure of control over their data (e.g., Articles 9, 15, 20). In the next section, I
will review how the GDPR has been implemented in the various national laws.

2.3.5.3 Implementations of the GDPR

European Union Member States were given two years to ensure that the GDPR was
implemented within their State. The easiest way to do this was to revise their existing data
protection legislation to meet the requirements of GDPR. Regulations override national laws
about the same issue or topic, and thus, to be enforceable, the national laws ought to be
consistent with, though not less strict than, the Regulation itself (Hestermeyer, 2014). While
there were fewer opportunities for the States to tailor their implementations as there were with the Directive, there are still articles wherein the States could and did add derogations and other exemptions that create inconsistencies between them. This section will briefly discuss the implementations across the European Union, and end with a description of certain components of the implementations in Germany, the Netherlands, Finland, and the Czech Republic, the locations of my field sites. The legislation described here reflect the Member State legislation in place by June 2020; both unofficial and official translations (sources of the translations are identified in the footnotes) of these laws into English from the relevant Member State authorities were reviewed.

As the European Commission succinctly put it in their report on the status of GDPR implementations one year after it went into effect, the goal of the GDPR was to have “one continent, one law” (European Commission, 2019, p. 2). At the time of that report (July 2019), three countries had not yet adopted updated national data protection laws to fully implement the GDPR: Greece, Portugal, and Slovenia. As of January 2022, both Greece 67 and Portugal 68 have adopted their new laws; the new Slovenian law is still being discussed, and its 2005 law remains in effect. The GDPR has also been incorporated (along with 68 other European Union laws) into

---


the European Economic Area Agreement as of July 2018; therefore, Norway, Iceland, and Liechtenstein have also implemented the GDPR in their own national legislation (European Commission, 2019; European Free Trade Association, 2018). One year after GDPR’s full adoption, the Commission found that there were four areas they felt more harmonization was needed on: ensuring the independence of national data protection authorities; limiting the ability of national legislation to allow additional conditions for personal data processing beyond what the GDPR required; addressing national restrictions of data subjects’ rights to their data; and reconciling the data protection with freedoms of expression with particular focus on not using data protection to “create[e] a chilling effect on journalistic work” (European Commission, 2019, p. 4).

A significant deviation in the Directive’s national implementations was how the different implementations transposed the key terms and definitions into their laws. These differences trickled down into later components of the laws, further distancing the national implementations from the original text and aims of the Directive. These definitions, and others, have also been included in the text of the GDPR under Article 4, and with limited updates to wording, all of the Directive’s definitions have been included essentially verbatim. Most Member States have explicit statements within their own laws which state that all or most of the definitions provided

---


by the GDPR apply without deviation, though, Denmark, the Netherlands, Portugal, and Spain neither included the definitions in their laws nor made it clear that they applied. All of the Member States which do apply or include the GDPR definitions to all of the definitions did so without significant deviation, with a few exceptions, which I will summarize below.

The incorporation of “controller” into the national laws is one where there are a few interesting differences. For example, the Czech law does not include “controller” in its list of terms included in “mutatis mutandis” or as a separate component of the law, even though that role is referenced throughout the law. The Hungarian law did not revise its previous definition from the 2011 Directive implementation beyond adding the potential for a “joint controller” to the definition. The Czech, German, Greek, and Italian laws did not include a definition for

72 Lov 68 Forslag til lov om supplerende bestemmelser til forordning om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger. (23 May 2018). (Act on supplementary provisions to the regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (Data Protection Act)). Retrieved from https://www.datatilsynet.dk/media/6894/danish-data-protection-act.pdf. Unofficial translation from the Danish Ministry of Justice.
75 With the exception of updating “Community” to “Union or Member State,” this definition is essentially unchanged from the Directive.
77 “Mutatis mutandis” is defined as “with the necessary changes having been made or with consideration of the respective different” (Thomas Reuters, 2020). In the case of the Czech Republic, it is implied that the definitions apply, those using the terminology used in the Czech law; however, how this is to be done for terms not listed in § 24(2) ZSOÚ is unclear.
“third party.” The GDPR’s definition of “consent of the data subject” builds on the Directive’s, adding in the requirement that consent must be unambiguous, a requirement which was included in many of the Directive transpositions, though not in the Directive itself. Whether the GDPR’s definition of “consent” was included, or how, varied more with this term than others. Austria, Italy, Ireland, and the Netherlands did not define it, though as with the Czech law and “controller,” consent is referenced through their laws. The German law also specifies that consent must be obtained “in a particular case” (§ 46(17)); thus, research consent forms, for example, must include separate consents for research participation and future data archiving (Section 4.4).

In contrast to the variety of transpositions of the Directive’s definitions, there is a great deal more consistency and harmonization of how the GDPR’s important terms defined. No individual definition has been brought into every national law completely verbatim, but where there are differences in the definition, for the most part, they are minor matters of wording. There also is not the variation in labeling of the terms seen in the Directive implementations, minimizing confusion when national laws are consulted. This consistency in definitions leads to greater harmonization overall in the application of the GDPR articles between Member States, as the terms given in the different official language versions of the GDPR aligned with the national data protection laws. However, the fact that several States do not define these terms and do not clearly indicate within their national laws whether and how the GDPR’s definitions apply does not

---

81 Article 4(11), GDPR, defines “consent of the data subject” as “any freely given, specific, informed and 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.”


leave lingering inconsistencies that will need to be clarified as the States and their institutions continue to adapt to the GDPR.

Despite this harmonization of terms, there are several articles which do provide opportunities for different implementations. Article 9(2)(a)-(j) provided the exemptions to the prohibition on processing special categories of data, while Article 9(4) allowed the States to add further conditions and limitations specifically for genetic, biometric, and health data processing. However, many states have included additional derogations and exemptions beyond what the GDPR intended (European Commission, 2019). The Dutch law, for example, provides exemptions for these data when processed for scientific, historical, and statistical purposes if consent is either impossible or the effort to do so is disproportionate (Art. 24(c)).

Articles 44-50 provide the criteria for third country personal data transfer adequacy decisions as well as the exemptions which allow for transfers to countries without an adequacy decision. For the most part, the national implementations have adopted this part of the GDPR without deviation. There are some differences, however, regarding the requirements to allow those transfers. In Cyprus,84 for example, any data transfers to a third country require both a data protection impact assessment and consultation with the data protection commissioner (Art. 18(1)), something not explicitly required in the GDPR for third country data transfers, whether they have an adequacy decision or not.

With regard to processing for scientific and historical research or statistical purposes, many Member States have gone beyond the requirements given in Article 89 of the GDPR. For

---

84 Ο περί της Προστασίας των Φυσικών Προσώπων Έναντι της Επεξεργασίας των Δεδομένων Προσωπικού Χαρακτήρα και της Ελεύθερης Κυκλοφορίας των Δεδομένων αυτών Νόμος του 2018 (125(I)/2018). (31 July 2018). (Law Providing for the Protection of Natural Persons with regard to the Processing of Personal Data and for the Free Movement of Such Data (Law 125(I) of 2018)). Retrieved from https://www.dataprotection.gov.cy/dataprotection/dataprotection.nsf/2B53605103DCE4A4C225826300362211/$file/Law%20125(I)%20of%202018%20ENG%20final.pdf. Unofficial translation provide by the Office of the Commissioner for Personal Data Protection.
example. Luxembourg\textsuperscript{85} requires a data privacy impact assessment (DPIA) before any such processing is carried out (Art 65(2°)), as well as that anonymization or pseudonymization be carried out by a third party “functionally independent of the data controller” (Art 65(4°)). Under the Finnish law,\textsuperscript{86} if special categories of data are to be processed for these purposes, a DPIA must be conducted and submitted to the national Data Protection Ombudsman (§31 para. 3).

Several laws incorporate recommendations from GDPR’s Article 32; the Czech law recommends, though does not require, a “process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing” (§ 16(j)), as does the German law (§ 22(2)(9)). The latter law also says that Article 15 of the GDPR “shall not apply if the data are necessary for purposes of scientific research and the provision of information would involve disproportionate effort” (§ 27(2)). Similar derogations exist in other laws. Unlike the previous laws which have extensive lists of recommended actions, the law in the Netherlands does not provide examples of recommended or required safeguards, simply allowing for processing in difference contexts as long as safeguards are such that “the data subject’s privacy is not disproportionately compromised” (e.g., § 23(b), § 24(d), § 28(2)(b)(2°)).

I will finish this section with a brief discussion of the GDPR implementations in Germany, the Netherlands, Finland, and the Czech Republic, reflecting their status as research sites for my dissertation. In terms of how key terms are defined, as discussed above, the Dutch law does not include the definitions or an explicit statement that the Article 4 definitions apply in


the law. In the Czech law, with the exception of the undefined “controller” as discussed above, the Article 4 definitions apply *mutatis mutandis*, though how this occurs in practice is unclear. The Finnish law states that anything in the GDPR not included in the text of the Finnish law applies directly; the definitions are included in this category. The German law incorporates the GDPR’s definitions essentially verbatim, with the exception of the undefined “third party.” In all four cases, the terms are used throughout the national laws, leaving some ambiguity in the exact meaning and intent of the following articles.

Under the German law, a data protection officer must be designated if at least 20 people “[deal] with the automated processing of personal data” or if any personal data processing which requires a DPIA is conducted, no matter the number of people involved (§ 38(1)). The Netherlands, Finland, and the Czech Republic do not have any additional requirements for a data protection officer beyond the GDPR’s Article 37, which does not provide a minimum employee requirement for designating that position. In the Czech Republic, this is an entirely new requirement; even informally there was rarely someone acting in this role, in contrast with the pre-GDPR situation at the Finnish Social Science Data Archive and DANS, for example (Sections 4.2 and 4.3).

While not provided in the law itself, the Czech Data Protection Authority provides a detailed list of which data processing operations do and do not require a DPIA (Úřad pro ochranu osobních údajů, n.d.). The German BDSG does provide guidance for when a DPIA must be conducted and what should be included (§ 67), and a similar list of operations requiring a DPIA.

---

87 Article 37(1), GDPR states that “The controller and the processor shall designate a data protection officer in any case where: (a) the processing is carried out by a public authority or body, except for courts acting in their judicial capacity; (b) the core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale; or (c) the core activities of the controller or the processor consist of processing on a large scale of special categories of data pursuant to Article 9 and personal data relating to criminal convictions and offences referred to in Article 10.”
has been provided by the Conference of Independent German Federal and State Data Protection Supervisory Authorities (Datenschutzkonferenz, n.d.). However, as will be discussed in Section 4.4, German state-level data protection legislation differ on when a DPIA is required; for example, the Datenschutzgesetz Nordrhein-Westfalen, the state law for one of GESIS’ locations, does not require a DPIA if one has previously been completed for a service that has not changed, while the BDSG and the state law for Baden-Württemberg (the other GESIS location) do not have such exemptions. These inconsistencies between the federal and state-level laws, and between the state laws themselves, caused and continue to cause challenges for German organizations operating across the entire country. There is no such current delineation in the Dutch law, nor does it discuss DPIAs in any way. However, in November 2019, the Dutch Autoriteit Persoonsgegevens published their list of processing operations which require DPIAs, based on their GDPR implementation as well as their criminal justice legislation. Under the Finnish law, a DPIA is included on the list of possible measures to safeguard the rights of the data subject (§ 6(2)(10)), though there is no indication that it is required except for what the GDPR mandates.

Article 89(2) GDPR allows national legislation to limit data subject rights when personal data are processed for scientific and historical research and statistical purposes. The German law utilized several derogations related to personal data processing for research. For example, § 27(1) allows for processing special categories of data for these purposes without data subject consent as long as the processing is “necessary for these purposes and the interests of the controller in processing substantially outweigh those of the data subject in not processing the data.” The GDPR rights of access (Art. 15), rectification (Art. 16), restriction of processing (Art. 18), and objection (Art. 21) can also, according to § 27(2), be limited if these rights “are likely to
render impossible or seriously impair” the research. § 16(3) of the ZZOÚ allows for similar limitations in the Czech Republic, and § 31 of the Finnish law allows the same. The Dutch UAVG (§ 44) specifically allows limitations on Articles 15, 16, and 18 (but not 21) GDPR for research and statistical purposes.

In terms of processing for scientific and historical research and statistical purposes, the Czech, Finnish, and German laws both list similar recommended, not-all-inclusive data protection measures to be taken. The Czech law separates these requirements into their own section of the law specific to these purposes (§ 16(a-l)), while the German law refers the reader to § 22(2), which provides the recommended safeguards for the derogations to processing special categories of data. The recommended measures are similar, with minor differences.

Additionally, the Czech law recommends limits on transmitting the data for these purposes to third countries (with no further clarification) (§ 16(k)). There is no related limitation in the German law. In contrast to the lists of measures, the Dutch law allows for processing special categories of data for these purposes only if the processing is in accordance with GDPR 89(1), if it serves the public interest, if it is impossible or requires disproportionate effort to get consent, and safeguards are in place (§ 24). All four of these are required, and no specific details in the law clarify the required safeguards. The Finnish law allows for sharing personal data with third countries and international organizations against Articles 44-50 GDPR if restricting such processing and transfer “would infringe on the right to freedom of expression or information” (§ 27).

---

88 These seven articles of the GDPR cover the principles for international data transfers (Art. 44), transfers based on the adequacy decisions (Art. 45), transfers subject to safeguards when there has been no adequacy decision (Art. 46), the binding corporate rules for third country data transfers (Art. 47), the requirement for international agreements to support court or tribunal ordered third country data transfers (Art. 48), derogations for third country data transfers when Article 45(3) or Article 46 (Art. 49), and the European Commission and Supervisory Authorities development of international cooperation mechanisms for these data transfers (Art. 50).
In terms of personal data archiving, a particular consideration for the archives of interest in this research, Article 89(3) GDPR allows national laws to authorize derogations from Articles 15, 16, and 18-21 GDPR when that archiving is in the public interest (and allows the Member States to determine what “public interest” is). The Finnish law used this derogation to allow for personal data archiving in the public interest (§ 32), which, as will be discussed in Section 4.2, is one of the bases FSD uses to support personal data archiving when informed consent was not the legal basis for data processing. § 28 BDSG implemented this derogation as well, though § 28(3) does mandate that, even when the right of rectification (Art. 16 GDPR) is limited at German institutions, a data subject is allowed to provide a correction which must be archived along with the original version. The Dutch UAVG limits the Article 15, 16, 18(1)(a) and 20 GDPR rights in this situation. The Czech law also allows for some limitations to Articles 18, 19, and 21 (in § 20, § 21, and § 22, respectively), though it clarifies that meeting these exercised GDPR rights can also be limited if so doing requires “disproportionate effort.”

2.3.5.4 Limitations of the GDPR

Taking into account the new and extended protections, this section will present the impact and limitations of the GDPR. The biggest change from the GDPR by far is, as previously discussed, that as a regulation, it is directly applicable to the Member States. Therefore, the biggest impact must be the standardization that it imposes (Albrecht, 2016; Aldhouse, 2018; Mondschein & Monda, 2019). As Albrecht wrote before Brexit and the full implementation of GDPR, the regulation would “bring more legal certainty and coherence than [in 2016], where 28 different legal systems as well as 28 different judicial and enforcement cultures define the regulatory environment” (2016, p. 288). As discussed above, there are still some allowances for distinct national implementations, including certain derogations of the requirements, therefore
the desired harmonization and standardization is not total (Kuner et al., 2017; Phillips & Knoppers, 2019). However, within the European Union, Junge (2018) argues that there will be greater transparency and protection which will reduce the effect of the Directive’s limitations. And, because of the extended scope of the protections, applicable outside of the European Union as long as the data are of Member State residents, those principles will in time become the norm globally (Albrecht, 2016; Bhaimia, 2018; Kuner et al., 2017; Mayer-Schönberger & Padova, 2016).

When it comes to research as a use of data, Article 5(1)(b) identifies processing for scientific or historical research as not being incompatible with the initial data collection purpose. Article 89 further identifies obligations for safeguarding (technical as well as organizational) to support that research use, and allowances for Member States to allow derogations of data subject rights to support research:

Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner (Art. 89(1)).

In addition to pseudonymization, Article 6(4)(e) also identifies “encryption” as a possible appropriate safeguard; however, beyond these two suggested safeguards, the GDPR leaves considerable flexibility for how these safeguards, whether technical or organizational, are actually put into action. GDPR also does not explicitly define what type of research falls under the umbrella term “scientific”; however, Boeckhout et al. (2018) argue that in not doing so, the GDPR “leaves more leeway for scientific research than for other forms of data processing” (p. 934). Much of the discussion in the literature so far about the impact of GDPR on research
focuses on health and biomedical research (e.g., Dove, 2018; Jacobs & Popma, 2019; Marelli & Testa, 2018; Stockdale et al., 2019). Chassang (2017) identifies research (both medical and other types of research) as a special type of data processing where ethical issues are triggered by the need to balance rights to privacy and freedoms to do research. Failure to comply with the GDPR requirements can lead to loss of research funding, fines, and other legal penalties individual Member States can impose.

There are also several critiques of GDPR as it relates to research. The first refer back to earlier concerns about informed consent (e.g., Corti et al., 2000; Law, 2006), particularly the difficulties inherent in identifying potential future data uses which data subjects can consent to. Van der Ree et al. (2019) have reported that, in the Netherlands, as a result of GDPR, obtaining consent for new research uses has become more difficult. The Dutch requirements for informed consent for the use of medical records, for example, allow only the original treating physician to contact patients to request their consent to share their medical records for research purposes. This is problematic in the case of patients who have relocated, as under the GDPR, this physician’s access to municipal records (to obtain the new address) is prohibited. Elsewhere in the European Union, data intensive industries, to include health, biomedical, and administrative data research, were also worried about this consent problem (Mourby et al., 2018; Phillips & Knoppers, 2019; Stevens, 2015). Explicit fears are not clearly defined in the literature, though many do revolve around multinational research collaborations and the ability to do data-driven research if Member State GDPR implementations do include variations, especially in the case of derogations to the requirements (van Veen, 2018; Vestoso, 2018). Webb and Sargent (2018), Härting and Gössling (2018), and Kuzio et al. (2022) after the GDPR join Barocas and Nissenbaum (2014) in highlighting an overreliance on consent as a *the* method by which privacy
is protected as an on-going concern that the GDPR perpetuates, even as it allows for other legal bases for processing. While from an ethical standpoint, informed consent for participation in research is a necessity, consent for other processing, such as for using online services, can be influenced by how the consent form is presented to a user who is faced constantly with consent requests (Utz et al., 2019).

Another important critique of GDPR is just what is meant by “public interest,” an exception provided along with research as not being incompatible with the original data collection. At a data protection roundtable after GDPR went into effect, Demotes-Mainard et al. (2019) found that clarification was still needed about just what public interest actually refers to, and to what contexts it is applicable. As an example from before the GDPR, Joyce (2011) questioned how to balance public interest in information about criminal offenders and the privacy rights of child offenders. When some nebulous concept of public interest is given specific context, the power of public interest as a motivator for data sharing which leads to privacy violations comes into question. Gavison (1992) also critiqued the defense of “legitimate public interest” as a valid reason for data processing even before the Directive was adopted. Van Deursen and Kummeling (2019) questioned what Article 9(2)(g), which provides an exception to the prohibition of processing of special data categories for “substantial public interest” (separate from public health interest, which are covered under Article 9(2)(i)), actually means, given this lack of definition.

The data protection roundtable also identified the need to clarify what “anonymization” means under the GDPR (Demotes-Mainard et al., 2019). GDPR is only applicable to personal data which are identifiable; this includes data which have been pseudonymized because if the subject could be re-identified “by the use of additional information, [the pseudonymized data]
should be considered to be information on an identifiable natural person” (Preamble para. 26). This proved particularly challenging for German institutions, which since 1987 have utilized a “looser” interpretation of “anonymization” than the GDPR, and if that interpretation was no longer legal, the impact on German institutions, including its data archives and research institutions, would have been quite severe (Section 4.4). The GDPR does recognize that pseudonymization can reduce the risk to data subject privacy by lowering the ability to link the data, but pseudonymization is not itself a perfect solution to address the risk of re-identification (Mourby et al., 2018; Ohm, 2009; Van den Eynden et al., 2015). Therefore, anonymization is seen (both in reaction to the GDPR and in other data protection discussions) as just one of the methods of data protection the enable to use of data outside the purview of GDPR. However, “full anonymization,” or even “adequate anonymization,” itself raises the risk of the data no longer being of value for reuse (Andersson & Sørvik, 2013; Carusi & Jirotka, 2009; Haeusermann et al., 2018; Jarolimkova & Drobikova, 2019; Kleiner et al., 2011; van Veen, 2018).

Wachter (2019) further argues that “truly” anonymized data can still be used for discriminatory practices, through the development of community or user profiles even if specific individuals are not themselves identified. GDPR does not clarify what constitutes pseudonymous or what is sufficient to make GDPR not applicable to a given data set. Zibushka et al. (2019) argue that anonymization is not feasible to achieve data protection and that the discussion about data protection techniques should instead focus on other privacy techniques. Mourby et al. (2018) also argue against using a definition of “pseudonymization” as the foundation for data being protected, instead recommending a two-stage process that first identify
if, using all reasonable methods, natural persons are identifiable, and then, if yes, if pseudonymization has been applied in accordance with Article 4(5) GDPR. 89

Other limitations of the GDPR have been identified since it was first proposed in 2012. Koops (2014) found the process of data protection reform in the European Union to be flawed because it was focusing too much on solving the information and communication technology challenges in a single legal framework and not enough on the data processing practices. He identified three fallacies in the then-forthcoming GDPR: objectives that data protection laws will give data subjects control of their data; that the processes will be simpler; and, that a comprehensive, omnibus law like the GDPR is the best solution. Following Koops’ criticisms, in the post-GDPR era Wachter (2019) argues that the GDPR focuses too strongly on the “input stage” of data collection, and ignores the privacy risks that come from the inferential and algorithmic processing. Waldman similarly criticizes organizational GDPR compliance efforts as merely a “focus on records and documentation [that] offers a convenient veneer of legitimacy to a process of technology design, data use, and information flow that remains unaltered and harmful to consumers” (Waldman, 2021, p. 142). Dove (2018) criticizes the GDPR for continuing to “regulate data on the basis of its form (‘personal’ and ‘special’) and content, rather than its function (i.e., regulating on the basis of use and whether the processing activity is capable of identifying or individualizing a data subject)” (p. 1016). This is best seen in the revision to the Directive’s Article 1(1), which gave one objective of the Directive as “… Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.” The GDPR’s

89 Article 4(5), GDPR defines pseudonymization as “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”
corresponding objective removed the reference to the “right to privacy” and instead says that the GDPR “protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data” (Art. 1(2)).

The consensus about the GDPR seems to be that, relative to the Directive, it or something like it was necessary to correct the limitations stemming from the Directive. The advancement of computing and communications technology in just two decades resulted in a new environment where the lack of standardized national Directive transpositions was creating, or would soon create, confusion and road blocks for the data sharing necessary for research and other data processing objectives. However, significant limitations and confusion remain or were created by it, and therefore, risks to data subject private as well as questions about what types of data sharing are allowed remain. It has been four years since the GDPR went into effect, which makes a study of the effect of the Regulation and these limitations on data sharing at European Union data archives particularly timely.

2.4 Data Archives and Organizational Change

The data of interest in this dissertation are those data used in the social sciences, specifically those archived at social science data archives. These data are often surveys, including national censuses and other research surveys. They can also be qualitative data, in the form of interview transcripts, documentation, and field notes which, according to Corti (2011), “capture lived experiences of the social world and the meanings people give these experiences from their own perspectives.” The key to these data, and what makes their archiving, sharing, and reuse challenging, is that the data are about people. Often, the very characteristics of data subjects that make them interesting and valuable for primary and secondary research are those characteristics which leave the data subjects potentially vulnerable to harm (see, Art. 9(1)
GDPR). In the next section, I will begin by introducing data archiving. Then, in Section 2.4.2 I will discuss data the current research on organizational changes in response to the GDPR. I will end in Section 2.4.3 with a discussion of the privacy concerns facing research data archives.

2.4.1 Data archiving

Archiving research data is not a new phenomenon of the 21st century—for example, the Inter-university Consortium for Political and Social Research (ICPSR) was founded in 1962 (Inter-university Consortium for Political and Social Research, 2019), the predecessor of Germany’s GESIS-Leibniz Institute for the Social Sciences was founded in 1960 (GESIS Leibniz Institut für Sozialwissenschaften, n.d.b), while in Austria, the Wiener Institute for Social Science Data Documentation and Methodology was founded in 1985 (Smioski, 2011). In Green and Gutmann’s social science research life cycle, the data archive has a key role in the final two steps of the cycle: publication and sharing, where the researcher shares data with the wider community through the medium of the data repository or archive, and long-term management, where the archive’s role is to both ensure that the data are “exposed for use and learning by others” and the data’s long-term preservation (2007, p. 38).

However, the last decade has witnessed a growing call not necessarily from researchers wanting to deposit their data but rather from governments, funders, and journal publishers requiring data deposit in data archives or repositories as a pre-requisite for funding and publication (Bishop & Kuula-Luumi, 2017; Bishop & Neale, 2011; Holdren, 2013; Information Commissioner’s Office, 2011; National Institutes of Health, 2003; Sterett, 2019). In practice it is still relatively new for many disciplines, especially when it comes to qualitative data (Cheshire et al., 2009). There are many benefits to archiving data: so that research findings can be validated, that data will be kept secure and safe long-term, to reduce the research burden on data subjects,
encourage scientific innovation, meet funder requirements, recognize the public right to publicly funded research results, maximize the utility of data, promote transparent research, and train the next generation of researchers (Antes et al., 2018; Arzberger et al., 2004; Bishop, 2009; Bishop & Neale, 2011; Carlson & Anderson, 2007; Carusi & Jirotka, 2009; Hammersley, 1997; Karcher et al., 2016; Mauthner & Parry, 2009; Moedas, 2015; Pasquetto et al., 2017; Tenopir et al., 2011). These benefits and calls for researchers to make their data publicly available have also grown alongside greater data subject desire to control their own data and information.

For that, however, there needs to be some form of data archiving and data sharing infrastructure, something which has only slowly been developing across Europe. In many cases, that slow development or lack of archival infrastructure is related to the research practices and cultures related to data sharing, specifically, a lack of common practice. In Poland, researchers are reluctant to share and reuse data, which has a direct effect on funding and support for archival infrastructures. They have concerns particularly with secondary users’ lack of personal experience from collecting the data they would reuse, raising methodological concerns about the quality of research that the reusers could do with those data (Binder & Filipkowski, 2011). Danish researchers only share qualitative data informally, if at all, though there is archival infrastructure for quantitative data for social science, medical, and history researchers (Kjeldgaard, 2011). Titarenko and Tereschenko (2011) describe the Belarusian research culture as closed, and therefore “data sharing is not prevalent as there are no traditions or official (or professional) policies to propagate research information” between researchers or through archives which provide access to archived data (p. 60). In the Czech Republic, even though the Czech Social Science Data Archive has existed since 1998, researchers have been slow to access data through it (Čížek, 2011). In some countries, such as Germany, quantitative data archiving
and sharing is better developed than qualitative data (Medjedović & Witzel, 2011). Over the last decade, the Czech Social Science Data Archive has seen an increase in data set holdings (Český sociálněvědní datový archiv, 2020; Jarolímková & Drobikova, 2019; Vávra, 2019), while the Danish Data Archive, now a collection of survey data within the Danish National Archives, as of 2013 has made that survey data available for reuse (though with limited qualitative data) (Kjeldgaard, 2011; Rigsarkivet, n.d.; UK Data Service, 2020).

Data archives and repositories offer more than just long-term storage and preservation of research data, however. They also provide access to the data held there; access to data can be unrestricted or moderated through various levels of restricted access to identifiable data or to anonymized versions of the risky restricted data. The increasing prevalence of online digital data archives further contribute to increasing access and reuse of research data (Borgman et al., 2019). While many archives do offer the opportunity for researchers to self-deposit their data,90 wherein it is the responsibility of the researcher to clean, format, and describe their data sets and documentation in return for a very quick addition of the data to a publicly accessible archive, the more common process places that responsibility on archive staff. As researchers share their data with archives, archive staff review the data deposit, create the metadata record, and verify any required anonymization or pseudonymization for the given data access method, among other processing activities. Archive policies and practices surrounding data acquisition and ingest, including any curation to the data after the data are deposited, differ in a number of significant

---

90 For example, openICPSR is the self-deposit data archive at ICPSR, whereby “ICPSR does not approve or alter data sets deposited in openICPSR in any way. Data are distributed in the same condition and format submitted by the depositor. Data approved for public use are distributed at no cost to users. ICPSR may charge a fee for the costs of managing data that have been restricted to protect confidential information about research subjects.” It can be accessed at https://www.openicpsr.org/openicpsr/. The Archive of the Indigenous Languages of Latin America (AILLA) provides the AILLA Self-Deposit Tool for researchers who do not have the funding to archive their data with archive support and curation. It can be accessed at https://aila.utexas.org/site/depositors/Self-Depositing. At these archives, data depositors clean, describe, structure, and edit their data. GESIS also offers the self-deposit system SowiDataNet|datum for researchers to self-archive their data and impose one of four different access categories, in addition to the Data Archive.
ways which have ramifications for data reusers, especially those who want to do comparative research with data from multiple archives (e.g., Levenstein et al., 2018; Tyler, 2020). Depending on their capabilities, some archives archive and provide access to both de-identified data and identifying/identifiable access files for the data, which allow certain data reusers to access the identifiable personal or special data under controlled access conditions. These archives, which include ICPSR, DANS, GESIS, and the United Kingdom Data Archive, regulate and restrict access across different access levels through data access requests, data use agreements or other contracts establishing penalties for misuse of the data from the archive.

Whether archives provide access to identifiable data under specified access conditions or allow researchers to self-deposit data into a public archive, concerns remain about the privacy risks to data subjects if those data are combined with other data. Some archives, such as the Finnish Social Science Data Archive and the Czech Social Science Data Archive, do not provide access to identifiable data, and have implemented archive policies for anonymization prior to data ingest. How the archive determines an appropriate level of anonymization and who is responsible for ensuring it, however, remains up to the archive. One of the critiques of the GDPR has been the reliance on anonymization for data protection, because anonymization can limit the usability of the data, reinforcing the challenges in protecting privacy while promoting data use.

To ensure compliance with the GDPR, data archives likely would have to update their own data acquisition policies about what data can be acquired, what condition the data can be

---

91 For example, ICPSR provides access to publicly-accessible data with no restrictions, as well as data through increasingly restrictive dissemination methods up to a secure physical enclave, based on a number of data features including the risk to the data subjects, funder requirements, and so forth. Data at DANS can be open to the public, requiring registration, have restricted access requirements, or other limitations based on the data, funder, or researcher concerns. GESIS data in the Data Archive are categorized into categories 0, A, B, and C, where the accessibility is based on depositor permission to access a given data set. The UK Data Archive offers access to data at “open,” “safeguarded,” and “controlled” data access levels, where the safeguarded data are potentially disclosive if combined with other data, and the controlled data are those which are disclosive on their own.
deposited in, and how the data can then be accessed and reused. What form these changes took and what effects they had on social science data archive operations has not been studied. In the medical and health research fields, in contrast, the impacts of the GDPR have already been reported. For example, the GDPR has impeded international research conducted by the United States’ National Institutes of Health and European partners relying on European donor samples for cancer and diabetes studies (Eiss, 2020). Van der Ree et al. (2019) similarly report GDPR-related barriers to recruiting participants in medical research because of restrictions on who can contact a potential participant about consenting to take part (the treating physician, and not a researcher). This is one of the motivations for this dissertation—to understand what happened at these data archives both within the archives themselves and to their operations. The flexibility that GDPR allows for research means that, just as national legislative implementations are not fully standardized across Member States, research institutions and data archives working with personal data may not implement standardized, consistent policies either (Bell et al., 2019; van Veen, 2018). Even with multi-institutional and multi-national organizations like the Consortium of European Social Science Data Archives (CESSDA)92 whose aims include the implementation of national standards for data archives, Member States and institutions do not necessarily approach their data acquisitions and data provision roles in the same way (Statutes of CESSDA, 2017, Art. 3(5)(d)).

2.4.2 GDPR and organizational change

As will be discussed in the individual case reports in Chapter 4, while the GDPR brought clarity to some data protection and privacy concerns held over from the pre-GDPR regime, it

also caused significant confusion for researchers and the data archiving community. As discussed in Section 2.3.5, one of the aspects of the GDPR where individual Member States could derogate from the GDPR’s requirements is in terms of personal data processing and archiving for scientific and historical research or statistical purposes (Art. 89 GDPR). Because of this, research institutions could not simply review the GDPR and make changes; they needed to wait for the national data protection laws to be adopted. This process was complicated as well because not every national law addressed research data processing in the first publicized versions of the new laws (as will be discussed in Section 4.3, this was a particular challenge for DANS in the Netherlands, for example). Therefore, while plans for potential changes could be created, and some changes made relatively early in the process, many institutions found themselves in an uncertain position, without clear guidance from other similar institutions, parent organized, or national entities about what was necessary to become GDPR compliant.

Since the GDPR was first announced, there have been numerous discussions about what it would mean for research across many disciplines and what challenges the laws, and potential differences between the GDPR and national laws, might or have already resulted in. For example, in genetics research and biobanking (Boeckhout et al., 2018; Molnár-Gábor et al., 2022; Šolc, 2020), medical and health research (Chassang, 2017; Demotes-Mainard et al., 2019; Jacobs & Popma, 2019; Marelli & Testa, 2018; Phillips & Knoppers, 2019; Rumbold & Pierscionek, 2017; Stockdale et al., 2019; van der Ree et al., 2019; Yuan & Li, 2019), anthropology (Yuill, 2018), social media research (Breuer et al., 2020), use of administrative data in the social sciences (Mourby et al., 2018; Stevens, 2015), big data research (Mayer-Schönberger & Padova, 2016; Vestoso, 2018), and administrative records management at research institutions (Popovici, 2018). Several of these have also focused on concerns around
the increased strictness regarding informed consent as a legal basis for processing. For example, Yuill’s discussion of the GDPR’s impact on anthropological research highlighted that archiving data in this field requires “a more thorough consent process, which must now detail exactly what participants are giving consent for and must encompass participation, personal data usage and future information usage by others” (2018, p. 38).

One of the rare discussions about GDPR considerations for data archiving can be found in a volume about urban drainage systems and stormwater management, where the authors identify data about sewage users, used water quantities, user habits, water user commuting patterns, and location of data collectors as the personal data to be protected in data archives and databases in this field (Prodanović & Branisavljević, 2021). However, even in that work, they do not discuss specific changes or concerns at a specific data archive. Lee et al.’s (2022) report on the progress of a project to set up a data archive for corpora of speech and communication disorders likewise indicates that they held workshops which included GDPR experts, recognizing that there are privacy risks in archiving and sharing their data, but did not report on what specific concerns they had or how they affected the development. A related work reporting on the CLARIN ERIC93 development refers to GDPR-related ethical considerations, without identifying them or discussing how they have impacted the ERIC’s development (van den Heuvel et al., 2020).

There have also been a number of reports of how commercial businesses have responded to the GDPR (e.g., Freitas & Mira da Silva, 2018; Sørøm & Presthus, 2020; Urban et al., 2019). However, as highlighted above, there have been no studies about how social science research

---

93 The CLARIN (Common Language Resources and Technology Infrastructure) European Research Infrastructure Consortium (ERIC) supports the use of language data and tools for humanities and social science research and of the Digital Research Infrastructure for the Arts and Humanities (DARIAH). Staff at DANS, one of my research sites, are heavily involved in the various work packages developing this infrastructure.
infrastructures themselves have responded, even though they “are leading the way toward overcoming barriers to data access, within and across countries” (Farago, 2014, p. 29). One result of this is that there were no social science research infrastructure-focused exemplars or models of organizational responses to the GDPR that the data archives could look to, or that their later decisions could be compared against. Almeida Teixera et al. (2019) identified these non-existent exemplars and guidance as one factor that could negatively affect how efficiently and effectively an organization adapted to the GDPR. Therefore, a discussion of possible changes or factors for success in responding to the GDPR must draw on literature from other disciplines and industries.

Brodin (2019) presented a three-phase framework to support small- and medium-enterprise GDPR compliance: analysis→design→implementation. The analysis phase comprised analyzing the organizations information and information flows, updating how personal data were classified under the GDPR, identifying the legal bases for personal data processing, and evaluating existing IT security measures (Brodin, 2019, p. 253). The design phase comprised updating organization routines and procedures for personal data processing, creating procedures for responding to data subject rights request, revising or creating processes for responding to data breaches, updating policies, and create standardized templates for these processes and procedures (Brodin, 2019, p. 254). In this phase, Brodin also highlighted the need for a consistent template for a processing agreement for all processes where the organization is or uses a processor. The implementation phase then should focus on creating the organization climate for compliance through the new processes, establishing or appointing staff to compliance roles, and addressing the existing pre-GDPR data to bring it to GDPR compliance. While testing his framework, Brodin (2019) identified a number of specific changes or processes his three
cases instituted: risk assessments (including data protection impact assessments (DPIAs)), mapping information flows, updating processes and policies to be GDPR compliant (including privacy policies, responding to data subject requests, and processing and storage policies), appointing a data protection officer, implementing processing agreements and processing registers, and creating and conducting GDPR training.

Lopes and Oliviera’s (2018) model of the stages of GDPR implementation also follows three stages: gather→analyse→implementation. In this model, the “gather” stage is where the organization gathers information about all personal data and personal data processing in the organization, current or potential data sharing, agreements, retention periods, etc. The “analyse” phase identifies where there may be GDPR non-compliance and develops the implementation plan. The “implementation” phase is where policy, process, and technical changes are made, and training is provided. They also present an Accountability Life Cycle, modeling compliance maintenance in line with the GDPR’s Accountability Principle (Art. 5(2) GDPR); this model directly follows the GDPR implementation: 1. prepare, 2. operate, and 3. maintain (Lopes & Oliveira, 2018). This model also identifies specific activities aligned with each phase. For the “prepare” phase, the organization should: establish a GDPR readiness team, identify and assess third party processing, establish a processing register, distribute updated policies, and conduct internal and external training. The “operate” phase activities include distributing external privacy policies, documenting the legal bases for processing, processing data subject rights requests, and managing data breach incidents. The final phase, “maintain,” involves demonstrating understanding of data protection policies, maintaining the process register, conducting DPIAs when necessary, and demonstrating compliance and effectiveness of data processing practices (Lopes & Oliveira, 2018). The authors applied these models in their
analysis of medical clinics in Portugal, and determined that “the implementation of the [GDPR] implies the definition of procedures, records and policies” and that “both people and technologies represent critical success factors” (Lopes & Oliveira, 2018).

Sirur et al. (2018) interviewed senior personnel working in data protection, IT security, and other roles directly connected to GDPR compliance at a number of United Kingdom-based organizations. They found that the key activity for GDPR compliance efforts was “data flow mapping,” because “without understanding where their data was transmitted and stored, organisations felt they could not hope to have enough control over their data to protect it” (Sirur et al., 2018, p. 92). Other important factors included data protection training, to counter what their respondents considered to be a lack of general awareness across entire organizations, and compliance verification measures, such as maintaining clear records, data mapping, and monitoring the outcomes of changes.

Almeida Teixera et al. (2019) conducted a systematic literature of thus far published studies and explorations of GDPR implementations across European Union organizations, and developed two lists of what they referred to as “critical success factors” for GDPR implementations. The first list represents “enabling” success factors: an organizational implementation roadmap, awareness and analysis of the GDPR, identification of organizational risks, data management, documenting processes, the presence of a data protection officer, data security and privacy protection measures, and training (Almeida Teixeira et al., 2019, p. 416). The second list represents barriers to a successful GDPR implementation: the extensive and complex nature of the GDPR, the subjectivity in how organizations can comply with the GDPR, lack of privacy-related expertise, lack of financial and staff resources, lack of necessary technology, lack of guidance and standard procedures (Almeida Teixeira et al., 2019, p. 416).
Many scholars have modeled the mechanisms of organizational change. Van de Ven and Poole (1995) present four broad categories of process theories that explain how organizations develop and change over time: life-cycle, teleological, evolutionary, and dialectic. Of these, teleological models are likely the best fit for understanding the mechanisms by which organizations approached adapting to the GDPR. These models are targeted at addressing a specific goal, implementing planned changes, evaluating the results, and making modifications based on those evaluation. Martin et al. (2009) and Sune and Gibb (2015) highlight the end-state goal setting nature of teleological approaches, and the flexibility to not require specific sequences of events (in contrast to life-cycle models) to achieve the goal. The mechanisms of GDPR adaptations at organizations, including data archives, may thus be considered as a teleological process: goal setting (becoming GDPR compliant)→implement goals (develop and/or change policies, processes, and mindsets)→dissatisfaction with results (through evaluation and auditing of outcomes)→modification (new developments and/or changes to address results of evaluations).

Returning to the intentions of the GDPR for a moment, the GDPR is intended to standardize the data protection processes across the EU and resolve the increasing variation stemming from the different Directive transpositions. Presuming that organizations follow a teleological approach to their responses, we must question then whether organizational operations and approaches to personal data protection indeed converge. But, what is convergence? Does convergence equate to an identical outcome, or simply increasing similarity and uniformity? Kerr defines convergence as “the tendency of societies to grow more alike, to develop similarities in structures, processes and performances” (Kerr, 1983, p. 3), while Dunlop et al. (1975) consider convergence to be a process of change, and not a definitive end state.
Thus, the GDPR’s goal of data protection harmonization has convergence as a goal, but written into the GDPR are opportunities for divergence (through the articles allowing for national derogations discussed in Section 2.3.5). There have yet to be any studies of social science data archive adaptations to the GDPR; therefore, this dissertation will be the first to investigate whether this desired harmonization has occurred in this personal data processing context.

2.4.3 Privacy concerns for data archives

In this last section, I will bring together the discussions of privacy risks and the data archive, to identify the necessity of regulations about data that prioritize the safety and privacy of data subjects. From the moment of data collection, data about people, whether they are collected through surveillance or through more formal data collection instruments (i.e., surveys, interviews, and records), present a privacy concern to the data subject. What is considered to be private may vary based on culture or individual perception, but all people have information about themselves that they wish to keep, if not secret or hidden, at least under their control. Depending on who is controlling or processing the data, stigmatization or discrimination can be the result (Black, 2001; Citron & Solove, 2022; Hayes, 2017; Schneier, 2015; Seltzer & Anderson, 2001). Recent decades have seen a change in the type of data which can be collected and in its volume and speed of processing, which increases the risk of data misuse and privacy violations (C. J. Bennett, 1992; Borgman, 2018; Doyle & Veranas, 2014). With the growing requirements for government funded researchers to make their data available (e.g., Holdren, 2013) or as a condition of journal publication (e.g., American Economic Association, 2020; International Journal of Digital Curation, n.d.), these data are often made available through data archives and repositories. Data archives are tasked, therefore, with implementing the policies and legal requirements for protecting the data, even as they encourage their further use.
Threats to privacy come from many directions. As Cockerft and Rekker (2016) describe, the increasing digitization of commerce, and the accompanying abilities to store and mine commercial data threaten consumer privacy, as evidenced by the numerous consumer data hacking events in recent years (e.g., the 2017 hacking of Equifax (Benner, 2020, p. A1); the 2019 Capital One hacking (Brewster, 2019); and the 2014-2018 hacking of Marriott International (Perlroth et al., 2018, p. A1)). With data about people collected in many different ways, by many different entities, Hummel et al. (2018) argue that individuals cannot be data sovereigns if they cannot enforce their own control over data about them or control the flow of that data; data privacy, therefore, relies on the ability to control or influence by whom, why, how, and in what context data are used. This fear among data subjects moves beyond just the data collected about them by social media, smart technologies, and commercial interactions; data they may have provided in the past can now be linked to other data provided elsewhere, in ways they did not anticipate, and consent to this new processing is difficult to obtain.

Even when these data are anonymized or pseudonymized, there are well known limitations to how effective these methods are at protecting privacy. Casabona (2004) claims that it is “practically impossible” to have irreversible anonymization of data without losing the utility of the data, arguing instead that a definition of “anonymization as ‘reasonably’ irreversible is the one that must prevail” (p. 42), for some nebulous threshold of “reasonably irreversible.” Other scholars are also concerned about these limits of true anonymization on data utility (Williams & Pigeot, 2016) and the ability to reverse engineer de-identified data (Wachter, 2019). Techniques for pseudonymization likewise may still result in identifiable data, or if not identifiable, at least “reachable” (Barocas & Nissenbaum, 2014). One way in which this might occur is through the ability to link or merge data sets and, though the data themselves are
anonymized, create identifiable data (e.g., Mayer-Schönberger & Cukier, 2014; Ohm, 2009; Sweeney, 2000). Data archives such as GESIS, one of the research sites for this dissertation, are currently working on developing processes and policies for linking social media data, geospatial data, and survey data (e.g., Beuthner et al., 2021; Breuer et al., 2020, 2021; Stier et al., 2020).

Privacy concerns also exist as a result of inconsistent, inharmonious implementations of data protection legislation and practices. The Directive, for example, was intended to promote standards of data protection that would both enable the sharing and transfer of data and preserve and protect the data subject privacy. However, the lack of consistency in how the Directive was imposed created an environment where different countries had different levels of protection, raising the risk to data transferred to more lenient legal and policy regimes (Dove, 2018; Junge, 2018; Poullet, 2006; Stevens, 2015; Tikkinen-Piri et al., 2018). Research institutions and researchers within the different Member States faced challenges negotiating the different requirements, whether they wished to do multinational, collaborative studies collecting new data or to reuse data from elsewhere in comparative research. Europe as a whole, with its 20th century history of discrimination, persecution, and violence against peoples, aided by the ability to identify and locate them through data, is particularly concerned about protecting privacy by controlling and protecting that data. The GDPR, with its broad geographic applicability and scope of data types, is the largest attempt at preserving and regulating that privacy. Its effect on data use has been heavily debated in the medical field, where it is seen as potentially limiting the utility of data because of the increased requirements for consent, but how GDPR will affect data in other fields, including the social sciences, has not been heavily discussed.
2.5 Conclusion

The General Data Protection Regulation has two potentially conflicting goals: to protect natural persons through regulating the processing of data about them and to enable the processing and transfer of those data between and beyond European Union Member States. Understanding the outcome and impact of the GDPR requires more than just interpreting the requirements that it set into place in the European Union; it must be understood in the context of a century of data misuse, privacy violation, and rapid technological advancement that have now, in the 21st Century, enabled re-identification, privacy intrusion, and reputation damage in ways unimaginable in the late 1800’s. In this chapter, I have presented the historical background for the GDPR, in order to identify just why such a regulation has been deemed necessary. I have also introduced the challenges provided by data and data behaviors which the GDPR is intended to address, as well as the challenges that remain.

I have also described the different ways in which the concept of “privacy” has been conceived, in both conflicting and interweaving manners, that lead to the theoretical framework I will use to understand how informational privacy is experienced and maintained in the context of social science data archives. This model (Fig. 2.1) presents informational privacy as a broader conceptualization of privacy not grounded in direct individual control over their information (Clarke, 2016), but as the combination of three dimensions of privacy. The first of these builds on Nissenbaum’s Theory of Contextual Integrity, where the right to privacy is the “right to context-appropriate [information] flows” (Nissenbaum, 2010, p. 187). Instead of imposing a strict definition of privacy based on a specific harm that results, this theory allows me to understand how GDPR disrupts and permits the flow of personal information from data producers to data archives and data repositories to data reusers, by examining how the “context-
appropriate flows” of information have been developed by and responded to the GDPR’s imposition of norms of privacy and data protection that previously were not standardized.

The second dimension focuses on the privacy harms inherent in data sharing. Drawing on Solove’s (2006, 2008) taxonomy of harmful activities and Citron and Solove’s (2022) typology of privacy harms will allow me to understand how data archives are addressing and promoting privacy protection from a risk minimization perspective. I will be able to see this through the policies and practices the archives put in place for the archive as well as for the data and data reusers. Given the type of data the three archives contain—social science data, predominantly survey data, across multiple social science disciplines—many of the data at the archives contain the special categories of data. However, even data without those especially risky data elements can still be processed or disclosed in ways that cause harm, whether to an individual or to a community. Understanding how the archives identify potential risks and mitigate against them will ground how I evaluate the effectiveness of the archive’s response to the GDPR in protecting privacy while encouraging data sharing and reuse.

The final dimension addresses privacy protection as a response to technological risks to privacy and allows me to understand how archives are adapting to the advances in computing, storage, and communication capabilities. These advances are not just the tools available to data reusers for data analysis. They include the growing global nature of research, with multi-organizational and multi-national research collaborations, where data are shared across national borders. Safeguarding measures are required by the GDPR when personal and special data are processed; therefore, this dimension will structure my evaluation of the technological methods of privacy protection that the archives use and how staff are trained to recognize the risks in the data and processing at their organizations.
This dissertation will also address a gap in the literature surrounding the discussions about the effect of the GDPR on research in the European Union. What literature there is discusses the legal or policy ramifications of the GDPR through panel discussions and legal reviews (e.g., Albrecht, 2016; Aldhouse, 2018; C. J. Bennett, 2018; Bhaimia, 2018; Blume, 2015; Chassang, 2017; Custers et al., 2018; De Hert & Papakonstantinou, 2016; Hayes, 2017; Kuner et al., 2016, 2017; Lievens & Verdoost, 2018; Linden et al., 2019; Lindqvist, 2018; Marelli & Testa, 2018; Mondschein & Monda, 2019; Mourby et al., 2018; Poullet, 2018; Rüeger et al., 2018; Tikkinen-Piri et al., 2018; Wachter, 2019), or presented the results of data management or data protection infrastructure development (e.g., Dallmeier-Tiessen et al., 2014; Jacobs & Popma, 2019), and policy workshops or roundtables (e.g., Clark et al., 2018; Demotes-Mainard et al., 2019; Fukushima et al., 2019; Hajduk et al., 2019; Manhas et al., 2015). The vast majority of non-law review literature published after the GDPR was proposed have been commentaries either describing potential outcomes or critiquing the Regulation (e.g., Cornock, 2018; Cvik et al., 2018; Dove, 2018; Lachauffe, 2018; Mayer-Schönberger & Padova, 2016; Stevens, 2015; van Deursen & Kummeling, 2019; van Veen, 2018; Williams & Pigeot, 2016). A few were comparative studies that investigated differing legal regimes as they related to data sharing: between China and the European Union (van Deursen & Kummeling, 2019); between Canada and the European Union (Zalnieriute, 2018); and between European Union Member States themselves—Germany, pre-Brexit United Kingdom, Ireland, Sweden, and Italy (Custers et al., 2018). There have been a small number of interview and survey studies about data subject perspectives on privacy (e.g., Haeusermann et al., 2018; Junge, 2018). None of these studies, commentaries, or legal reviews took on the operation of data archives as their focus in response to the Regulation.
The literature that does discuss research data with personal data in the post-GDPR environment has been from the medical and biomedical fields, with a particular focus on genetics research (e.g., Boeckhout et al., 2018; Chassang, 2017; Demotes-Mainard et al., 2019; Fukushima et al., 2019; Haeusermann et al., 2018; Jacobs & Popma, 2019; Marelli & Testa, 2018; Nurmi et al., 2018; Perez Vallejos et al., 2019; Phillips & Knoppers, 2019; Stockdale et al., 2019; van der Ree et al., 2019; van Veen, 2018; Yuan & Li, 2019). There have been other studies since 2016 which have looked at the anticipated impact of GDPR on data use or reuse that were focused on commercial, humanitarian organizational, or administrative data (Hayes, 2017; Mourby et al., 2018; Rüeger et al., 2018; Tikkinen-Piri et al., 2018). However, there are no published studies of how social science data archives in the European Union changed, either in terms of adaptation of repository practices or in the measurable operational outcomes.

Therefore, there are two gaps in the literature that this dissertation contributes to:

- Informational privacy in data archives
- Operational impacts of new data protection legislation

### 2.5.1 Informational privacy in data archives

Information privacy literature prioritizes the concept of individual control over their personal information. It discusses privacy as resulting from an individual’s ability to restrict access to that information. In the context of research data archives, however, that individual’s ability to control that access becomes more limited (even after giving their initial consent to participate and for their data to be archived), as once they give their information to a researcher, that researcher then has some measure of control over the information (granted, ideally, through the individual’s informed consent). Once the resulting data set is then given to a data archive, that control may also be granted to the data archive, which, along with the researcher, is now
responsible for preserving that individual’s informational privacy. Therefore, informational privacy as experienced only from the perspective of the individual is not sufficient to understand how it is maintained by data archives whose missions is to make data available for reuse.

### 2.5.2 Operational impacts of new data protection legislation

There is a growing body of literature discussing the outcomes of GDPR adaptations on organizations which work with personal data. However, this literature does not address social science data archives. There have been commentaries and discussions about its effect on research, including in the humanities and social sciences, but literature focusing on organizational change because of the GDPR instead focus on medical research and health care, business IT compliance, and social media companies. While many of the findings and reported changes may be applicable to aspects of research data archive operations, the work of data archives is distinct enough from medical care and corporate personal data usage that the lack of published research in this area is a limitation for the data archiving community.

### 2.5.3 Addressing the gaps in the literature

In this dissertation, I address these gaps by investigating:

- How data archives identified what aspects of their operations were not GDPR-compliant,
- What changes (including new developments) were made to address the new data protection legal regimes,
- What motivated specific decisions about addressing non-compliance,
- What challenges data archives faced and how they affected operational outcomes (including compliance achievement), and
- The impacts and outcomes of the changes on use of data archives.
This dissertation focuses on social science data archives in four different European Union Member States. Despite supporting similar research communities, I intentionally sought to compare the experiences of data archives which differed on a number of criteria, including: size, type of data supported, ability to provide secure access to personal data, relative independence from parent or supervisory organizations in operational and strategic planning, differences in national GDPR implementations, and other criteria (discussed further in Chapter 3).
Chapter 3 Research Design and Methodology

In this chapter, I present my research design, the rationale behind the case study method, the four cases and the rationale behind their selection, the data sources I collected, and the data analysis methods applied to those data. In Chapter 4, I provide a more in-depth analysis of the cases.

3.1 Rationale for a qualitative, multi-case study

I conducted a qualitative, multiple case comparative case study to investigate the effects of privacy regulations on data sharing. In particular, I focused on the General Data Protection Regulation, because of its direct applicability across multiple nations and its potential to influence privacy policies and data sharing and reuse beyond the European Union. My guiding research questions were:

How do the ways in which General Data Protection Regulation has been enacted increase our understanding about approaches to privacy in data archives?

How has the full adoption of the General Data Protection Regulation affected data sharing and subsequent data reuse at four European Union data archives?

In so doing, this dissertation will also address the following subsidiary research questions:

1. What factors determined the four social science data archives’ responses to the General Data Protection Regulation?

2. What do the different archive implementations of the General Data Protection Regulation indicate about what aspects of privacy are prioritized?

3. What are the factors motivating the archives’ data protection policy and practice adaptations to the General Data Protection Regulation?

4. What are the mechanisms used at the data archives to address their stated goals of General Data Protection Regulation implementation?
5. In what ways have the different data archives’ policies and practices converged and diverged as a result of the General Data Protection Regulation?

6. How have the data archives conceptualized “contextually appropriate” information flows in response to the General Data Protection Regulation requirements?

This study aimed to address the lack of research available on the direct effects of GDPR, especially as it related to a topic of European Commission importance: the sharing of scientific research and innovation to the world (Moedas, 2015). I focused on one aspect of the European research community that had been left on the sidelines of discussions of the potential and actual impact: social science data archives.

The focus on organizational practices and policies in response to privacy regulation and the extent to which those practices and policies changed the use of the data archive as both a site for data deposit and protection and a source of data for reuse called for a qualitative approach to my data collection and analysis. Qualitative research is suited to explorations of problems or specific issues in order to better understand them, and allows for identification of patterns that emerge from the data during the research process (J. E. Creswell & Poth, 2018; Glaser & Strauss, 1967). Qualitative research generally relies on multiple methods of data (such as interviews and document analysis), seeks to understand research participant perspectives, and is situated within specific contexts that shape the experiences under investigation (J. E. Creswell & Poth, 2018; Maxwell et al., 2020).

This qualitative study was structured as an explanatory comparative case study (Maxwell et al., 2020; Yin, 2018). The role of explanatory case studies is to investigate and explain “complex phenomena that may not lend themselves easily to quantitative research methodologies” (Harder, 2010), and is also appropriate for the novel nature of that phenomena. The multiple case structure enables the investigation of a single phenomenon from multiple
perspectives (Maxwell et al., 2020; Stake, 1995), which in this dissertation allowed for comparison across multiple data archives in different European Union Member States, including large, long-standing archives as well as smaller, newer archives, and the explanatory structure was appropriate for the novel nature of the phenomenon under study. Each case in this study focused on an individual data archive. Within each case, data collection encompassed interviews with the staff of the archive and affiliated organizations, a review and comparison of documentation (archive policies, contracts, information sheets, etc.) and logs and records of data deposit and access where possible (See Section 3.3.2). The full list of documentation reviewed per archive can be found in Appendix E (Tables E.1-E.4). The multiple sources of data for each archive provided rich sources of evidence and allowed for triangulation of findings (Benbasat et al., 1987; Eisenhardt, 2002; Stake, 1995; Yin, 2018).

As discussed in the previous chapter, one of the motivations for this study was to address the lack of research investigating the impact of the GDPR on social science data archives in the European Union. What literature exists consists of legal reviews (e.g., Bell et al., 2019; De Hert & Papakonstantinou, 2016; Etteldorf, 2019; Guzewicz, 2020), reports on infrastructure development, commentaries about potential outcomes (especially for medical and health data) (e.g., Boeckhout et al., 2018; Chassang, 2017; Demotes-Mainard et al., 2019; Dove, 2018; Marelli & Testa, 2018; Molnár-Gábor et al., 2022), and GDPR critiques (e.g., Demotes-Mainard et al., 2019; Mourby et al., 2018; van der Ree et al., 2019; van Deursen & Kummeling, 2019). There have been pre-GDPR studies of data archives that included interviews, surveys, focus groups, document analysis, and even case studies. These studies have looked at data sharing and data reuse behaviors of researchers (e.g., Alter & Vardigan, 2015; Broom et al., 2009; Lagoze et al., 2013), the role of data archives as knowledge infrastructures (e.g., Borgman et al., 2019;
Hazeleger et al., 2018), and the ethical concerns over data sharing (e.g., Karcher et al., 2016).

However, with regard to the impact of privacy and data protection legislation on data archives and their operations, the literature predominantly comprises commentaries or reports on data archive innovations.

Therefore, as the impact of GDPR on the data archives is a unique phenomenon that has not yet been studied, case studies are appropriate for these types of phenomenon (Benbasat et al., 1987). The phenomenon itself, GDPR implementation, is the focus of the research, not an intervention or manipulation of research subject behavior. The case study seeks to answer the “how” and “why” questions about a specific, contemporary phenomenon. As Yin (2018) describes, a case study is appropriate when “you want to understand a real-world case and assume that such an understanding is likely to involve important contextual conditions pertinent to [the] case” (p. 15).

3.2 Case Selection

The GDPR’s applicability to personal data necessarily limited the scope of this dissertation to archives which process, store, and make accessible research data about people. These personal data are used in numerous academic fields; the two most prominently discussed in the literature have been in the social sciences and medicine. These data encompass a broad range of types of data which the GDPR has included in its definitions of personal data and special categories of personal data (for full definitions see Section 2.1). Data categorized as “personal” and as “special” are important for research investigating social issues and phenomena (e.g., the impact of social welfare programs on academic achievement or cyberbullying and sexuality); they are also important for an organization’s operations (e.g., to maintain a record of users to facilitate access to services) or a government’s functions (e.g., to budget for future
education needs based on localized economic or health factors, or to respond to health crises). I
focused in this dissertation on social science data archives because, while similar concerns over
data privacy and data subject protection exist in the medical field, the potential and experienced
impact of GDPR on medical research and medical care has been heavily discussed in academic
and legal literature. This is not yet the case for social science data archives.

The four social science data archives were selected based on several criteria, following a
purposive sampling approach using replication logic. Purposive sampling “is useful when the
research is oriented towards a specific case” (Barglowski, 2018, p. 163), and enables the
selection of cases based to achieve maximum variation (Flyvbjerg, 2006). Social science data
archives in the European Union generally work with data applicable to research in various social
science disciplines, though not every archive necessarily contains data applicable to all
disciplines. Many data sets are also used across multiple disciplines; for example, national
census data are used by economists, sociologists, public health researchers, political scientists,
and others. Because of this commonality, I selected my four cases based on more criteria than
just which disciplines they supported, with the intention to maximize the differences I
hypothesized would be relevant to understanding the choices and decisions made at each data
archive. The cases for this research were selected based on: relative age of the archive; number
of data sets made available; number of employees; whether an independent institute or affiliated
with a parent organization; type of data archived; services offered besides the data archive with
personal data considerations; ability to ingest and disseminate data through secure data access
methods; data access restrictions; external research collaborations; whether the archive is
positioned as a processor or a controller and for which data; historical and cultural influences on
privacy; and unique aspects of or derogations in the GDPR implementation (see Table 3.1).
<table>
<thead>
<tr>
<th>Criteria</th>
<th>ČSDA</th>
<th>DANS</th>
<th>FSD</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Czech Republic</td>
<td>The Netherlands</td>
<td>Finland</td>
<td>Germany</td>
</tr>
<tr>
<td># Employees</td>
<td>12</td>
<td>50+</td>
<td>27+</td>
<td>300+</td>
</tr>
<tr>
<td># Data Sets</td>
<td>960+</td>
<td>195,000+</td>
<td>1600+</td>
<td>6,700+</td>
</tr>
<tr>
<td>Format of Data</td>
<td>Quantitative</td>
<td>Qualitative</td>
<td>Quantitative</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Services Besides Data Archive</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Secure Data Access</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Archives Personal Data</td>
<td>No*</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Access Restrictions</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Affiliation</td>
<td>Department of Institute of Sociology</td>
<td>KNAW Research Support Service Provider</td>
<td>University</td>
<td>Independent Research Institute</td>
</tr>
<tr>
<td>External Research Collaborations</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controller</td>
<td>Administrative/operational data External research</td>
<td>Administrative/operational data External research</td>
<td>Administrative/operational data External research</td>
<td>Archived data Administrative/operational data External research</td>
</tr>
<tr>
<td>Processor</td>
<td>External research</td>
<td>Archived data DataverseNL</td>
<td>Archived data</td>
<td>External research</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>National GDPR Derogations and Implementations</td>
<td>Exemptions for providing information for “academic” purposes “in public interest”</td>
<td>Exemptions for consent when effort is “disproportionate” Data archiving derogation</td>
<td>Freedom of expression derogation</td>
<td>Factual anonymization State-level data protection laws</td>
</tr>
<tr>
<td>Historical/Cultural Influences on National Privacy Perception</td>
<td>Communist regime (1948-1990) Constitutional right</td>
<td>Nazi abuse of personal data registers Open society reputation Constitutional right</td>
<td>Employer abuse of union membership Constitutional right</td>
<td>Nazi regime East German communist regime (1949-1990) Constitutional right</td>
</tr>
</tbody>
</table>

These criteria were chosen because they reflect key concepts on which to compare and contrast the cases. The differences between the four archives allowed me to also test the theoretical model of privacy in the context of each data archive.

I based my decision on these criteria for several reasons. Data archives have had to respond to GDPR’s requirements and expectations for data protection and privacy just as companies do, but have done so under different funding and personnel limitations. For this reason, I did consider that the staffing and volume of data in the data archives could be a factor; if the necessary changes were extensive but the staff few, the potential workload burdens could be higher, while a larger staff might be better able to address the needs. However, because a larger staff also often supports additional services, this could also be a factor, as additional services working with personal data would also need to be addressed along with research data and the archive’s administrative and operational data. Whether the data archives were reliant on external parties, such as a parent organization, advisory board, or national data protection authority, to set or approve their response plans, was also considered as a possible contributor to how they responded.
to the GDPR. Similarly, if the data archive participates in or leads multi-national or multi-institutional research projects, whether the resulting data are then archived with them, they must consider whether and how the GDPR would affect those data.

The research data in the archives were also a consideration. Beyond just the disciplinary areas, the number of data sets potentially affected by changes could play a role in how readily an archive can become GDPR compliant, and the anonymization challenges for qualitative data (such as interview transcripts or video recordings) are different than those for quantitative survey data and often require different processes or extra considerations for data ingest into the archive. The ability to provide access to personal data under restricted access conditions means at those archives affects more than just the technical data protection measures or archiving platforms. GDPR-related changes or new requirements could also affect the agreements between the archives and their data providers or data users regarding deposit or use of restricted data.

Under the GDPR, there are two roles available to organizations and the personal data they work with: controller and processor. The controller has responsibility for not only deciding what processing will be done to personal data and why, but also for demonstrating compliance with the legal requirements. The processor processes\textsuperscript{94} personal data only under the direction of the controller, and while the processor must comply with the GDPR, it is not responsible for demonstrating that compliance. The research data in the data archives are not the only personal data whose processing must be GDPR compliant; the data resulting from any associated research projects and the administrative and operational data (e.g., employee personal data, user data from account registrations, etc.) used for the organizations’ operations may also fall under the GDPR’s

\textsuperscript{94} Art. 4(2) GDPR defines processing as operations performed on personal data, including “collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.” At data archives, for example, whether for archived research data or other personal data, archive staff may perform any of these types of processing, in addition to data curation tasks applied to research data during and after a deposit review.
purview. Processes and policies for these additional personal data sources would also need to change, and which, if any, processor or controller role applied to those data could affect what changes were necessary. I determined which role—processor or controller—applied to which data—e.g., the archived research data, archive-involved research project data, administrative and operational data used for archive operations—during my initial reviews of the data archives policies in 2019 and 2020. These roles are stated in several publicly accessible parts of the data archives’ web pages, including privacy policies, data deposit guidance, deposit, processing, or usage contracts, and/or other legal information. I verified the chosen roles during the interviews.

Outside of the data archives, there were two other criteria I considered to be possible factors in how the data archives approached the GDPR. The first relates to the national GDPR implementation laws and how they interpreted or derogated from the GDPR (Section 2.3.5.3). The motivation for the comparative case study nature of this dissertation was to compare across different data protection regimes; therefore, the differences between the national legislation and the GDPR, and each other, were considered in my selection process. Finally, while 20th Century personal data abuses in Europe are significant motivators for the development of data protection legislation over the last 50 years (Section 2.3.1), these are not the only influences on how, at a broad national and cultural level, the issues of privacy and data protection are conceived of across the European Union Member States. This dissertation aims to understand the factors that influenced the GDPR response at each archive, and differences in how privacy in particular is conceived of and prioritized could be one of those factors. Therefore, during the selection process, I also sought differences in this area.

These criteria were used to identify possible cases among the European Union social science data archives. In the end, I selected four sites: Germany’s GESIS Leibniz Institute for
the Social Sciences (GESIS), the Netherlands’ Data Archiving and Networked Services (DANS), the Finnish Social Science Data Archive (FSD), and the Czech Social Science Data Archive (ČSDA) as the cases for this dissertation. This selection was based on maximizing the differences between the cases based on those criteria. Data were collected from fall 2020 through the end of 2021; to support this research, I was awarded the Fulbright Schuman research grant to support data collection and analysis at the four data archives from September 2021-June 2022. Of note, FSD was the site of my pilot study in summer 2020. It was added as a fourth full case in this dissertation in fall 2020, wherein a fuller evaluation of the site was conducted.

3.2.1 Czech Social Science Data Archive (ČSDA)

The Czech Social Science Data Archive (Český sociálněvědní datový archiv) (ČSDA) is the data archive for the Czech Academy of Sciences (Akademii věd České republiky) (AV ČR) in Prague. The AV ČR is the leading Czech research body, established under national mandate by Act No. 283/1992 Coll.\(^95\) The Academy is funded under the national budget, and is considered a “central administrative body (ústředního orgánu České republiky)” of the Czech Republic (§ 3(2)); a Czech central administrative body “control[s] state administrative performance in [their] relevant [competency] area” (Bureš et al., 2004, p. 18). First established in 1998 as the Sociological Data Archive, ČSDA is a small department (with 12 staff members as of early 2022) within the AV ČR’s Institute of Sociology (Sociologický ústav AV ČR) (SOÚ). It is focused on data archiving and promoting secondary use of social science data.

ČSDA primarily archives quantitative social science survey data through the Nesstar-based catalogue, though qualitative data are made available through the MEDARD catalogue (viewable through the same Nesstar platform) under stricter rules and conditions for deposit and reuse (see also Antes et al., 2018). As of November 2021, ČSDA is transitioning to a Dataverse-based archive, though the Nesstar platform will remain in operation for the foreseeable future. There are currently 960+ data sets in the Nesstar catalogue.

ČSDA does not have the infrastructure to support remote access to data sets with personal data. The only data publicly available are considered “fully anonymized.” There is a small collection of data sets with personal data that ČSDA physically stores; these data are from SOÚ researchers and are only available for SOÚ researchers to reuse in a secure room within the SOÚ’s Sociological Library in Prague. The data at ČSDA cover such topics as elections and politics, youth issues, economics, housing, social change, and public opinions; the majority of the data come from SOÚ researchers, though data from non-SOÚ sources is increasing as outreach to external researchers and data providers has increased. With a few exceptions, the data are available for free. As an SOÚ department, ČSDA staff support the Czech component of the European Social Survey, and as a data archive, ČSDA staff participate in external research and development collaborations with CESSDA.

Because the research data available online through Nesstar are considered anonymous, ČSDA has adopted no processor or controller role for those data. In terms of the personal data from SOÚ researchers, the SOÚ is the controller, and any processing that ČSDA staff carry out for these data is done in their SOÚ support role. Similarly, ČSDA is the controller for the

---

Nessstar is a “software tool for web data publishing and online analysis. Through Nesstar, users can search data, analyze data online, and download them to their computer. In addition to the data themselves, it also contains metadata... as well as associated documentation” (Czech Social Science Data Archive, 2022). The ČSDA interviewees, and both the Czech and English versions of the ČSDA website refer to the data archiving platform as a “catalogue,” which I have retained in this dissertation.
administrative and operational data it relies on for its operations (e.g., account registration information), because the SOÚ has adopted the controller role for all administrative and operational data used across the Institute. Whether ČSDA is the processor or controller for the external collaborations depends on the specific project.

Within the Czech Republic, privacy is a constitutional right, and has been prioritized since the end of the communist regime in 1990. The reuse of personal data has been highly regulated since then as well; this is most apparent in that the GDPR has enabled the reuse of personal data for research purposes in the Czech Republic; the previous Data Protection Act (2000) made no such exemption due to parliamentary concerns stemming from experiences during the communist regime. The Czech implementation of the GDPR also includes exemptions from GDPR requirements about providing information to data subjects when personal data are used for academic purposes in the public interest (§ 18(2)(d)), though the practical implications of how this “academic exemption” apply to data archives such as ČSDA are unclear. § 16 covers personal data processing for “scientific or historical research or for statistical purposes,” but there is no related exemption as there is for § 17, which regulates processing for “journalistic purposes or purposes of academic, artistic or literary expression.”

3.2.2 Finnish Social Science Data Archive (FSD)

The Finnish Social Science Data Archive (FSD) is an archive operating within Tampere University, in Tampere, Finland. It is funded by the Finnish Ministry of Education and Culture, with additional, project-specific funding from the Finnish government and the European Union. Though it is organizationally part of Tampere University, it is considered a national research
center. Under the Universities Act (558/2009), Tampere University has administrative control and responsibility for the archive, despite the Finnish government support both for and by the archive. It has been in operation since 1999, and supports researchers, teachers, and students across Finland, as well as multinational researchers whose data or research are relevant to the Finnish context. There are currently 29 employees of the archive, as well as temporary and student employees as needed.

The archive holds 1,686 data sets as of December 31, 2021, of which 2,282 are qualitative data sets and 1,403 are quantitative. Data in Finnish, Swedish, and English can be deposited at the archive, and Finnish data can be translated to English upon request. FSD data are available to users at no cost, though there are various levels of accessibility. While FSD does focus primarily on the social sciences, the data reflect a broad spectrum of subject areas: e.g., agriculture, business, cultural issues, economics, health behavior, international politics, medical and health sciences, psychology, physical fitness, religion, etc. The research data are accessible through the Aila portal. FSD does not have, and has never had, the infrastructure to make identifiable data available to data users through various restricted data access methods.

While FSD primarily archives anonymized data sets, it does archive data sets with personal data under two derogations from the GDPR: a “freedom of expression” exemption (§ 27) and a “public interest” exemption (§ 4) in the 2018 Data Protection Act. Since this dissertation began, FSD has also started archiving pseudonymized data. FSD is the Finnish node of CESSDA; however, while FSD staff collaborate on CESSDA infrastructure development projects, they do not conduct research. In terms of its GDPR role, FSD has adopted the

---

98 There is one exception to this. FSD is funded to conduct the Finnish component of the ISSP survey; for this external project, FSD is the controller of these data. Otherwise, FSD is the processor for any personal data resulting from these projects.
processor role for the research data in Aila, and the controller for FSD’s administrative and operational data.

As discussed above, there are two particular derogations from the GDPR in the national implementation law that are relevant to FSD. These are the exemptions for personal data processing in the public interest, and under the constitutional right of “freedom of expression.” This second exemption was used before the GDPR to support personal data archiving related to public figures in Finland. Historically in Finland, privacy and personal data protection have been highly valued. In addition to more than thirty years of data protection laws, early 20th Century abuses of records of union membership to deny employment have heavily restricted any processing of employee data across Finland, even when the data subject has or would consent to that processing.

3.2.3 Data Archiving and Networked Services

The Netherlands’ Data Archiving and Networked Services is an institute within the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW). KNAW is the umbrella organization overseeing 11 national research institutes as well as DANS, the research infrastructure support service. In addition to supporting and promoting the various research institutes, the KNAW is also the Dutch government’s scientific advisory board. It receives funding from the Ministry of Education, Culture and Science, but is not itself a part of the Dutch government in any other capacity than as an advisor or collaborator on initiatives. Established in 2005 in The Hague, DANS serves as the digital research infrastructure for Dutch researchers and institutions. In addition to supporting DataverseNL instances for in-progress research, DANS provides the EASY online archival system for dissemination of archived data. As of November 2021, DANS is transitioning to a series of thematic, Dataverse-based “Data Stations” providing
access to both research data and research support for archaeology, the social sciences and humanities, life, health, and medical sciences, and the physical and technical sciences.

This archive contains the social science data collections of the previous Steinmetz Archive (an archive for Dutch quantitative social science data), as well as further data from research in the humanities, medical and health sciences, social science, behavioral science, oral history, spatial science, and archaeology (Koninklijke Nederlandse Akademie van Wetenschappen, n.d.). It ingests and provides access to both qualitative and quantitative data. DANS does provide access to personal and identifiable data, as well as data with other restrictions, through a restricted data access process. All data are available free of charge, no matter the access level. As of the end of 2021, DANS provides access to nearly 200,000 data sets through the EASY portal at four levels of access restriction (though the remaining two legacy categories will be phased out), including nearly 75,000 which are archived at DANS but from other institutions. Data producers and users are supported by 50+ employees in administrative, data acquisition, data dissemination, legal, IT, and other roles.

In addition to EASY, DANS hosts DataverseNL, a consortium of sixteen Dutch institutions who contract with DANS to provide the technical infrastructure for institutional Dataverses. DANS staff also collaborates on external research and development projects through its membership in CESSDA and in support of European open science initiatives. DANS has adopted the processor role for the archived research data in EASY and for DataverseNL, while it is the controller for its administrative and operational data. DANS is also the controller for data from many of the external projects it is involved in, depending on the contract for the project.
The Netherlands is known for its relatively open society. While Dutch people are growing more concerned about data and informational privacy, the prevalence of “open-curtained windows” facing the street and providing a clear view into the home have also given the Dutch a reputation for not being overly concerned with privacy, despite its status as a constitutional right. They also have an 800 year-long history of creating detailed population registers. After the end of World War II when the role of these population registers and the religious and ethnic identifiers documented on them became known, these data points (which are included in the GDPR’s list of “special categories of personal data”) were removed from these records.

3.2.4 GESIS — Leibniz Institute for the Social Sciences

The GESIS — Leibniz Institute for the Social Sciences (GESIS – Leibniz-Institut für Sozialwissenschaften) is the largest social science data archive in Europe. It was established in 1986 as the German Social Science Infrastructure Services (Gesellschaft Sozialwissenschaftlicher Infrastrukturreinrichtungen), comprising three institutes in Bonn (InformationsZentrum Socialwissenschaften), Cologne (Zentralarchiv für Empirische Sozialforschung), and Mannheim (Zentrum für Umfragen, Methoden und Analysen). In 2007, the three institutes merged into the single GESIS institute, though the three locations were maintained for the time being, and in later 2008 was renamed the GESIS — Leibniz Institute for the Social Sciences. In late 2011, the Bonn and Cologne locations merged into the current Cologne location, while the headquarters remain in Mannheim. There are more than 300 employees supporting “advice, expertise and services at all stages of scientists’ research projects” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.b). It is an independent research
institute receiving funding support from the federal government, the states of North Rhine-Westphalia and Baden-Württemberg, and research grants.

Within GESIS, the Data Services for the Social Sciences (DSS) department serves both data producers and data reusers, making over 6,700 studies available through two archives: the Data Archive and the self-archiving SowiDataNet|datorium. GESIS provides access to German government- and European Commission-collected data as well as data from university or research institute-affiliated researchers. The topics covered by the data include employment, religion, social behavior, mass politics, and more within the social sciences. In addition to its research services, GESIS is a member of the Consortium of European Social Science Data Archives (CESSDA) and the International Data Access Network (IDAN), and collaborates on research and infrastructure processes with these and other partners. GESIS staff also conduct research outside of these collaborations.

As the oldest social science data archive in Europe, GESIS has played an important role in establishing standards for data management, curation, and dissemination; it also develops and disseminates training on data management and preservation for all CESSDA members. It only collects quantitative data (any qualitative data collected in the non-self-deposit archive are those supplementing the primary quantitative data being deposited). GESIS does accept data sets with personal data and provides access to identifiable data under a tiered restricted data access system, including the on-site Secure Data Center in Cologne. Data with access restrictions generally require the data producer to authorize further data use, and may also require a fee. GESIS has adopted the controller role for the research data in both archiving systems. It is also the controller for the administrative and operational data. Whether GESIS is the processor,
controller, or joint controller for data from its external collaborations depends on the contract for the project.

The data protection regime within Germany is the oldest in the world, and a right to privacy is guaranteed in the first two articles of the German constitution, the Basic Law. Germans have a reputation for being deeply suspicious of government use of personal data; this has its foundations in the experiences under Nazism and East German experiences under the communist regime. There are multiple levels of data protection within Germany; in addition to the GDPR and the national implementation law, the 2018 Bundesdatenschutzgesetz, each German state has its own data protection law for which organizations operating throughout the country must negotiate compliance. Germany has also defined a particular interpretation of what “anonymization” means which differs from the description provided in the GDPR. This definition of “factual anonymization” has been retained after a period of confusion about whether “factually anonymized data” were actually to be considered personal data under the GDPR’s purview (Rat für Sozial- und Wirtschaftsdaten (RatSWD), 2018). I will discuss this further in Section 4.4.

3.3 Data Sources

This case study’s research questions address the policies and practices of data archives in response to privacy regulations that have, on the surface, conflicting goals: to protect data subject privacy and to enable the free flow of data within the European Union. Therefore, the primary sources of data came from the data archives themselves. Yin (2018) has identified six common data sources for case study research: documentation (e.g., internal administrative documents, policies, news, memoranda, etc.), archival records (e.g., organizational charts, organizational records, services records, etc.), interviews (e.g., talking with key respondents about facts and
opinions), direct observations (e.g., of relevant behaviors or conditions), participant-observation (e.g., actively take part in the event under analysis), and physical objects and artifacts (e.g., a piece of physical evidence). This study formally incorporated two of these types of data sources from the data archives: 1. Semi-structured interviews with data archive staff and stakeholders (including advisory boards and affiliated data protection officers); and, 2. Administrative documentation, including archive policies for data acquisition, data use, and data protection from before and after the GDPR was adopted in 2016, as well as the data archive’s usage statistics (including data acquisition, data sharing, secure data applications, and other metrics of archive use from before and after the GDPR was adopted in 2016). I describe these data sources below.

To establish the national contexts for privacy and data protection for each of my four field sites (discussed in the third section of each case report in Chapter 4), as well as to develop related questions for the interview protocols (Section 3.3.1) I also relied on data from two Eurobarometer99 surveys: the 2015 Eurobarometer 83.1 (European Commission, Brussels, 2018) and the 2019 Special Eurobarometer 487: Charter of Fundamental Rights and General Data Protection Regulation (European Commission, Brussels, 2021). Additionally, in Section 4.1, I report on Czech-specific results from Eurobarometer 91.1 (European Commission & European Parliament, Brussels, 2019), which collected EU-wide public opinions on the European Union itself three months before the 2019 European Parliament elections.

---

99 The Eurobarometer surveys, conducted since 1974, are surveys “used by the European Commission, the European Parliament and other EU institutions and agencies to monitor regularly the state of public opinion in Europe on issues related to the European Union as well as attitudes on subjects of political or social nature” (European Union, n.d.). Topics of interest include politics, health, foreign/external relations, science, culture, education, climate, etc.; “special” Eurobarometer surveys can also be conducted to investigate opinions on specific issues. The Special Eurobarometer 487 (European Commission, Brussels, 2021) is one of these “special topic” surveys, focusing on the GDPR and the EU’s Charter of Fundamental Rights, while the Eurobarometer 83.1 (European Commission, Brussels, 2018) was a broader survey of attitudes towards the European Union.
### 3.3.1 Interviews

In order to better investigate what GDPR’s requirements mean to institutions that must implement them and how those decisions were made, I conducted interviews. Rapley (2013) states that “interviews are, by their very nature, social encounters where speakers collaborate in producing retrospective (and prospective) accounts or versions of their past (or future) actions, experiences, feelings and thoughts” (emphasis in original) (p. 16). Specifically, these interviews were semi-structured interviews, a structure which asks all respondents the same questions while enabling flexibility within the interview process for probing further into interviewee responses and adapting to the unique characteristics of the different cases and the individual expertise of each informant (Dearnley, 2005; Hesse-Biber & Leavy, 2005; Rubin & Rubin, 2012). Yin (2018) identifies the benefits of interviews as a component of case study research: interviews can be targeted to address the specific topics of interest, and they can be insightful, providing an opportunity for extensive discussion and explanation, as well as for probing for further information that other methods, such as survey questionnaires and document analysis do not allow for (see also Creswell, 2014).

Interviews were requested from a total of sixty-five potential respondents (31 from GESIS, 21 from DANS, 6 from ČSDA, and 7 from FSD). Forty-one interviews were completed with forty-two respondents (one interview was conducted with two respondents). I conducted all but two of the interviews with respondents at the four data archives remotely via Zoom, Skype, or GoTo Meeting. These interviews (totaling 70 hours) were recorded and transcribed. I verified that all transcriptions were accurate. For the remaining two interviews, the respondents requested to respond to a written questionnaire due to time constraints and concerns over possible language limitations. Each respondent completed their questionnaires as well as an
additional questionnaire with follow-up questions based on their initial responses. All interviews were completed in English, with some responses given in Czech and Dutch which required translation and confirmation of the translation with the respondent. At the end of each interview, I wrote a reflection that summarized the responses, identified emerging patterns, themes, and relationships, and listed what, if any, additional resources the respondent provided (e.g., links to affiliated institutions, policy documents, etc.).

Table 3.2: Interviews by archive and respondent category

<table>
<thead>
<tr>
<th>Respondent Categories</th>
<th>GESIS</th>
<th>DANS</th>
<th>ČSDA</th>
<th>FSD</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive stakeholders</td>
<td>1</td>
<td>0</td>
<td>1$^6$</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Senior leadership</td>
<td>4$^4$</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Archive staff</td>
<td>10$^7$</td>
<td>8$^3$$^5$</td>
<td>1</td>
<td>2$^1$</td>
<td>21</td>
</tr>
<tr>
<td>Legal advisor</td>
<td>1$^2$$^7$</td>
<td>2$^3$$^5$</td>
<td>1</td>
<td>1$^1$</td>
<td>5</td>
</tr>
<tr>
<td>IT support</td>
<td>1$^4$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>17$^2$</td>
<td>15</td>
<td>4</td>
<td>5</td>
<td>41</td>
</tr>
</tbody>
</table>

1 At FSD, the data protection officer is also a member of the user support staff in the archive.
2 At GESIS, one interview was conducted with two respondents, for a total of 42 respondents in 41 interviews.
3 At DANS, the legal advisor is also an active member of the archive staff.
4 At GESIS, one senior leader provided extensive insights into IT needs and concerns.
5 At DANS, one archive staff member also provided extensive insights into the legal considerations.
6 At ČSDA, the stakeholder respondent also provided extensive insights into legal considerations.
7 At GESIS, one archive staff member had previously been the data protection officer and provided extensive insights into legal considerations.

Within the data archives, I identified five categories of respondents who could provide insights into this research. These categories were: archive stakeholders (including advisory boards), senior leadership (including directors and department leadership), archive staff (including acquisitions, data curation, administration, and user services), IT support, and legal advisors (including data protection officers). Table 3.2 above provides a breakdown of these categories and the number of interviewees from each archive. I identified potential interviews for each category first from each data archive’s public facing web sites, and then through
snowball sampling (Biernacki & Waldorf, 1981; Coleman, 1958). At the end of each interview, I asked for recommendations for other potential interviewees, who were then invited to take part as well. Four respondents were at the time of the interview or had been in the past both a legal advisor or data protection officer as well as an archive employee and provided legal and archive insights into the GDPR experience. One respondent in a leadership role was the head of the IT department, and provided IT-specific considerations as well as a higher-level, strategic perspective.

Three interview protocols were developed to address three broad categories of respondents: staff of the archive, legal advisors, and external stakeholders. The full interview protocols can be found in Appendix A (Data Archive Staff Interview Protocol), Appendix B (Legal Interview Protocol), and Appendix C (Archive Stakeholder Protocol). The questions were designed to capture respondent perspectives on the archive’s GDPR response experiences, specific changes and challenges faced, the effectiveness of the GDPR, and the respondent’s own views on privacy and privacy regulation. The legal advisor protocol also asked about legal considerations and perspectives on the GDPR and national implementations, while the stakeholder protocol also asked about the role of archive stakeholders in the archive’s response.

The interview questions were initially developed out of questions raised from the literature. For example, the literature reporting the impacts of the GDPR on other research areas, such as medical and health research, and the literature discussing fears about what impacts the GDPR might have on research or potential issues for organizational operations, guided the questions addressing “institutional policies and procedural changes.” The questions about the interviewee’s perspectives on privacy, the GDPR, and various aspects of their archive’s experienced were developed out of the literature on privacy and a review of the Eurobarometer
83.1 (2018) and Special Eurobarometer 487 (2021) national responses to questions about data privacy and the GDPR. Questions specific to the legal interview protocol (Appendix B) were developed based on review of the texts of the relevant legislation and legal discussions about the GDPR (Section 2.3.5).

During the pilot study with the Finnish Social Science Data Archive over the summer of 2020, the Data Archive Staff Interview Protocol was tested and revised based on how the pilot study interviews flowed. Follow-up questions that came out of the semi-structured nature of those interviews became permanent questions in the protocols. Minor revisions to it were made over the first 20 interviews; where possible, I did follow-up via email with the previously interviewed participants to address additional questions. The separate Legal Interview Protocol and Archive Stakeholder Interview Protocol were developed out of the pilot study. One pilot study respondent recommended that I speak with FSD’s legal advisor; in that interview, I used the Data Archive Staff Interview Protocol with additional questions about the legal context based on the legal texts and legal commentaries on the GDPR. The revision of the original protocol to a form specific to the legal context came out of that first legal interview, and was not further revised. The same pilot study respondent also indicated that advisory board members did provide feedback on the GDPR transition process at FSD; for this reason, I added archive stakeholders to my interview plan and created an additional tailored version of the interview protocol that focused on understanding the stakeholder’s role with the data archive.

3.3.2 Administrative documentation

A study of institutional responses to policy changes should include documentary evidence reflecting those changes. Yin states that for case study research, “the most important use of documentation is to corroborate and augment evidence from other sources” (2018, p. 115).
In a multiple case study framework, administrative documentation also supplement the comparative analysis between the cases. There are two distinct types of administrative documentation that I collected for this dissertation: documentary evidence and archive usage statistics.

3.3.2.1 Documentary evidence

In this study, the categories of administrative documentation I collected are:

- Archive information (general information about the archive, its operations, and/or services)
- Archive policy (policy documents about data deposit, data acquisitions, data use, data curation/quality control, etc.)
- Archive publications (data archive-produced news releases, press releases, blog posts, and research publications about archive operations, data protection, privacy, the GDPR, etc.)
- Archive training materials (internal and external training materials about data management, data protection, privacy, the GDPR, etc.)
- Deposit agreements (any agreements or contracts required for data deposit)
- Deposit guidance (guidance or requirements for data providers related to data deposit, including anonymization requirements, preferred file formats lists, etc.)
- External agreement (contracts or policies between the data archive and external parties)
- Governance and oversight (policies, evaluation protocols, etc. related to external oversight and supervisory bodies)
- Legislation (relevant national or sub-national legislation)
- Privacy policy (data protection and privacy statements or information sheets)
- Terms of use (data use agreements, end-user licenses, and/or terms of use)
- User guidance (guidance or requirements for data users related to accessing and using research data from the data archive)

The full breakdown of collected administrative documentation by archive can be found in Tables E.1-E.4 in Appendix E. Where possible, I collected multiple versions covering the period 2011-2020, though this was not possible for all archives. This time frame recognizes that the
European Commission formally proposed the GDPR in 2012, beginning a several year period of
debate and negotiation over the final version of the Regulation before its 2016 adoption and 2018
implementation. I collected some documentation dated back to 2007 in instances where the
previous version of a policy document pre-dated the proposal. Doing so enabled a comparison of
archive policy and practice changes over time both before and after GDPR was adopted on April
14, 2016.

I referred to available literature on organizational GDPR changes to identify possible
types of documentation at my research sites (e.g., Almeida Teixeira et al., 2019; Brodin, 2019;
Lopes & Oliveira, 2018; Sirur et al., 2018). These guided my initial review of publicly available
documentation; as I systematically identified types of documentation, I revised my list, and
returned to previously reviewed data archives to check for documentation not found on the initial
search. Many documents were available on the data archive public-facing web sites. Historical
information (pre-GDPR as well as post-GDPR but pre-data collection) were identified through
the Wayback Machine; 100 if the desired information was not accessible (e.g., a document only
available as a .pdf file which was not included in the web capture), I requested it from the data
archive or was granted access to the archive’s intranet where the documentation were available.
All four archives support English language data deposit and data reuse, and most documentation
is available in English. ČSDA was the exception to this; less documentation existed there
overall, though there are some with English versions. I evaluated what documentation was not
available in English and coordinated with the archive to confirm translations.

With the exception of internally-focused materials in the “Archive training” category, all
four data archives have documentary sources in all categories listed above. While interviewees

100 https://archive.org/web/
at all four sites discussed improvements and intensive revisions to internal training materials, these were the one category of documentation that they were less willing to provide. I only have internal and external training materials from DANS; the training materials from ČSDA and GESIS are exclusively externally focused, to national or CESSDA-wide audiences, while I received and found no training materials from FSD. ČSDA, as the smallest data archive and the one which has not claimed a processor or controller role for its research data, does not have the breadth of documentary sources seen at the other three data archives (Table E.1). This is not a limitation based on language, though the ČSDA interviewees acknowledged that creating English language translations are a forthcoming priority. As will be discussed in Section 4.1, there were fewer formal changes in response to the GDPR; therefore, the multiple versions of policies, agreements, and regulations are less common. GESIS has the largest number of documents due to the structure of its depositor and user guidance (spread out on individual web pages or embedded tabs on web pages, instead of collated into more readily accessible documents or files).

3.3.2.2 Archive usage statistics

An important aspect of understanding an archive’s response to the GDPR is whether there is any discernable effect on the operation of the archive as an archive. Would the increased focus on data protection change data acquisitions, due to stricter compliance with legal requirements, changes in policies and practices, or other effects of the implementation? Would the use of, or attempted use of, data from the archive change? Archive usage statistics provide a way to see the practical implications of the GDPR on actual archive operations, without solely relying on interviewee perceptions. I sought the following from the four archives:

- Annual data deposits and/or data set publications (by access category if available) 2011-2020
• Annual end-of-year total data sets (by access category if available) 2011-2020
• Data download and/or usage statistics (by access category if available) 2011-2020
• Requests for access to secure data 2011-2020
• Archive account registrations 2011-2020

Not all of the requested data were available at all of the archives. The data specifically used for each archive are discussed below.

Czech Social Science Data Archive

At ČSDA, the data I used are the following:

• ČSDA End-of-year Total Data Sets (2011-2020)
• ČSDA Annual Data Set Publications (2012-2020)
• ČSDA End-of-year Total Registered Users (2011-2020)

I created all three data sets by aggregating the data from the Institute of Sociology’s (the organization ČSDA is a department of) annual reports (Sociologický ústav Akademie věd České republiky, v.v.i., 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022b), available through the public facing web page. These annual reports are only available in Czech; therefore, I created a list of likely terms which might be used to describe the data I was interested in and confirmed possible variations with a native Czech speaker. I then searched through several years’ reports, and refined my list based on those results. I reconfirmed the relevant search terms with the native Czech speaker and had them also review two reports from different years, to validate that the data I was interested were the data that I found using those search terms. The structure of the Nesstar catalogue does not facilitate obtaining total counts of data, whether in Czech or in English; therefore, the annual totals and publication numbers are only viewable in these annual reports. ČSDA does not report data download or other possible data reuse statistics; therefore, I only have anecdotal reports of possible data use changes from the ČSDA interviews.
At FSD, the data I used are the following:

- FSD Data Reuse 2015-2020 (Version 2.0) (University of Tampere, 2021)
- FSD Downloads and Deposits totals (2010-2021) (Kuula-Luumi, 2022)
- FSD Archived Data Totals by Access Category (as of January 1, 2022) (Finnish Social Science Data Archive, n.d.g)
- FSD Data Deposits (2008-2021)
- FSD Data Downloads and Access Requests (2012-2021)

“FSD Data Reuse 2015-2020 (Version 2.0)” is an open access dataset available at FSD under a CC BY 4.0 license. It documents all data set downloads by month from 2015-2020, and identifies whether the data are qualitative or quantitative, the access category, and the purpose of the download if provided, among other variables. From this data set, I identified annual data download totals by qualitative/quantitative, access category, and download purpose. The “FSD Downloads and Deposits totals (2010-2021)” and “FSD Downloads by Access Category (2018-2021)” were e-mailed to me in January 2022 once the annual numbers had been reported for other internal FSD documents (Kuula-Luumi, 2022). They provide the annual (not broken down by access category) data downloads and data deposits per year from 2010-2021 and the annual data downloads by access category for 2018-2021, respectively. “FSD Archived Data Totals by Access Category (as of January 1, 2022)” represents a snapshot of the archive totals as of the end of 2021 of the total data sets in FSD and was documented from the main page of the Aila Data Catalogue on January 1, 2022.

---

101 The Aila Data Catalogue ([https://services.fsd.tuni.fi/catalogue/index?limit=50&study_language=fi&lang=en&page=0&field-publishing_date&direction=descending](https://services.fsd.tuni.fi/catalogue/index?limit=50&study_language=fi&lang=en&page=0&field-publishing_date&direction=descending)) can be filtered by access category, whether the data are qualitative or quantitative, and whether the data are available in Finnish, English, or Swedish. To determine the totals by the end of 2021, the Finnish language data catalogue was used (the data catalogue can be viewed in either English or Finnish, based on whether the data are available in English) because all data sets in FSD are listed in the Finnish catalogue, while the English catalogue is missing 33 data sets in its count. The Finnish total was confirmed to be accurate after consultation with my FSD liaison. The data catalogue was filtered based on the access categories and the qualitative/quantitative distinction on January 1, 2022, and re-verified on
I created the final two data sets, “FSD Data Deposits (2008-2021)” and “FSD Data Downloads and Access Requests (2012-2021)” by aggregating the usage numbers reported from several sources available on the FSD website: the annual news releases reporting the previous year’s FSD usage (covering 2016-2021) (Finnish Social Science Data Archive, 2017b, 2018d, 2019a, 2020c, 2021b, 2022b) and the Annual Reports (in Finnish) for 2010-2020 (Finnish Social Science Data Archive, 2011, 2012, 2013b, 2014c, 2015b, 2016, 2017a, 2018c, 2019b, 2020b, 2021a, 2022a). The news releases (available in both English and Finnish) provided the following per year: Total number of data sets published in Aila per year, data publications by qualitative and quantitative (the numbers for 2017 and 2018 here were not reported, and the numbers counted instead from the Data Catalogue webpage), annual totals (archive total, totals by qualitative, quantitative, and openly accessible (Category A)), applications for access to data (for Categories B, C, and D), total data sets delivered for reuse, total Aila registered users, total unique Aila register users who downloaded data (only for 2016-2019). The breakdown of the downloads of qualitative and quantitative data were only given in percentages; the actual download numbers for 2015-2020 here were pulled from “FSD Data Reuse 2015-2020 (Version 2.0)” (University of Tampere, 2021).

In terms of the annual reports, the Annual Report for 2014 (Finnish Social Science Data Archive, 2015a) is the only one available in English, and it is a complete translation of the Finnish report for 2014 (Finnish Social Science Data Archive, 2015b). The English document was compared with the Finnish report to build a list of Finnish-language variables and labels which were used in every report to ensure that the data collected from the other reports represented the same variables. The Annual Reports always present the current year’s number January 15, 2022, to account for any additional data sets which may have been published in 2021 which did not appear in the catalogue immediately (there were none).
for a given variable in comparison to the previous year’s, which also aided in tracing the values through the Finnish reports (for example: “In 2014, FSD received 1,068 data access applications (2013: 258) and submitted 1,403 data sets for reuse (2013: 597)” (Finnish Social Science Data Archive, 2015a, p. 5). From the 2015-2020 Annual Reports, only the total number of data sets published per year (separate from the data sets deposited), with break downs by qualitative, quantitative, and open access, were added to the aggregated data sets (the other data provided in the Annual Reports had already been collected from the other sources listed above, and the 2015-2020 Annual Reports only confirmed them).

The 2010-2014 Annual Reports (Finnish Social Science Data Archive, 2011, 2012, 2013b, 2014c, 2015b, 2015a) mainly provided the following: the total number of data sets published per year (separate from data sets deposited), break downs by qualitative, quantitative, and open access, applications for access to data, total data sets delivered for reuse, and total registered FSD (reflecting both the pre-Aila Data Service totals) users. However, 2014 was the year that the Aila Data Service came into operation (beginning in May), and so the numbers for all of these variables for 2014 are not directly comparable to either the 2010-2013 or the 2015-2020 numbers, which are themselves not directly comparable with each other. Before 2014, for example, a single application to access data could request multiple data sets; through the Aila system, each application is for a specific, individual data set. Additionally, before October 2014, a small number of data sets which were archived outside of FSD had149 the data description in the FSD data catalogue, and counted among the annual published data sets. In October, these descriptions were removed from the catalogue, and the only data in Aila now are those which are actually archived with FSD. The Annual Reports do identify how many data set descriptions only were published per year.
From DANS, I received and used the following administrative data:

- DANS Data sets in EASY (2007-2020), per year and cumulative
- DANS Data sets in EASY by discipline by year (2007-2020)
- DANS Distribution of Data sets by Access Category (biannual, 2012-2018)
- DANS Annual Visits to EASY (total and registered users) (2007-2019)

All these data were e-mailed to me in January 2022 (Doorn, 2022), and are the underlying 2007-2019 data reported in Doorn’s (2020) article describing the archiving services at DANS. While Doorn plans to deposit these data formally with DANS, as of this writing, the data have not yet been deposited, and cannot be independently verified. For the “Data sets in EASY” and the “Data sets in EASY by discipline” data, I added the 2020 values from the NARCIS portal (DANS, 2019a) reflecting the DANS EASY numbers, according to instructions provided in the email communication with Doorn. In 2020, the DANS access categories changed, and the numbers provided in the e-mail do not include the counts from 2020 or reflect the recategorization of many data sets (see Section 4.3.5). The count totals by access category for 2020 are not available; the count totals for 2021 are, and will be reported in Section 4.3 and discussed with the 2012-2018 data in Section 4.3.7. The 2020 data for Annual Visits and Data set Downloads have not yet been reported and cannot be included in these data.

From GESIS, I received and used the following administrative data:

- Download statistics GESIS Data Archive (GESIS - Data Archive for the Social Sciences, 2019)
- Current archive data set counts by access class (2022) (Prinz, 2022a)
• GESIS Data Archive data set counts by access class (2017-2022) (Zenk-Mölten, 2022)
• GESIS Data sets in SowiDataNet|datorium by year (2014-2021) (Prinz, 2022b)
• GESIS Data sets in Data Archive and SowiDataNet|datorium by year (2010-2021)
• GESIS Data sets published in Data Archive and SowiDataNet|datorium by year (2010-2021)
• GESIS Data sets published by access class in SowiDataNet|datorium by year (2011-2021)

The “Download statistics GESIS Data Archive” data set is a published data set available for all GESIS users (GESIS - Data Archive for the Social Sciences, 2019). It consists of aggregated download counts for all data sets in both the Data Archive and SowiDataNet|datorium from 2004 to 2018, based on the four different access points to those data sets: the Data Catalog (DBK), HISTAT, the data service, and SowiDataNet|datorium. From this data set, I identified the number of Data Archive data sets published by access category per year, annual Data Archive data set download counts by access category, and annual SowiDataNet|datorium data set download counts. Due to GESIS’ response to the 2019 hack of its account registration databases, GESIS no longer records the usage statistics as reported in this data set, and these statistics are not reported in another form elsewhere.

I received four data sets through email from GESIS staff. The first is the usage statistics for the Secure Data Center (Wiltshire, 2022). These are: the number of Secure Data Center visitors, the number of requested output checks, and the total number of days the Secure Data Center was used by year, from 2014-2020. The 2020 data are only from January 1, 2020, to March 5, 2020, because the Secure Data Center closed in 2020 due to the Coronavirus lockdowns in Germany, and did not re-open until spring 2022. The second email contains the current (as of April 2022) counts of Data Archive data sets by access category (Prinz, 2022a). The third contains the annual end-of-year total counts for the Data Archive, as determined
through the data catalog log files (Zenk-Möltgen, 2022); these counts provided the end-of-year totals by access category for the Data Archive for 2019-2021 to supplement the other data. The fourth email contained the annual end-of-year total counts by access category for SowiDataNet|datorium (Prinz, 2022b). While I did create three aggregated data sets with SowiDataNet|datorium data (see below), I was notified that there were discrepancies in the reported totals, and they provided a revised count by access category based on the log files for the self-archiving platform.

Figure 3.1 GESIS Data Archive and SowiDataNet|datorium data are searched for through the main GESIS webpage’s search function. There is no separate research data search function (GESIS Leibniz Institut für Sozialwissenschaften, n.d.f).

I created two data sets, “GESIS Data sets in Data Archive and SowiDataNet|datorium by year (2010-2021)” and “GESIS Data sets published in Data Archive and SowiDataNet|datorium by year (2010-2021),” by aggregating results from the GESIS search results (Fig. 3.1) (GESIS Leibniz Institut für Sozialwissenschaften, n.d.f). From this aggregation, I documented the end-of-year total numbers of data sets (by setting the “publication year” range to 1945 to [given year]), as well as the year-by-year publications (by setting the “publication year” range to [given year] to [given year], as seen in Fig. 3.1), for each archive. It is not possible to search for or
filter data sets by access class. For the Data Archive counts, there is some mismatch between the numbers I aggregated in this process for 2010-2018, and the data set published by GESIS described above. When in doubt, I used the counts provided in these aggregated data sets, as recommended in the email exchange referenced above, due to conflicts related to versioning as recorded in the log files for the Data Archive (Prinz, 2022a).

As referenced above, I created the “GESIS Data sets published by access class in SowiDataNet|datorium by year (2011-2021)” data set from the SowiDataNet|datorium webpage, which has its own portal to access the self-archived data through (GESIS Leibniz Institut für Sozialwissenschaften, n.d.n). This portal allows users to search for SowiDataNet|datorium data by publication year and by access class (the self-archived data access classes are not the same as the Data Archive access classes; this will be discussed in Section 4.4.1). While I ultimately refer in Section 4.4.7 to the log file-based data (Fig. 4.23), this is a late revision to this dissertation, and I find it important to clarify how I aggregated the data from the platform.

### 3.4 Data analysis

My data were primarily textual: interview transcripts and the documentation I collected from the websites and the respondents, though as described in the previous section, I also collected archive usage statistics. In accordance with recommended qualitative data analysis best practices (Miles et al., 2014; Yin, 2018), my analysis of these data began with the pilot study at FSD; I revised and expanded the interview protocols and developed a preliminary qualitative code set, as well as identified potential respondents.

All interviews were coded and analyzed in the computer-assisted qualitative data analysis software, NVivo 1.5.1. The development of the qualitative codes, which were used to identify themes in the interviews, was done through inductive coding of the data (LeCompte & Schensul,
2012; Miles et al., 2014). During the data collection phase, as discussed above, I wrote a reflection after each interview that summarized the responses, identified emerging themes and patterns, established the relationships between respondents, and helped develop a timeline for each archive’s particular experiences. These reflections were used to guide the refinement of the interview protocols and the development of the qualitative code set. The qualitative code set used for analysis can be found in Appendix D.

The qualitative code set was developed in several phases and continued to be revised until the end of the interrater reliability process. The initial code set, developed during the pilot study, primarily came through the inductive coding discussed above. There was a baseline code set based on the questions asked; the pilot study responses were used to develop child codes. However, I did identify potential codes about potential organizational changes (e.g., Brodin, 2019; Freitas & Mira da Silva, 2018; Lopes & Oliveira, 2018; Sirur et al., 2018; Urban et al., 2019) and challenges (Almeida Teixeira et al., 2019; Sirur et al., 2018) related to the GDPR from the literature review, which were incorporated into the code set formally after the pilot study to bolster the codes developed through the inductive coding. As part of the solidification of the qualitative code set, interrater reliability (IRR) was established with the assistance of a colleague experienced in qualitative data analysis. Three interview transcripts were independently coded, and I compared the results. Using the statistic Scott’s Pi (W. A. Scott, 1955), a measure for interrater reliability of qualitative data coding between two coders, with the third transcript we achieved a score of 0.728, demonstrating that the data and the qualitative code set were understandable and usable by others. The IRR process also resulted in further revisions to the code set; these were clarification of definitions, renaming several codes to prevent confusion, and addition or recategorization of codes based on my colleague’s feedback and perspective on
the data and the code set. Once sufficient IRR was achieved, I coded all remaining interview transcripts, and re-coded the pilot study transcripts from 2020 under the new code set.

The documentation that I collected from each archive included, where possible, pre- and post-GDPR versions. From this, I was able to review the different versions and identify where and in what ways changes were made to the various policies, legal agreements, and information for staff, data producers, and data reusers which were traceable directly to the archive attempts to respond to the GDPR. Examples of changes I looked for include: where text specific to the GDPR was added (and what it replaced, if anything); whether that revised text correlated with changes in policies, processes, or obligations for any parties; whether that revised text added clarification to prior policies, processes, or obligations; and, whether the document was a new creation to address a specific GDPR requirement. I also used the qualitative code set to review the documentation in NVivo; codes from the “Data at the archive,” “Data protection and security measures,” “Impact of the GDPR,” and “GDPR organizational changes” sections of the code set (Appendix D) were used in the analysis.

The data from the archive usage statistics were plotted over time to provide trends in archive usage over the time period of interest. These graphs are provided in the relevant case reports in Chapter 4. As discussed in Section 3.3.3, while I sought to collect the same usage data for data deposits and publications, annual end-of-year total numbers of data sets, download or other usage statistics, requests for access to restricted data, and account registrations from all four data archives, these data were not available from all data archives. Indeed, I was not able to obtain all five types of data from any of the data archives. There are several reasons for this. At ČSDA, this information is not available through the data archive itself; the end-of-year total data set counts, records of annual data set publications, and end-of-year total number of register
accounts are only reported on an annual basis through the Institute of Sociology’s annual reports. These reports do not include any other record of ČSDA or Nesstar platform usage besides the number of registered users, and there are no other publications or reports discussing ČSDA usage. Therefore, the operational metrics discussed in Section 4.1 refer only to these data. FSD also reported annual total data set publications (including by qualitative and quantitative), requests for access to restricted data, total data sets delivered for reuse, and total registered Aila users in its Annual Reports and annual news releases.

FSD and GESIS have both archived data sets containing usage statistics; however, they do not reflect the same types of use. “FSD Data Reuse 2015-2020 (Version 2.0)” contains data set downloads counts by month from 2015-2020 (which allow for a breakdown by qualitative or quantitative, access category, and download purpose, a required field when downloading data through the Aila portal) (University of Tampere, 2021). The “Download statistics GESIS Data Archive,” on the other hand, presents aggregated download counts for both the Data Archive data sets (which allow as well for a breakdown of data set downloads and publication year by access category) and SowiDataNet|datorium from 2004-2018 (GESIS - Data Archive for the Social Sciences, 2019). However, these data do not include any data after 2018, due to policy changes related to usage statistics made after the 2019 hack of the registration system (Section 4.4.6).

FSD, DANS, and GESIS responded to requests for additional data not available in the above referenced sources. The additional FSD data supplemented the data available in the archived data set and what I aggregated, and where there was overlap between my aggregated data set and the data deposit and data download counts from the emails, there were no inconsistencies. At DANS, these data were the underlying data for a 2020 publication about
DANS’ archiving services (Doorn, 2020, 2022) and will eventually be collated and deposited in the DANS archive as FSD and GESIS have done. Due to the design of the EASY platform, the ability for data providers to change the publication dates of their deposited data, and the on-going recategorization of the data sets to the new access categories (Section 4.3.5), the EASY platform does not provide stable by-year filtering of archived data sets. Therefore, the reported values in Section 4.3.7 reflect the snapshot of values captured for the publication, and the additional 2020 values I captured using the instructions Doorn provided.

There were inconsistencies between the aggregated annual end-of-year total data I created from the GESIS search results (the primary access point to the data in both the Data Archive and SowiDataNet|datorium\(^{102}\)) and the counts from the deposited data set discussed above. I contacted DSS staff about these inconsistencies and was provided with another version of the annual end-of-year total data set counts based on the data catalog files, which more closely aligned with the aggregated values than the published data set. As discussed with the Data Archive technical staff, these inconsistencies are related to data set recategorization that had occurred since the data set was first published in the Data Archive. Therefore, in consultation and agreement with the technical staff, I used the aggregated values where there were inconsistencies.

Prior to beginning this aspect of my data collection, I intended to compare changes in annual data deposits across the four data archives, as opposed to data set publications. “Data deposit” in this context refers to the provision of individual data sets and associated documentation by the data provider to the data archive, while “data set publication” refers to the

\(^{102}\) As referenced in Section 4.4, there are two access points to the self-archiving platform SowiDataNet|datorium: the platform’s own portal and regular GESIS search feature. The counts are the same in both platforms; however, more specific information about the data sets (e.g., access category, publication date, funder, geographic coverage, etc.) are available as filters in the portal than in the GESIS search results.
point where the data set has been processed and made available for reuse (unless under embargo, in which case the “publication” refers to the metadata record) through the data archive’s web portal. Due to the time required for data archive staff to review the data, carry out any curation activities, confirm any necessary or recommended changes, and review the data set, the data may not be made available in the same calendar year as they were deposited with the archive. However, while all four data archives reported their annual data set publication numbers, only FSD also reported data deposit counts per year. For consistency, therefore, this dissertation will refer only to the data set publications in each of the case reports in Chapter 4.

3.5 Validity

Within social research, the question of how to evaluate the quality of research is an important one; it is even more so for case study research. Four tests of validity—construct validity, internal validity, external validity, and reliability (J. W. Creswell, 2013; Miles et al., 2014; Newman & Benz, 1998; Yin, 2018)—have been employed in this dissertation to ensure that I present an accurate and representative account of the experiences of the four archives.

A common criticism of case study research is whether the research instruments and measures address the concepts and questions under investigation (construct validity). Flyvberg (2006) identifies the subjectivity inherent in case study research as an influential criticism of the approach (see also Yin, 2018). Yin (2018) recommends three tactics to improve construct validity in a case study: base the analysis and findings on multiple sources of evidence, preserve the chain of evidence, and conduct member checking of the findings and interpretations. I implemented these recommendations throughout my research: triangulating the data that I collected through the interviews, document analysis, and evaluation of the usage statistics; maintaining a secure file of all documentation, transcripts, recordings, interview reflections, and
archive statistics in a secure Dropbox folder, with a backup copy on a password protected
Google Drive; and, requesting clarification of statements and events in the case reports from the
interviewees. During the data analysis phase, I requested clarification and asked follow-up
questions where necessary, and the completed case reports were shared with my interviewees.
Their feedback was incorporated into the final versions of each report.

Internal validity tests “whether the researcher provides a plausible causal argument,
logical reasoning that is powerful and compelling enough to defend the research conclusions”
(Gibbert et al., 2008, p. 1466), and whether inferences can be supported (Gibbert et al., 2008;
Yin, 2018). As I conducted the interviews and reviewed the documentation, I identified policy
and operational changes at the archives, and investigated further into whether those changes were
directly or indirectly a response to the GDPR, and where not, what the causes were. Yin (2018)
argues that this is the most appropriate tactics for explanatory case studies attempting to
demonstrate causal relationships.

A particular challenge of case study research is that case studies are not statistically
generalizable to wider populations. Yin (2018) presents analytic generalization as another type of
generalization that can be achieved in research (external validity), and Eisenhardt (1989, 2002)
recommends the use of cross-case analysis to enhance this analytic generalization. My data
analysis incorporated both within case and cross-case analysis to understand how four data
archives responded to and were affected by the GDPR.

The final test for reliability “assesses the extent to which the results and conclusions
drawn from a case study would be reproduced if the research were conducted again” (Mills et al.,
2010). Even though, when compared with experimental research or even survey research the
likelihood of case study research being repeated is very low, case study researchers must still be
concerned with the reliability of their research. To enhance the reliability of this research, I developed a case study protocol which follows Yin’s (2018) five components of case study research, and I developed a case study database to collect and store all case study notes, data, reflections, and other documentation in an organized manner that preserves the chain of evidence (Gibbert et al., 2008; Yin, 2018).

3.6 Limitations

There are three primary limitations to this study of European Union social science data archives. The first is that, as a case study, the results of this research cannot be generalized to all data archives, or even all data archives within the European Union. Instead, case studies can establish analytic generalizability and aid in the development of theory. Case studies, however, do provide opportunities for very rich, intensive understanding of the phenomenon of interest: in this case, how the GDPR was and is still experienced at four different data archives. As well, I am only investigating four social science data archives. It is possible that other data archives, whether focusing on social science data or not, in the European Union may provide insights and perspectives on the GDPR which will be missed in this study, further limiting the generalizability of the findings.

The second limitation is that I am only addressing the data archive in this dissertation. It is possible that changes within data archive operations are not the only motivators for any potential change in data archive usage. The same discussions and concerns about the GDPR’s effects on research could also influence whether and how primary investigators, data providers, and data users use the data archives. This dissertation is only focused on the data archive side, and it is possible that data providers and/or data users may have additional insights on the GDPR’s impacts on the data archive which will be missed.
The final limitation is linguistic. I am a native English speaker, and I have experience reading and speaking German at an upper-intermediate level, Dutch at an intermediate level, Czech at a beginner level, and have no experience with Finnish. While I was able to conduct all my interviews at the four archives in English, and most documentation overall is available in English as well as the original language, the language differences did occasionally prove challenging. Ultimately, it was not necessary to use non-English versions of my interview protocols, as all respondents, including the two respondents who provided written responses, preferred to take part in English. There were a small number of instances (eight question responses out of 70 hours of interviews) where a part of a response was given in Czech and Dutch, and translations were later confirmed with the respondents. As discussed in Section 3.3.3, I created several data sets related to data archive usage from published reports. For the Czech and Finnish data sets aggregated in this way, I developed potential lists of search terms for data of interest from various sources and translations, including documentation with English versions, and validated them with native speakers and against similar documentation.
Chapter 4 Case Study Reports and Findings

In this chapter, I present the case reports for each of my four research sites: the Czech Social Science Data Archive, the Finnish Social Science Data Archive, Data Archiving and Networked Services, and GESIS-Leibniz Institute for the Social Sciences. Each case report is structured as follows: I begin with a description of the data archive and its structure, its place within the national research community, and national views of privacy and reactions to the GDPR. Then, I describe the timeline for the data archive’s response to the GDPR, the changes that were made, the challenges that they faced, and conclude with the impacts and outcomes that the data archives experienced from the GDPR. The changes the data archives made are presented through the three dimensions of my theoretical framework: regulating information flows, minimizing harm, and responding to technology. While there were a multitude of changes made at each archive, some were simply updates to policy language to reflect the GDPR and the new national data protection legislation without any changes to archive operations or archive user obligations. Others were more substantial changes that directly changed archive practices and the archive user experience. The discussions that follow in the case reports will focus primarily on the latter. I end this chapter with a comparison of these experiences, and a discussion of my findings.
4.1 Czech Social Science Data Archive (ČSDA)

4.1.1 Archive description

The Czech Social Science Data Archive (ČSDA) is a department of the Institute of Sociology (SOÚ), one of 54 legally-independent public research institutes of the Czech Academy of Sciences (AV ČR). ČSDA was established in 1998, and is the primary data archive for the empirical data collected by SOÚ researchers (Sociologický ústav Akademie věd České republiky, v.v.i., 2022a). In addition to SOÚ research data, ČSDA archives and makes available data from the Czech Panel Household Survey, Czech Public Broadcasting-conducted public opinion polls, and research data from 1971-1989. ČSDA is the Czech national node and service provider for the Consortium of European Social Science Data Archives (CESSDA), and it supports the Czech components of large-scale, multinational survey programs, including the International Social Survey Program (ISSP103) and the European Social Survey (ESS104).

ČSDA is one of two national research infrastructures in the SOÚ; the other is the ESS. As such, ČSDA is not an academic department; instead, it is a service provider supporting the long-term preservation of primarily sociological research data from Czech researchers, though it is starting to branch out to other disciplines. ČSDA supports the Czech research community through the data archive, training courses, and its involvement in external collaborations. It received its CoreTrustSeal certification in 2016, and was recertified in 2018 (CoreTrustSeal, 2018).

103 The ISSP began in 1984 as an international collaboration to include a common set of survey modules in national surveys. The ISSP topics include the role of government, social networks, social inequality, gender roles, health, and more. The research institutes conducting these surveys fund the ISSP themselves; there is no central funder.

104 The ESS is a biennial survey that has been conducted across European countries (22 in 2002, 32 in 2020) since 2001. The ESS investigates changes in social structure, social and political attitudes, gender issues, attitudes about climate change, immigration, and more. Data and documentation for the ESS is available through the ESS website.
4.1.1.1 Data archive

As a data archive, ČSDA is a “scientific institution giving service to a scientific community” (ČSDA01). The research data it archives come primarily from the SOÚ. However, it also publishes public opinion research data from the Czech Public Broadcasting company and the STEM Empirical Research for Democracy (Empirický výzkum pro demokracii), as well as research data from universities, research arms of government ministries, non-governmental organizations, and private organizations (for a full list of external data providers as of 2020, see Sociologický ústav Akademie věd České republiky, v.v.i., 2021, p. 94). It also publishes the Czech components of large-scale international projects, such as the ISSP and the “RESpondIng to outbreaks through co-creaTIve sustainable inclusive equality sTRatEgies” (RESISTIRÉ) project studying the gendered impacts of COVID-19 political responses (European Science Foundation, 2021).

The majority of the ČSDA data are from various SOÚ projects and the public opinion research conducted by the SOÚ’s Center for Public Opinion Research (Centrum pro výzkum veřejného mínění) (CVVM). It is SOÚ policy that these data be archived with ČSDA, and of the 40-50 data sets ČSDA archives each year, at least ten of them come from the SOÚ (the ten data set minimum is because the CVVM deposits at least ten opinion polls per year) (ČSDA01). However, there are three other paths for data to be deposited at ČSDA. The first is through contracts with external data providers. The agreements with the Czech Public Broadcasting company and with STEM are examples of this contracted data sharing. The second is through active data acquisitions. A typical scenario for this second path is:

… some person from Institute to Sociology tells me that there is some very interesting research, some interesting survey, and he or she has no access to the data. ‘Do you have the data?’ And, we don’t have it. So, I’m trying to find the data source and asking potentially the depositors if it will be possible to
send us data. So, it’s rather persuasion role, trying to persuade, communicate with potential depositors who we know that they can have some interesting data (ČSDA01).

The third path is for a researcher to contact the archive directly to begin the data deposit process. In order to deposit data, the data provider contacts ČSDA and provides information about the project, institutional affiliation, data type(s) and formats, and additional documentation (Český sociálněvědní datový archiv, 2022a); based on this information, ČSDA staff determine whether these data will be accepted for archiving. If the data are accepted, it is the data provider’s responsibility to ensure the data have been anonymized and otherwise comply with legal requirements (Český sociálněvědní datový archiv, 2022c; ČSDA01). The data provider and ČSDA negotiate a deposit agreement (which is ultimately between the data provider and the SOÚ) authorizing the transfer of the research data and associated documentation to ČSDA and granting ČSDA the license to preserve and disseminate the data. The data provider also prescribes any additional usage conditions, including any requirements for data provider consent for reuse, in this agreement.

The majority of data available through ČSDA are survey data stored as ASCII files with aggregated data tables; these data can be exported in different formats or analyzed online through the Nesstar catalog (Český sociálněvědní datový archiv, n.d.). There is also a small collection of qualitative data sets, primarily interviews, held in the separate Medard archive. ČSDA users can access the metadata for both the quantitative survey data and the qualitative data through the Nesstar catalog (Fig. 4.1). Using any of the quantitative data requires first logging into the ČSDA account, and then into the Nesstar account. ČSDA changed their Nesstar user database in 2020 (Section 4.1.5.5), and there are now two distinct logins required. Most of the data are Czech language only, though around 100 data sets are also available in English (ČSDA01).
The quantitative data are available for use under three access categories:  A (data sets are available under “standard conditions”); B (data sets are available under “standard conditions” as well as any additional conditions set by the data provider); and, C (data sets which also require data provider permission) (Český sociálněvědní datový archiv, 2022b). There is technically a fourth category, 0 (data sets available online without registration), still listed on the ČSDA webpage (Český sociálněvědní datový archiv, 2022b); however, this is a legacy category:

This is a historical division. For example, the category ‘available online,’ in the past of the archive, there was some data sets produced in the Institute which were totally freely available, for downloading for everybody without any restrictions. And now, we don’t have any such data sets (ČSDA01). According to ČSDA02, ČSDA adopted these categories in 1998 when ČSDA was established, basing them off GESIS’ access categories (Section 4.4.1.1). The “standard conditions” are that: the data user is prohibited from attempting to re-identify data subjects; any publication must
include a citation to the data set; the data user is prohibited from further dissemination; and, the
data user must notify ČSDA of all publications (Český sociálněvědní datový archiv, 2022b).

Access to and use of the quantitative data is free through the Nesstar catalog no matter
the access category. If for some reason ČSDA must distribute the data to a user via another
method (e.g., email, mail, etc.), there may be additional associated costs. It is also free for data
providers to deposit their data with ČSDA, because “we are scientific institute giving service to
scientific community, and we are paid by Ministry of Education of Czech Republic [MŠMT ČR]
for giving these services as a scientific infrastructure” (ČSDA01). The baseline requirement for
all three in-use access categories (A, B, C) is that the user must register for a ČSDA account to
download or analyze the data through Nesstar. Once the user registers their account, the separate
Nesstar account is created automatically. The qualitative data in Medard are not available
through the Nesstar platform. Each Medard metadata record includes requirements for access to
a given data set, which may include simply ČSDA registration and a request to ČSDA archive
staff, data provider permission, additional citation requirements, or other requirements.

Beginning in fall 2021, ČSDA has begun transitioning to and testing a Dataverse-based archiving
platform which will eventually replace the Nesstar catalog. The reason behind this change is
because

Nesstar is not properly maintained. On one side, it’s better for users we
believe, because they can go into details on the level of variables. But it’s not
properly maintained, and there is more and more problems with that. It’s not
safe, also (ČSDA02).

The new Dataverse platform is not currently publicly accessible, and it is not known at this time
how this new platform will affect the access processes for the qualitative data.

All publicly available research data at ČSDA are considered fully anonymized. This has
been the policy since ČSDA was established, in part because
the kind of data were available for the archive was this kind of data. Data from large, anonymous, representative surveys, and one or two years after archive was established, the quite strict law of the personal data protection I was talking about was passed in Czech Republic. That was another reason to avoid some other kind of data with some personal data or some possibility of disclosure (ČSDA01, see also Krejči, 2002).

Anonymization standards became a data ingest concern as ČSDA began archiving data from external partners; for the SOÚ data, it was perceived that the data archiving and publishing data is safe, because it was produced by colleagues in Institute of Sociology, and archiving these data from the Institute, it was covered by some rules issued by head of the Institute of Sociology. … But when you take most of the data from outside of the Institute, from other institutions, you have to be more cautious, more correct (ČSDA01).

Therefore, all data listed in the Nesstar catalog are considered to be anonymous. However, ČSDA does physically store data sets with personal data which are not available to the public. These data are from SOÚ researchers; for example,

… some original versions of the data from the past of the Institute where there are personal names, or some specific survey of some persons who are taking care of children in public administration, and it’s quite small, basic population so it can potentially reveal the real identity of the person. So again, we have this very specific data set in the safe. It’s recorded on the hard disk, which is protected by password (ČSDA01).

These data are only available to SOÚ researchers, and only in an on-site, secure study room within the SOÚ library; other than this secure study room, ČSDA does not “have the structure to provide public access to confidential data” (ČSDA02). In general, as discussed above, anonymization is the data provider’s responsibility. Since ČSDA has not adopted a processor or controller role (see Section 4.1.5.2) for the data available through Nesstar because it considers the research data to be anonymized, “we don’t want to do anonymization because it means that you are working with personal data and we don’t want to work with personal data, especially when they are from the outside of the Institute” (ČSDA01). In the SOÚ service role, however,
ČSDA is “sometimes working on anonymization or collaborating on anonymization quite extensively” (ČSDA01) with the SOÚ researchers. This is possible because the SOÚ is the controller for its internally-produced research data, and ČSDA-as-SOÚ-department is a controller.

4.1.1.2 External collaborations

In addition to its role as a data archive, ČSDA is involved in a number of external collaborations, on its own as well as an SOÚ department. ČSDA has been the national node for CESSDA since 2001 and supports research and development such as: the BeYond-COVID (BY-COVID) project to develop an online data portal to access, share, and analyze infectious disease data; technical and organizational support for the Czech components of the ISSP; and, working with the SOÚ’s Value Orientations in Society (Hodnotové orientace ve společnosti) department to support the Czech node of the ESS. Within the Czech Republic, ČSDA has been heavily involved in the Czech Household Panel Survey (České panelové šetření domácností), a study tracking Czech social trends since 2015. The data are available to registered ČSDA users though the Nesstar catalog.

4.1.1.3 Organizational structure

ČSDA is a department of the Institute of Sociology (SOÚ); as an SOÚ department, ČSDA is both a data archive and a service provider for the SOÚ (Sociologický ústav Akademie věd České republiky, v.v.i., 2022a). Unlike the other departments, ČSDA is not considered an “academic” department, although ČSDA01 feels that occasionally ČSDA is pressured to produce more academic (i.e., research) outputs. As a service provider as well, especially during the GDPR transition period, “we are not only data archive, in fact, we are department which should,
in the Institute of Sociology, be responsible for giving advice on personal data protection or GDPR” (ČSDA01).

Within the SOÚ, in general, ČSDA is relatively free to determine its operations outside of its obligations as that SOÚ service provider. To that end, ČSDA is not significantly affected by Academy of Sciences of the Czech Republic (AV ČR) oversight either. Each AV ČR institute is its own legal entity, and while in some areas (e.g., certain Academy-wide strategies, policies, or targeted budgetary aims), “the [AV ČR] can make decisions which are binding for all these institutes” (ČSDA03), in others, the institutes direct their own operations. This is evidenced in that, while the AV ČR did establish a GDPR preparation team before the GDPR entered into force, that team focused primarily on the AV ČR, and the AV ČR also decided that each institute would direct its own GDPR response (ČSDA04). Despite this independence, the AV ČR does evaluate each institute once every five-year period based on a review of: research outputs; national, European Union, and foreign research grants and projects; research outreach; research infrastructures; strategy and mission; and more (Akademie věd České republiky, 2021).

The AV ČR funds the ČSDA through the SOÚ, though according to ČSDA01, the AV ČR funding is mainly dedicated to personnel. A significant portion of ČSDA’s funding is from the MŠMT ČR to support its services as part of the national research infrastructure (Ministerstvo školství, mládeže a tělovýchovy České republiky, 2010). ČSDA also receives some funding from non-governmental Czech research funders and international funders to support the different projects in which it takes part. ČSDA can also apply for any sociology or social science grants it deems appropriate without restriction from either the SOÚ or the AV ČR.

ČSDA has a Scientific Advisory Board (Vědecká rada Českého sociálněvědního datového archive) which advises the SOÚ director. Currently, the board comprises six members
representing two Czech universities, as well as data archive staff and university faculty from Romania, Germany, Lithuania, and Estonia (Český sociálněvědní datový archiv, 2022d). The Board’s responsibilities are to: discuss and advise on long-term strategic plans, provide independent advisory services, and address any SOÚ director concerns (Krejčí, 2016). Members are appointed by the SOÚ director for four-year terms. According to ČSDA01, the board did advise on planned major changes during the GDPR transition period; in particular, the decision to develop and transition to the forthcoming Dataverse-based archiving platform.

The Office for Personal Data Protection (Úřad pro ochranu osobních údajů) (ÚOOÚ) is the data protection authority for the Czech Republic. It was first established by the 2000 Czech implementation of the European Union’s Data Protection Directive as the independent central authority supervising personal data processing in the Czech Republic. It advises on data processing and conducting data protection impact assessments, investigates complaints and data protection violations, and raises public awareness of data protection and privacy. Recent investigations include: responding to complaints about a city’s public survey which collected “mandatory” personal data without providing information about how these data would be processed or for what purpose (Úřad pro ochranu osobních údajů, 2021a), and complaints against a Czech university’s mandate that applicants provide both personal and special categories of data during the application process without the appropriate and required legal bases (Úřad pro ochranu osobních údajů, 2021b).

4.1.1.4 Legal framework for ČSDA

The GDPR and the 2019 Act on Personal Data Processing (ZZOÚ) are not the only laws directly influencing ČSDA’s operations. There are two laws, the 1992 Act on the Academy of
Sciences of the Czech Republic\textsuperscript{105} and the 2005 Act on Public Research Institutes,\textsuperscript{106} which together establish ČSDA’s parent institute, the Institute of Sociology, as a legally independent entity, allowing ČSDA significant freedom in its own strategic planning. While there are other laws related to personal data protection (e.g., laws targeting health services, social security, health insurance, etc.), these laws do not apply to ČSDA or its operation.

### 4.1.2 Data archiving in the Czech Republic

ČSDA is the only social science-focused data archive in the Czech Republic. According to the Registry of Research Data Repositories, there are ten currently active Czech data archives (re3data.org, 2022). Of these, there are two linguistics repositories and two archives of map collections for archaeology and cartography. The interdisciplinary nature of these archives means that they are classified as “humanities and social sciences” on the Registry; however, ČSDA is the only archive for research data from sociology and other fields of social science research. Within the Czech Republic, ČSDA’s parent organization, the Institute of Sociology (Sociologický ústav Akademie věd České republiky) (SOÚ), is considered the focal point of Czech social science research, and ČSDA is “a unit of national relevance” (Illner, 2002, p. 413). This perspective is shared at ČSDA itself; as ČSDA01 described, “our position is good, well-established, strong. We have quite a well-defined role, not only as a part of the Institute, but as a part of Czech social sciences.” According to Jarolímková and Drobiková, the “social sciences are also the only discipline in the Czech Republic that has its own national data archive” (Jarolímková & Drobiková, 2019, p. 556).


172
There is no legal mandate in the Czech Republic for research data to be deposited in a data archive. According to ČSDA01, there were several attempts to make data sharing a formal mandate for funded research to be shared, for example, through the AV ČR, the Ministry of Education, Youth and Sports of the Czech Republic (Ministerstvo školství, mládeže a tělovýchovy České republiky) (MŠMT ČR), or the Czech Science Foundation (Grantova agentura České republiky) (GA ČR). None were successful. There is a formal policy for data sharing from the SOÚ, however:

… archiving these data from the Institute, it was covered by some rules issued by head of the Institute of Sociology. It was decided that all data which originated in the Institute, after some short time after the end of the project, has to be moved to the archive (ČSDA02).

Otherwise, this lack of a formal mandate is considered a barrier to data sharing and archiving, though both ČSDA01 and ČSDA02 feel that sociologists are more open to the idea than other social science disciplines, due to the SOÚ’s role in social science research (see also Illner, 2002).

Thus, while sociology as a discipline has a particular historical tradition in the Czech Republic (see Section 4.1.2.1 for more detail on this point), archiving sociology data is relatively new. This is especially when compared to data archives such as ICPSR, GESIS, or the UK Data Archive, which all have their origins in the 1960s (see Section 4.4.1 for more detail about GESIS). Due to targeted restrictions on sociology under the Communist regime (1948-1989), Czech-produced sociological data were frequently removed from Czechoslovakia to archives and institutions elsewhere, and even today, many of these data sets remain archived at non-Czech archives. After the end of the Communist regime, sociology became a popular and supported research area again. However, in recent years, ČSDA02 has witnessed some not very nice developments in the social sciences, that many people doesn’t want to base their research on data. More and more people go to that
softer part connected with qualitative methods or theory, and there is less interest in the quantitative methods of analysis than in the past. Which is a bit controversial because on the other side, there is, I would say, more demand for this kind of analysis from the society.

Czech sociology had long been grounded in math and statistics, and there has been a shift towards a more qualitative or humanities-based approach. Since ČSDA as an archive primarily archives quantitative data, this shift, if it continues, would result in significant changes in how ČSDA operates as an archive.

4.1.2.1 History of sociology in the Czech Republic

The Czech Social Science Data Archive’s place in the Czech research environment is deeply intertwined with the role of Czech sociology, scientific learning on a broader scale, and nationalism in the Czech Crown lands (Bohemia, Moravia, and Silesia), dating back to the 1784 founding of the Czech Society of Sciences (České společnosti nauk), re-named in 1790 as the Royal Czech Society of Sciences107 (Královská česká společnost nauk) (Akademie věd České republiky, n.d.). At the time of its establishment, the Czech Crown Lands’ “schools were German, the towns Germanized, and thus, German was the language of academia” (Čapek, 1938/1990). This trend continued for the next century. In 1890, the Czech-language Czech Academy of Emperor Franz Josef for Science, Literature, and the Arts (České akademie císaře Františka Josefa pro vědy, slovesnost a umění) was established, followed by the Association for the Fostering of German Science, Arts and Literature in Bohemia (Gesellschaft zur Förderung deutscher Wissenschaft, Kunst und Literatur in Böhmen) in 1891, both of which remained separate from the Royal Czech Society of Sciences. This was part of the late-19th century Czech National Revival, which sought to bring the Czech language and culture back into prominence at

107 Also referred to as the Royal Bohemian Society, depending on the source.
a time when, across Europe, there was a “wider and more complex phenomenon—the rise of modern nationalism” (Auty, 1956, p. 242).

Illner connects the start of Czech sociology to a decade earlier, in “1882, when teaching in the Czech language was renewed at Prague University”108 (2002, p. 405). The first Czech sociological work was Selbtsmord als sociale Massenerscheinung der modernen Civilisation by Tomáš Masaryk (1881/1970), published first in German and then in Czech in 1904. In 1882, Masaryk was appointed a professor of philosophy, and is considered to be the father of Czech sociology (Illner, 2002). Masaryk and his student, Edvard Beneš, would become the first and second presidents of Czechoslovakia (Československo) after the end of World War I, in 1918 and 1935, respectively. Their combined academic and political treatises were influential in the development of Czech democratic thought (Illner, 2002; Kerner, 1921; Skovajska & Balon, 2017). From its beginning, sociology has been closely connected with social reform movements in the United States and Europe (for an in-depth discussion of this, see Turner, 2014), but in the Czech Lands, it has always had a more influential, national status: “very much unlike in the large countries of the West, Czech sociology was intimately wedded, through the person of Masaryk and his followers, to the project of the cultural, social, and political emancipation of the Czech nation” (Skovajska & Balon, 2017, p. 14). Even as Masaryk’s role as the founder of Czech sociology was overshadowed by his political status as the founder of Czechoslovakia (as was Beneš’s, as the first Minister of Foreign Affairs and the second President), their connection to

---

108 Various sources refer to the university as “Prague University” (Illner, 2002) or “Czech University in Prague” or “Czech branch of the University of Prague” (Skovajska & Balon, 2017). The university referred to is the present-day Charles University (Univerzita Karlova). One of the oldest continuously operating universities in the world, Charles University was founded in 1348 as the University of Prague. It has undergone a number of name changes and changes in its academic freedoms as regimes changed, including during and after Nazi occupation in World War II and under the communist regime until 1990. (Fried, 2000; Koucký, 1990; Univerzita Karlova, 2022).
sociology would also be its downfall after the end of the World War II. As ČSDA02 explained, “always, sociology was connected to politics before 1948, before communist regime.”

Czechoslovakia’s founding in 1918 also saw the renaming of the imperial Czech Academy to the Czech Academy of Sciences and Arts (Česká akademie věd a umění) (ČAVU) (Akademie věd České republiky, n.d.). Elsewhere, there was growing political differentiation, with 39 political parties divided on ethnic, religious, linguistic, and regional grounds taking part in the 1925 election (Rose & Munro, 2003). Several new universities opened in Brno, on the eastern edge of the modern Czech Republic, including Masaryk University (Masarykova univerzita) as the second Czech-language university in 1919. Other research institutes and academies were also established outside of the ČAVU’s control during this period for many reasons, including poor funding, public reactions against funding “Austrian” institutions, and Masaryk’s own disillusionment with the ČAVU (Skovajska & Balon, 2017; Winters, 1994a).

Czech sociology continued to develop over this period as well, especially at Charles University and Masaryk University. This growth across the country in all areas “was interrupted by the Nazi occupation of the Czech Lands in March 1939 and by the closing of Czech universities soon thereafter” (Illner, 2002, p. 406).

The ČAVU remained operational during the Second Czechoslovak Republic (Druhá Československá republika), which existed between the signing of the Munich Agreement on September 30, 1938, and the Nazi invasion on March 15, 1939, and then through the entirety of the occupation (though in a severely reduced capacity). Nazi persecution of Czech Jews and the Czech resistance, before and after the 1942 assassination of acting Reich-Protector of Bohemia and Moravia, Reinhard Heydrich, saw the imprisonment and/or murder of members of the ČAVU, as well as other academics and researchers (Skovajska & Balon, 2017; Winters, 1994a).
At the end of World War II, the Association for the Fostering of German Science, Arts and Literature in Bohemia was dissolved. Beneš returned from exile and was confirmed as president once more on October 28, 1945 and re-elected the following June. The Czechoslovak government-in-exile drafted a number of decrees, known collectively as the “Beneš Decrees” targeting post-war recovery, reparations, and development. Among them were a number that resulted in the de-population of at least three million ethnic Germans and Hungarians and property seizures from collaborators and traitors (Ahonen et al., 2008; Brown, 1958; Hámos, 2019; Korbel, 2015; Kuklík, 2015; Winters, 1994b). The effect this had on the liberated Czechoslovakia, and Czech sociology and academia in particular, would prove catastrophic to the Third Czechoslovak Republic (Třetí Československá republika).109 The demographic changes brought by the expulsions and the internal resettlements across the country, especially in the Sudetenland, provided a foothold for the Czech Communist Party (Komunistická strana Československa) (KSČ), which had already been growing in influence as part of Beneš’s post-war coalition government. In the Sudetenland, the KSČ received 75% of the votes in the 1946 general election, nearly half of the total national votes for the KSČ (38% overall, the largest of all parties) (Bernášek, 1970; Korbel, 2015).

During the war, Nazi collaboration by Czech sociologists was rare, and even during the Nazi occupation, sociology “became attractive for a number of left-wing young people” who later joined the KSČ after liberation (Skovajska & Balon, 2017, p. 39). Other Czech sociologists fully supported the Beneš government’s plans for post-war recovery, while within institutions, Czech sociology was shifting in political leanings from the pre-World War I socialist (but not

---

109 The Third Czechoslovak Republic lasted from the end of World War II in 1945 to the February 1948 Communist coup, even though the country’s name remained the “Czechoslovak Republic” until 1960 when it became the Czechoslovak Socialist Republic, which lasted until March 29, 1990. For consistency, after this except where relevant, I will refer to the country that existed from 1948-1990 as “Czechoslovakia” as that is how it was referred to by scholars and my interviewees.
Marxist) to Marxist-Leninist. Just before the 1948 coup, communist student groups dominated Czech universities (Connelly, 2000; Skovajska & Balon, 2017). At ČAVU, the new director had been a communist party member before the war and spent the war years in exile in Moscow. Under his tenure, and with the Beneš government’s compliance, ČAVU came to politically support the Soviet Union. In 1946, one ČAVU member, Charles University sociologist Josef Král, proposed establishing an Institute of Sociology at the ČAVU, which was supported but then abandoned due to the 1948 coup (“Král Josef,” 2018). This was the first attempt at establishing a national sociology research infrastructure in Czechoslovakia.

From 1948 to the 1960s, as ČSDA01 described, “there was no sociology in Czechoslovakia, because Communist regime and Stalinist era were enemies of sociology.” After the coup, communist student groups were at the forefront of student and faculty purges (Connelly, 2000), and sociology was a very early target (Musil, 2011; Skovajska & Balon, 2017). ČSDA02 described the communist regime’s view of sociology as follows:

If you look specifically to Czechoslovakian history, the first president Masaryk was… he didn’t call himself sociologist, but in fact, he was. … and also the second president, Edvard Beneš, was sociologist. There are a lot of social science research used for purposes of the government. So, always sociology was connected to politics before 1948, before communist regime. And then they called sociology as ‘bourgeois’ pseudoscience, and they forbid that, but at the same time their economy was based on planning. So, they needed the data, but these data were produced under different other labels than sociology. ČSDA02’s description of sociology from the KSČ’s perspective as being a “bourgeois” science is reflected in Musil’s discussion of why Marxism and Czech sociology were incompatible.

Sociology as a theory ‘system of society’ … was bound to come into conflict with historical materialism, which was not only a theory of society but which was also the ideological basis of revolutionary practice. For the Marxists, sociology was a bourgeois science, and for sociologists, historical materialism was a one-sided
interpretation of the functioning of particular societies. The sharpest conflicts arose, therefore, because both sociology and historical materialism were theories of the structure and dynamics of society (Musil, 2011, p. 393).110

As a discipline, sociology disappeared for several years in Czechoslovakia, though, interestingly, not in neighboring communist Poland (Illner, 2002; Vakakova & Strmiska, 1983).

In 1948, all non-university research institutions and societies (except, technically, the ČAVU and the Czech Society of Sciences) were closed by the new regime. In 1953, the Czechoslovak Academy of Sciences (Československá akademie věd) (ČSAV) was founded; § 1(2) of the Act111 establishing the ČSAV grounded its establishment in the historical traditions of both the 1784 Royal Czech Society of Sciences and the ČAVU, while also finally closing both institutions. The Slovak Academy of Sciences (Slovenská academy vied) was established later in 1953 as well, initially as a separate institution (it was incorporated into the ČSAV in 1960). Both the ČSAV and the Slovak Academy of Sciences highlighted their historical traditions despite serious opposition from those who favored “a fresh start using only the model of the Soviet Academy of Sciences” (Winters, 1994a, p. 76). The ČSAV expanded after 1953, incorporating or establishing numerous institutes focused on scientific and technological research innovation (in 1953, ČSAV comprised seven institutes; by 1969, there were 138), all under strict ideological control (Skovajska & Balon, 2017; Winters, 1994b).

Beginning in the 1960s, the return of sociology as a supported discipline was one of the outcomes of the delayed Czech de-Stalinization efforts that Soviet Premier Nikita Khrushchev


111 § 1(2), Zákon č. 52/1952 Sb. o československá akademie věd (Act no. 52/1952 on the Czechoslovak Academy of Sciences): “Akademie vzniká přebudováním Královské české společnosti nauk, založené r. 1784 a České akademie věd a umění, a navazuje na pokrokové tradice české a slovenské vědy” (The Academy is a rebuilding of the Royal Czech Society of Sciences, founded in 1784, and the Czech Academy of Sciences and Arts, and builds on the progressive traditions of Czech and Slovak science). (Translation by the author).
had called for in 1956 (Illner, 2002; Skovajska, 2008; Sommer, 2016). After a period of ideological realignment that also ended in 1956, a form of Marxist sociology developed in Czechoslovakia, with very close connections to philosophy, a discipline that had not been abolished in 1948; the Institute of Philosophy was a founding institute of the ČSAV in 1953 (Hladík, n.d.). The return to sociology was also prompted by the growing realization that societal reorganization required knowledge about society, though this was not universally accepted. Czechoslovakia in the early 1960s was in economic difficulty, and an empirically-focused sociology was viewed as a tool to help resolve it (Illner, 2002; Skilling, 2015; Skovajska & Balon, 2017; also ČSDA01). Research from both the Institute of Philosophy and the Institute of Public Opinion Research (established in 1965 along with the Institute of Sociology) focused on social stratification, modernization, and cultural contexts for development. 1967 also saw the first attempt to establish a social science data archive by Michael Illner (later director of the post-communist Institute of Sociology), who had been inspired by the work of Ekkehart Mochmann, director of the predecessor of Germany’s GESIS-Leibniz Institut für Sozialwissenschaften (Illner, 1968; ČSDA02). It did not succeed due to the Warsaw Pact invasion in August 1968, which ended Alexander Dubček’s period of “Prague Spring” (Pražské jaro). Dubček’s period of liberalized “socialism with a human face” had followed the Action Program of the Communist Party from April 1968, a plan that Czech sociologists and social scientists were heavily involved in developing (Skovajska & Balon, 2017; Vakakova & Strmiska, 1983).

Following the Warsaw Pact invasion, Dubček was expelled from the KSČ, and his replacement, Gustav Husák, reversed many of the liberal reforms. This period of “normalization” beginning in 1969 saw the purging of Dubček and reform supporters from government and academia, including all ČSAV institute directors. The Institute of Sociology
was closed in 1970 due to its status as a “hotbed of the revisionist epidemic” (Skovajska & Balon, 2017, p. 75), and the new Institute of Philosophy and Sociology was established in its place. As Illner described to re-vetting and purging of Institute staff over the next several years, “people were more targeted than institutions” during normalization (2002, p. 406; see also Vakakova & Strmiska, 1983 and the editor's note for Šiklová, 1982). In 1970, the KSČ established the official interpretation of the events surrounding the Prague Spring; among other factors leading to the “counter-revolution,” it identified

Serious problems and crisis phenomena in the ideological field, which have been emerging and growing in our society for a long time, resulted in a distortion, denial and later a sharp rejection of the basic principles of the Marxist-Leninist worldview in philosophy, sociology [emphasis added], in history, in economics, in the field of state, law and party building, and in culture and art, where there was a direct surrender before bourgeois ideology (Komunistická strana Československa, 1970/1978).

Despite the heavy post-Prague Spring backlash, sociology did not disappear entirely again (Vakakova & Strmiska, 1983); however, as it was not safe to retain the 1960s’ sociological research data under the normalization regime, many data sets were removed to the United States, the United Kingdom, and West Germany.112 As ČSDA02 described,

And many of these data, we don't have [in the archive]. It's from other sources. And there is a lot of data, which we don't have, located also the Federal Data Archive, of course… in GESIS, sorry, but also in the British archive, UKDA. ... Because it's mostly from old research projects, like from 1960s, and at that time it was impossible to, or even unreasonable to locate the data in Communist Czechoslovakia.

In 1973, students were once again admitted to university sociology programs, through the curricula were intensively focused on the communist ideology. The Czech Sociological Association as well lost its autonomy, but over the next 20 years, it grew in membership and in

---

112 For example, the Images of the World in the Year 2000; Czechoslovakia, 1967 (1976) data set available through the UKDA.
methodological expertise (Skovajska & Balon, 2017). In the years leading to the Velvet Revolution and the fall of the communist regime, the growing liberalization of younger researchers across the ČSAV and the Czech Sociological Association was very apparent and influential.

Sociology after the end of the communist regime changed fundamentally, as “all political restrictions and regulations on what can be investigated, how, by whom, and with what results [were] lifted” (Illner, 2002, p. 407). The Institute of Sociology (Sociologický ústav) (SOÚ) was re-established in 1990. On December 31, 1992, as part of the dissolution of the federalized Czechoslovakia into independent countries—the Czech Republic and Slovakia—the Academy of Sciences of the Czech Republic (Akademie věd České republiky) (AV ČR) replaced the ČSAV. The 1990s, along with rapid economic growth and reintegration into the capitalist and democratic Europe (it joined the North Atlantic Treaty Organization in 1999 and the European Union in 2004), saw a steady increase in students of sociology and the social sciences across the Czech Republic. Czech sociology was also increasingly incorporating previously inaccessible works of Bourdieu, Kuhn, and Giddens, among many others (Skovajska & Balon, 2017).

By the late 1990s, the Social Stratification research team at the SOÚ “had a lot of data. And there was also some long history of trials of establishing an archive in the Czech Republic. So, there was a great demand for that, but nobody succeeded” (ČSDA02). The Sociological Data Archive (Sociologický datovy archiv) (SDA) was thus established in 1998 as one of several SOÚ departments specifically to support the SOÚ researchers. The SOÚ policy for SOÚ researchers to archive their research data with the SDA (renamed the Czech Social Science Data Archive (Český sociálněvědní datový archiv) (ČSDA) in 2011) remains, and the SOÚ is still the primary data provider for the archive.
The SDA joined the then-Council of European Social Science Data Archives in 2001 and began to take part in international projects through that association. In 2010, the Ministry of Education, Youth and Sports of the Czech Republic (MŠMT ČR) included the SDA in its plans for large scale national research infrastructures, and began funding it to support those endeavors (Ministerstvo školství, mládeže a tělovýchovy České republiky, 2010). The MŠMT ČR specifically highlighted the on-going SDA and SOÚ collaborations with the European Social Survey (ESS), and SDA’s status as the national CESSDA node. When the “informal” Council of European Social Science Data Archive became a “legal” entity as the Consortium of European Social Science Data Archives (CESSDA) in 2013, ČSDA joined again. CESSDA became a European Research Infrastructure Consortium in 2017, and ČSDA remains the national node.

4.1.3 Privacy and the GDPR in the Czech Republic

The reaction to the GDPR in the Czech Republic was decidedly mixed, but even those mixed reactions tended to be negative. All ČSDA interviewees reported this negative trending reaction, which has also been reported in several post-GDPR studies. Panic and frustration were common responses. For example,

And at the beginning, even in social sciences, it was totally uninformed, we only knew that there is something like GDPR. It was, I remember, a strange abbreviation. What does it mean? There it will not be possible to do any of the surveys, because everything will be personal data (ČSDA01).

… there was a great panic because… out from nowhere many consultancy firms or businesses surged from the ground, ’et's say, threatening the enterprise and organization with fines. I mean, ‘use our services, otherwise you could be fined!’ (ČSDA04).

The frustration was tied to the uncertainty and panic over what the GDPR would mean, but there was also a frustration borne out of the negative views Czechs have come to have about the European Union. I will address this second source of frustration later in this section.
The negative views of the GDPR do not reflect a lack of concern over privacy in the Czech Republic. The right to privacy is guaranteed by the Charter of Fundamental Rights and Freedoms (Listina základních práv a svobod) (LZPS): “The inviolability of the person and of her privacy is guaranteed. They may be limited only in cases provided for by law” (Čl. 7(1)). The LZPS was first adopted by the Czechoslovak Federative Republic on January 9, 1991, and was adopted as part of the Czech Constitutional Code on December 16, 1992. It was developed in recognition (in part) of “the bitter experience of periods when human rights and fundamental freedoms were suppressed in our homeland” (Preamble, LZPS). In June 1992, the first data protection law in the post-communist Czechoslovakia was adopted (this would be replaced by the 2000 Data Protection Directive implementation law). While this law was seen as relatively weak, in part because there was no independent body overseeing its enforcement, it did establish precedent for what “abuses” would be identified in the later data protection laws (Banisar & Davies, 1999). In 2012, the Civil Code was replaced by the New Civil Code. The Civil Code, which had been in effect since 1964 during the Communist regime, only had a “general” clause about a right to privacy protection (Ondřejová, 2014). The New Civil Code, in contrast, grounds all legal interpretations in the rights listed in the LZPS, and reiterates a


117 § 11, Civil Code (1964): “Fyzická osoba má právo na ochranu své osobnosti, zejména života a zdraví, občanské cti a lidské důstojnosti, jakož i soukromí, svého jména a projevů osobní povahy.” (An individual shall have the right to protection of his or her personhood, in particular of his or her life and health, civic honour and human dignity as well as of its privacy, name and expressions of personal nature). Translation provided by Masaryk University.
person’s right to privacy several times, including in § 3(2)(a), which lists “everyone has the right to protect his life and health, as well as freedom, honour, dignity and privacy” as the first of six principles that private law relies on, and § 86, which declares that

No one may interfere in the privacy of another without a lawful reason. Without an individual’s consent, it shall in particular be prohibited to intrude into his private premises, watch or record his private life on audio or video recordings, use such or other recordings made by a third person about the private life of an individual, or distribute such recordings about his private life. Private documents of personal nature are protected to the same extent.

As Ondřejová reiterates, “demonstratively, the main principle is that everyone has the right to protect their life and health, as well as freedom, honour, dignity, and privacy” (Ondřejová, 2014, p. 55).

The “bad experience from communist times with its misuse of personal data” that ČSDA02 described not only made privacy a priority within the Czech Republic, it also affected how personal data were reused for research purposes before the GDPR. The 2000 Act on Personal Data Protection was interpreted to have no exemption for scientific research reuse of personal data by other researchers (according to ČSDA01, ČSDA02, and ČSDA04); the text of Art. 5(1)(e) states that data controllers shall

preserve personal data only for a period of time that is necessary for the purpose of their processing. After expiry of this period, personal data may be preserved only for purposes of the state statistical service, and for scientific and archival purposes. When using personal data for these purposes, it is necessary to respect the right to protection of private and personal life of the data subject from unauthorised interference and to make personal data anonymous as soon as possible).

Per the law, the data controllers could archive or further use the data, but the interpretation was that other users could not do so. ČSDA02 tied this lack of a research reuse exemption to the experiences under communism which made members of parliament “a bit nervous. So, if they
could delete some exceptions, they usually do that. So, GDPR makes it more safe from our nervous members of parliament.”

In the broader Czech society, before the GDPR, surveillance technologies, (e.g., CCTV cameras) were the most prominent privacy concern (Gawrecká, 2013). Gawrecká’s (2013) study was on how Czech media reported on the “security-privacy” debate in the Czech Republic, where the desire for increasing security in an era of rising international terrorism and organized crime created new risks to personal privacy and dignity. When the Eurobarometer 83.1 survey (European Commission, Brussels, 2018; see also Appendix F) asked how concerned respondents were about public surveillance, 37.31% of Czech respondents reported being fairly or very concerned. This was higher than the European Union (EU) average of 30.41%.118

In terms of feeling in control of their information online, the majority of Czech respondents felt they had only partial control (54.83%), while only 16.64% felt in complete control. At the same time, 77.38% felt concerned about this limited control, well above the EU average of 65.13%. When asked about whether explicit consent should be obtained before collecting or processing personal data, there were three possible “yes” answers, and multiple responses were possible: “yes, in all cases” (73.08%), “yes, in case of personal information required online” (13.70%), and “yes, in the case of sensitive information whether online or offline” (14.82%). The spread of “yes” responses varies on the EU level, but in all EU Member States, more than half said that consent should be required in all cases. The EU average for consent in all cases was 70.21%.

118 For claims that involve contrasting rates between one of the four countries of interest and the EU excluding the four countries of interest, the smallest sample sizes were 588 (single country) and 14406 (remainder of EU). In this worst-case scenario, the standard deviation of the difference of two percentages is no more than 2.1%, yielding a 95% margin of error of 4.2%. Thus, any difference between two percentages reported here that exceeds 4.2 percentage points is statistically significant.
Despite these concerns, when considering the importance of privacy compared to privacy-ensuring behaviors of people in the Czech Republic, the ČSDA interviewees had similar views:

I think when people are asked about it, they are quite sensitive about their privacy, but in their real behavior, probably it's the same everywhere in the world, in their real behavior, especially on the internet, they are not so much caring about privacy (ČSDA01).

I think majority of society does not care about issue of privacy at all. I think it is also question of education, social background, etc. (ČSDA03).

Privacy on a broader scale is perceived quite differently. According to ČSDA01, after the Velvet Revolution in 1989, there was a view that, in order to be free, the people had to be free from state surveillance. Even now, there is

very low, quite low trust in the state and in public administration. I am speculating, it's hard to tell what's behind? I think people are suspicious, especially in public sphere, suspicious that the state tries to know too much about them, to collect too much information (ČSDA01).

When the GDPR was adopted, there were two main public reactions: panic and frustration. The panic, as discussed above, was related to a lack of awareness and uncertainty about what the GDPR would mean. As late as May 24, 2018, the day before the GDPR entered into force, media reports were highlighting “the lack of information made available to the general public about what GDPR is and how it will affect them has resulted in scaremongering and rumours about alleged bans and excessively high fines” (Lazarová, 2018; see also Kramer, 2018). ČSDA01 took a similar view:

I remember that in newspapers, it was discussed, ‘what will happen? Will it be possible to do anything with the data after the GDPR will come into existence?’ And they were ‘lie’ debates, totally uninformed about what will happen. ‘Will it be possible for example, for schools to organize anything with parents? Would it be possible, any communication between school and parents
when there is no possibility to reveal any personal data?’ And, so on, it was quite hot, quite vivid, even among lay people in newspapers. The panic turned into frustration as time wore on because it remained unclear what would happen.

There was a second source of frustration, however, tied not to the GDPR as a legal obligation itself but its status as an instrument of the European Union. The ČSDA interviewees described this reaction to the GDPR as follows:

Something wrong comes from the European Union again. Because, from both sides, I think that some people were afraid that now there is more intervention into their privacy. … So, there were discussions which were against... very much against GDPR and very much against Europe… in connection also to attitudes against European Union. Simply too much work. Too much money (ČSDA02).

… people hate GDPR. And they find it as annoying, not important. Something from the EU we don't want to have (ČSDA04).

These views reflect the dramatic shift in Czech public opinion about the EU since the early 2010s. After the Czech Republic joined the EU in 2004, the views of the EU were typically high, especially during the six month period (January to June) in 2009 when the Czech Republic first held the presidency of the Council of the European Union (Hořejš et al., 2019). However, as reported in the Eurobarometer 91.1 (2019) survey addressing perspectives on the EU, only 33% of Czech respondents in the spring of 2019 considered Czech membership in the EU “a good thing.” This was the lowest of all surveyed countries; of note, it was lower than the United Kingdom’s 43% less than a year before Brexit took effect. Despite this, only 24% of Czech respondents would vote to leave the EU.

Hořejš et al. found that the main negative association to the EU was “the feeling that the EU devotes its time to banal topics, such as bans and regulations that intrude on citizens’ everyday life” (Hořejš et al., 2019, p. 13), while Cvik et al. (2018) highlighted the EU’s lack of
transparency, clarity, and consideration of the realities of data processing institutions. Both Hořejš et al. (2019) and Chlup (2020) considered the historical experiences of outsiders (particularly Germans and Soviets) imposing their views and policies on the Czech people as the likely foundation for this view. Therefore, any regulation is almost automatically considered too bureaucratic and burdensome, no matter the intended outcome (e.g., Kříž, 2017; ČSDA01; ČSDA04). For example, while Tahal and Formánek’s survey of Czech internet users six months after the GDPR went into effect did find that their respondents considered the GDPR’s measures themselves to be useful, they approached the GDPR itself “with certain amount of reservation, given the bothersome aspects involved with the procedure of providing data sharing consent” (Tahal & Formánek, 2020, p. 91). This “consent” burden includes the need to constantly accept cookies when using the internet, as well as the need to sign GDPR-related privacy policy forms seemingly everywhere, even when the GDPR did not apply. Kadlecová and Weiss found that the public Czech discussions of the GDPR focused on the bureaucratic burdens and not on the “need for privacy as an integral part of human dignity” that the GDPR intended (Kadlecová & Weiss, 2021, p. 8). ČSDA04 similarly gave the anecdotal story of being asked to read the GDPR privacy policy at their child’s nursery school while also being told that they did not really need to read it.

There is a closely related view that the problem with the GDPR is that it is too general, but also too focused on what ČSDA03 referred to as the “big data controllers”–the banks, IT companies, etc.–for whom the

duties, administratives and responsibility imposed on them by GDPR are not a major issue, as they have a lot of staff (experts) and departments that can deal with them for long time. But if you impose the same strict principles and rules on e.g. preschools, municipal authorities or small companies, these entities will often get into special (more or less unusual) situations.
For everyone else, according to ČSDA04, “those rules are pretty much useless or, they're too vague. For me, it's not very useful, those rules, and those poor organizations and institutes have to fit with those GDPR boxes and categories and those privacy policies.”

A frequent pre-GDPR complaint from the then-European Commissioner for Justice, Consumers and Gender Equality, Věra Jourová, was that Czechs were not interested in and did not care about their personal data (this view was shared in multiple interviews during this period, including: Hendrych, 2017; Lazarová, 2018). However, the Special Eurobarometer 487a (2021), conducted in March 2019 (see Appendix G), found that 49.53% of Czech respondents had both heard of the GDPR and knew what it meant, at a much higher rate than the EU average (35.85%). Overall, 84.46% of Czech respondents at least knew of it, whether they knew the details specifically, compared to the EU average of 67.09%. It also found that more than half of Czech respondents knew about the rights of access (68.73%), objection to marketing (54.58%), rectification (61.42%), to be forgotten (56.65%), while less than half knew about the right of portability (45.59%) or to not be subject to automated decision making (37.73%). Of these rights, by March 2019, only two had already been exercised by more than 10% of respondents: access (11.24%) and objection to marketing (11.99%). The Special Eurobarometer 487a also found that the percentage of Czech respondents feeling partially or completely in control of their online information had decreased from the 2015 Eurobarometer 83.1, from 54.23% to 51.56% (partially) and 16.64% to 16.39% (completely). However, there was a larger decrease in concern about the limited control, from 77.38% in 2015 to 55.55% in 2019 (European Commission, Brussels, 2018, 2021).
4.1.4 Timeline of GDPR response

After the GDPR was adopted by the European Parliament in 2016, ČSDA began reviewing its operations and processes in early 2017 to identify potential areas where changes were needed. Also in 2017, ČSDA staff began collaborating with fellow SOÚ department, the Center for Public Opinion Research, to advise the wider SOÚ on the GDPR (ČSDA02). ČSDA also contacted the ÚOOÚ with questions during this period, because there was no data protection officer (DPO) yet at either the AV ČR or the SOÚ. ČSDA also collaborated with SIMAR (Sdružení agentur pro výzkum trhu a veřejného mínění), the association for market and social research, and used templates from SIMAR in its evaluation of ČSDA’s operations.

Just before the GDPR entered into force in 2018, the AV ČR hired a DPO. According to ČSDA03, this position was created entirely because of the GDPR; there was no such requirement for a DPO under the 2000 Personal Data Processing Act. On March 12, 2019, the Czech GDPR implementation law was passed, and it entered into force on April 24, 2019. Also in 2019, the AV ČR published its GDPR-compliant privacy policy (Akademie věd České republiky, 2020). ČSDA then hosted a Czech and English language workshop on data management and data protection after the GDPR, sponsored and supported by CESSDA. In 2020, ČSDA changed its Nesstar user database in response to the hack of GESIS’ account registration database (Sections 4.1.5.6 and 4.4.6). In 2021, ČSDA began transitioning its archiving platform to a Dataverse-based platform. In early 2022, the SOÚ published its own privacy policy (Sociologický ústav Akademie věd ČR, v.v.i., 2022a) and hired its own DPO, both of which directly apply to ČSDA.

4.1.5 Priorities in the response

In the interviews, the ČSDA respondents reiterated that, because their research data are considered anonymous, before they began reviewing their operations and data they anticipated...
that few changes would be required. However, despite this, they were concerned with making sure that ČSDA was compliant, especially regarding the already-archived research data. ČSDA01 also identified greater transparency with data subjects (including research participants) as a priority.

4.1.5.1 Priorities and goals development

While the research data archived at ČSDA are considered anonymous and thus outside the scope of the GDPR, ČSDA did not ignore the GDPR or the 2019 Personal Data Processing Act (ZZOÚ). ČSDA staff began reviewing operations and policies in 2017 to identify potential areas where policy and practice changes were required, and ČSDA was largely independent in this process. According to ČSDA03, the AV ČR decided,

with regard to the wide range of research activities of research institutes (54 institutes), that each institute should have deal with the GDPR on its own. [The AV ČR DPO] should only have an advisory role for the institutes (on demand).

Within its parent institute, the SOÚ, ČSDA also found itself in an advisory role to the SOÚ in developing institute-wide policies—along with fellow SOÚ department, the Center for Public Opinion Research (Centrum pro výzkum veřejného mínění)—such as the GDPR-compliant “Ochrana osobních údajů” (hereafter, “SOÚ Privacy Policy”) (Sociologický ústav Akademie věd ČR, v.v.i., 2022a). As well, the “ČSDA Expert” in data protection (discussed further in Section 4.1.6), one of the research data-focused staff, was the primary ČSDA staff member planning ČSDA-specific work, advising the ČSDA director, and advising the wider SOÚ. The AV ČR’s DPO had not yet been appointed; therefore, ČSDA had no other data protection expertise in the AV ČR to turn to. The priorities ČSDA identified were related to the deposit review process, informed consent, addressing already-archived data, and data protection efforts.
4.1.5.2 Processor versus controller

Whether any organization needs to adopt the processor or controller role in a given context depends on whether it processes personal data in accordance with Article 4(2) GDPR (See Section 2.1). ČSDA is an interesting case in this dissertation in this capacity. As described in Section 4.1.1.1, ČSDA considers the data made available through Nesstar to be anonymous, and thus, not regulated by the GDPR. Even though ČSDA does have data from SOÚ researchers which contain personal data, these data are stored in a separate physical safe and are only available to SOÚ researchers in a secure area within the SOÚ library (ČSDA02). There is no online access to these data, and they are not available outside of the SOÚ. Therefore, for the research data ČSDA archives and makes available through the Nesstar platform for reuse, ČSDA itself has not adopted any role for these data because they are considered anonymous.

Instead, the SOÚ is the controller for the SOÚ-produced research data which do contain personal data, and any processing that ČSDA does in its support role on these data (anonymization, storage, dissemination to SOÚ researchers), it does as a “controller” as part of the SOÚ. ČSDA itself made no changes related to these data after the GDPR was adopted, though as discussed in Section 4.1.5.4, the SOÚ as a whole has implemented requirement for proof of consent for archiving for personal data collected by SOÚ researchers and physically stored in the department. The “Conditions of Access” (Český sociálněvědní datový archiv, 2022b) also clarify to data users that ČSDA is not the owner of any data it makes available, and is not responsible for their accuracy. My ČSDA interviewees did acknowledge that this may be a challenge if insufficiently-anonymized data are deposited and data curation begins. According to ČSDA01

When we receive from them in some first version a data set with personal information, we seek, we communicate with the depositor, and we give the
depositor advice, how to anonymize, but in most cases, we ’on't want to do anonymization because it means that you are working with personal data and we ’on't want to work with personal data, especially when they are from the outside of the Institute.

The situation is the same when it comes to the administrative and operational data, and the data from the ČSDA-involved projects (Section 4.1.1.2). The SOÚ is the controller and the processor for all administrative and operational data across the Institute, and therefore, the ČSDA is in that way the controller and processor on its departmental administrative and operational data. This role is stated in the “SOÚ Privacy Policy” (Sociologický ústav Akademie věd ČR, v.v.i., 2022a), which identifies the SOÚ as the controller (správce) and processor (zpracovatel).119 These roles are the responsibility of the SOÚ and not the AV ČR because, as discussed in Section 4.1.1.3, each AV ČR Institute is its own, independent legal entity, and “from the GDPR view, they are separate controllers” (ČSDA03). For the projects that ČSDA staff are involved in in partnership with CESSDA or through the SOÚ (such as the ESS or the ISSP), whether ČSDA is the controller or processor (through the roles the SOÚ adopts) depends on the projects: “we have such contacts regarding our own research activities and our other work. So, in that other work, we are boss, processors, and controllers as well, it depends on the job” (ČSDA02).

4.1.5.3 Regulating information flows

Even though ČSDA has not assumed the role of processor or controller for the research data it makes available through Nesstar, it did implement changes affecting the flow of

119 As will be discussed in Section 4.1.5.4, there is a discrepancy between the Czech and English language versions of the “SOÚ Privacy Policy.”. The Czech version says the following: “Správce a zpracovatel osobních údajů” (Controller and processor of personal data), using the Czech terms from the official Czech language translation of the GDPR (https://eur-lex.europa.eu/legal-content/CS/TXT/HTML/?uri=CELEX:32016R0679&from=CS#d1e1366-1. Článek 4(7) and 4(8)). The English version says the following: “Data controller.” As with all my analysis, and in accordance with translations provided in other (but not all) archive documentation with multiple linguistic versions, I am accepting the original language versions of all policies as “official.”
information. These changes specifically focus on data coming into the data archive. The “Archivační řád” (Archiving Rules) (2016) themselves, which direct the data acquisition and access provision processes from ČSDA’s perspective, were not revised in response to the GDPR. This was because

our acquisition policy, which we set formerly, not because of GDPR, but because we became a member of CESSDA infrastructure, was specified enough and strict enough. Our archiving principles, archiving guide was strict enough to be able to cope with GDPR legislation (ČSDA01).

However, the deposit review process that begins once ČSDA receives a data set has become stricter. This was not a formal, documented change, but a change in archive staff attitudes towards the deposit review process:

… there was no need to change formally anything, but maybe informally, we are more strict, and we are taking more care in dealing with data, especially at the beginning of the process when checking the data. Or before publishing the data, checking the data before publishing, again, checking if there is anything which can be clashing with personal data protection, so probably it's more strict, but it's not written anywhere in formal rules to be more strict, but probably it's only a consequence of the GDPR discussions (ČSDA01).

Not only are staff stricter and more precise about checking for personal data, ČSDA02 argues that the expectations are clearer now than before the GDPR:

… the other thing is that we now better know what to do, thanks to GDPR. While, because of lack of description under the former Czech law, we were not sure if we are misbehaving or not when handling data which were not perfectly anonymous.

The legal clarifications thus support the increased awareness and focus that ČSDA01 described.

The presence of personal data is not the only thing that ČSDA looks for in greater detail now than before the GDPR. ČSDA looks for whether, if there are personal data that must be archived (such as in the case of SOÚ-produced research), the data subjects have consented, either to the research or to the archiving. The data deposit will be rejected if “they don't have consent,
or they don't have any other reason why they collected their data, or they can't provide any
evidence about it” (ČSDA02). ČSDA01 thinks this is more of a concern for the qualitative data
than the quantitative data. If issues do arise during the deposit review, ČSDA staff will stop the
process and try to resolve them with the data provider and/or the head of the archive.

4.1.5.4 Minimizing harm

ČSDA, or the SOÚ on ČSDA’s behalf, made several changes intended to minimize
potential harms to data subjects. These changes regard informed consent, data already in the
data archive, and processing ČSDA’s administrative and operational data. Changes to informed
consent practices apply to both the data in the data archive and to the projects that ČSDA takes
part in. The first is related to the change discussed in the previous section (4.1.5.3). When
ČSDA looks for proof of consent during the deposit review, it also looks at whether the consent
of the data subjects would be “sufficient” (ČSDA01). The GDPR is stricter in this regard than
the previous Czech Personal Data Processing Act (2000). The pre-GDPR Czech law did require
the controller to inform data subjects about any processing or dissemination of their personal
data (§ 11(1)). However, the required information shared with data subjects is more extensive
under the GDPR (Art. 13), and “must be given separately from the rest of the document, it must
be given separately for each purpose of processing and the type of personal data” (ČSDA03).
That includes that their data will be shared with an archive and reflects the importance of
transparency and data subjects’ rights to their data. The SOÚ now enforces this requirement for
research that SOÚ researchers conduct, and ČSDA staff now also assist those researchers with
preparing their “informed consent or other document and information in relation to data
protection during their research work” (ČSDA02). This transparency with data subjects is also a
consideration with the external projects ČSDA collaborates on. Reiterating the importance of
data subjects’ awareness of their rights and what will happen to their data, ČSDA01 describes the requirements for informed consent thus:

… when you are collecting the data, you have to be more precise when you are asking for consent from the respondents, for example. So, yeah, it's expected that the consent will be really understandable for the person, for the respondent, and so on. That you will have to be clear, when you explain the way… how we will deal with the personal data.

The consent forms provided for these projects were revised to be GDPR-compliant and more transparent to potential survey respondents.

The next change ČSDA made was to data sets already in the data archive. During the review of the data archive, ČSDA identified data sets which contained potential indirect identifiers. Before the GDPR, “it was considered that it’s okay, that it’s anonymous. After GDPR, we were not so confident ourselves regarding this” (ČSDA02). When they found a problematic data set, ČSDA staff further anonymized the data. If they determined that sufficient, GDPR-compliant anonymization was impossible, they removed the data set from the archive entirely, prioritizing the potential privacy risk over reuse utility.

The final change in this section address the administrative and operational data; specifically, the implementation of a GDPR-compliant privacy policy, though the policy itself is somewhat complicated. This policy, the “SOÚ Privacy Policy,” was developed by the SOÚ with input from ČSDA staff, and applies across the entire Institute (Sociologický ústav Akademie věd ČR, v.v.i., 2022a). The AV ČR had already implemented its own privacy policy in 2019, the “Informace o zpracováni Vašich osobních údajů” (Akademie věd České republiky, 2020). The “SOÚ Privacy Policy” mirrors a significant portion of the AV ČR policy and (as discussed in Section 4.1.5.2) identifies the SOÚ as both the controller and processor, while the AV ČR only claims the role of controller. The “SOÚ Privacy Policy” (2022a) informs data subjects of the
purposes and legal bases for personal data processing, personal data retention periods, and the data subject’s rights under the GDPR (specifically, the rights of access, rectification, erasure, objection, and to revoke their consent).

This policy was only implemented in early 2022, four years after the GDPR came into effect, and is entirely new. The policy is also only directly accessible through the Czech language SOÚ web page. There is an English language version of the policy (Sociologický ústav Akademie věd ČR, v.v.i., 2022b); however, there are no links to the policy from any English-language SOÚ web page. There are some textual differences between the English and Czech versions of the “SOÚ Privacy Policy,” but the significant change is that the English version only identifies the SOÚ as the controller (Sociologický ústav Akademie věd ČR, v.v.i., 2022b). In line with other documentation with both original language and English versions available, I accepted the Czech version of the “SOÚ Privacy Policy” as official and accurate, and confirmed its accuracy with my interviewees. Of note, the “SOÚ Privacy Policy” is also not linked (in either language) through the ČSDA web page, or any other SOÚ department web page, so how widely-known this policy is is unclear.

4.1.5.5 Response to technological risks

The changes at ČSDA that reflect concerns over technology risks to privacy focused on administrative and operational data protection efforts. The first change for these data was a rule change that the SOÚ made for working with the data. For example,

\[\text{if you handle personal data, you have to do a lot of records... We have that safe place. So, if you don't use the data, you have to put them to the safe place. There are more tough rules regarding where you can keep the data, on which equipment, which computers, notebooks, and how to handle with these equipments, what to do with the notebook. You have new rules also regarding}\]

---

120 The English version of the “SOÚ Privacy Policy” can be accessed by adding “en” to the end of the link to the Czech version: https://gdpr.soc.cas.cz/en/.
passwords, for example. And I think also there are many very small things which are different in comparison with the past (ČSDA02).

Among these rules are also restrictions on using services located outside of the Czech Republic. For example, ČSDA staff use an SOÚ-owned and -operated cloud server for any work with personal data and cannot “use external services without any correct contract” (ČSDA02).

The second change affecting ČSDA’s administrative and operational data is one that ČSDA itself implemented. Until 2020, ČSDA’s data access catalog Nesstar also contained the ČSDA user account database. Initially, after reviewing their system operations, ČSDA staff determined that the database was secure and their processes for it were GDPR compliant.

However, in December 2019, GESIS’ account registration database was hacked (see Section 4.4.6 for more details). This event did not have any consequences for ČSDA, as there are no connections between the two archiving systems. Even so, ČSDA staff decided to err on the side of caution, and “in collaboration with [the SOÚ] IT department took some measures to make our database of users more secure” (ČSDA01), including removing the database from the Nesstar platform. As a consequence, ČSDA users now have two separate logins, one based on their registered email address to log in to their ČSDA account, and one specifically for analyzing data in Nesstar, which the user only provides when they begin selecting variables for analysis.

For the final change, ČSDA created GDPR-specific data protection training, not just for ČSDA staff but for the entire SOÚ, and took part in several GDPR training events aimed at the wider AV ČR and social science research community (e.g., Krejčí, 2019a, 2019b). These trainings introduced important data protection and privacy concepts and requirements of the GDPR and the 2019 ZZOÚ for social science research and what the new policies and practices at the SOÚ were.
4.1.5.6 Other changes

There was one additional change that affected ČSDA specifically because of the GDPR. As with creating the “SOÚ Privacy Policy,” the SOÚ was ultimately responsible for establishing and filling the position of data protection officer (DPO). However, this was a complicated process, and an external DPO firm was only contracted in early 2022. This is because of a several years-long debate between the AV ČR and its institutes about whether each individual institute needed its own DPO, a debate that is still on-going. As ČSDA04 described in 2021,

It is not in general agreement, whether the institutes of the Academy of Sciences as individual legal entities, do need a DPO, Data Protection Officer. So, there was a great discussion, and I don't think that the discussion has been resolved, or the problem. So, some institutes do have a data protection officer and some not.

Before the SOÚ hired a DPO, according to ČSDA02, because ČSDA (and the SOÚ) is “part of the structure of the Academy of Science, and there is a data officer. So, we [could] contact data officer at the Academy.” However, AV ČR’s DPO was only appointed a month prior to the GDPR entering into force; there was no AV ČR advisor available who could specifically address GDPR-related concerns prior to that date.

4.1.5.7 Changes remaining to be completed

ČSDA02 considers that ČSDA is 90-95% compliant with the GDPR: “concerning the data, it’s okay. But, I think that we are not so punctual concerning organization of events and regular operations.” There are several reasons for that perspective which will be discussed in Section 4.1.6. Accordingly, there are several changes that the ČSDA interviewees feel remain. These are: policies for “atypical” data; transitioning to the new archiving system; and, revising data protection training. This first issue, the atypical data, are the qualitative data in the Medard archive. ČSDA is predominantly a quantitative archive, and the deposit review processes are
well established for these data. However, qualitative data are trickier and present different anonymization challenges, and according to ČSDA01,

… we need some better specification of archival processes in relation to qualitative data or non-anonymous data. And not only in the relation of GDPR. But the GDPR is part of it, because the fact that the big part of this qualitative data has some personal information, this is a big part of the problematic status in our archive. So, qualitative data, I think is one field or one issue we will get probably in the future, to improve our qualitative archive, or qualitative part of our deposits.

As discussed in Section 4.1.4, ČSDA began transitioning from its Nesstar-based archiving and analysis platform to a Dataverse-based archive in 2021. This process is still on-going, and the Nesstar platform will remain active for a while after the Dataverse archive becomes operations for users who are more familiar with the Nesstar analysis tools (ČSDA01). The ČSDA interviewees were not sure whether the current access categories would remain once the new platform was operational; that is an on-going discussion as well. The final change that remains is to further update the data protection training. During the GDPR transition period, ČSDA staff were seen very early on as the data protection experts within the SOÚ, and regularly provided GDPR-related training to SOÚ researchers and the wider social science community in partnership with CESSDA. However, further changes to the SOÚ-focused training have been delayed because SOÚ policies were still under review and not yet finalized; “based on that, we will also provide more schooling to the researchers in the Institute, different kinds of forms for use during their research, those kinds of things” (ČSDA02).

There were two additional unresolved issues the ČSDA interviewees identified during their fall 2021 interviews: the lack of a GDPR-compliant SOÚ privacy policy, and the lack of a dedicated data protection expert. As discussed in Section 4.1.5.4 and 4.1.5.6, respectively, both issues were resolved as of early 2022. However, the delays in these changes at the SOÚ level
have also contributed, for example, to the delay in further revising the data protection training materials.

4.1.6 Challenges at ČSDA

Despite considering its accessible research data to be anonymous, ČSDA faced several challenges during the GDPR transition period. These challenges can be classified into three categories: challenges related to the data protection laws; challenges related to ČSDA’s organizational structure; and, challenges for ČSDA’s operations. There was a lot of confusion at ČSDA, within the SOÚ more broadly, and within the Czech social science research community about just what the GDPR would mean, and which requirements would apply to whom. ČSDA01, ČSDA02, and ČSDA04 connect this confusion to the long delay before the new Czech Act on Personal Data Processing (ZZOÚ) was passed in 2019. For example,

… there were a lot of discussions about what will happen, and if it will be total catastrophe, or how to prepare. And we were waiting for the Czech law, new Czech law, because there was... the European law was passed, but there was expectation that Czech legislation will take the European law, but add some Czech-specific rules to it, which are in accord with European legislation … And the process in government, and then in Parliament was very, very slow. And it was clear that Czech legislation will not be prepared before the time when the European law, the GDPR, came into power in Czech Republic (ČSDA01).

There was a great discussion the time of implementation of the new law, everybody were afraid, and the major problems from perspective of research I think is continuity from the times before that, and operation with old data, because sometimes you can find some problems which is not easy to deal with them under the new law. And other problems are of course, connected to costs. So, it's a lot of work now. … Maybe one more thing that the GDPR was implemented in 2018, at the date where it should be, but the specific Czech law came later, much later, I think, 2019. So that time of provisional situation was quite long (ČSDA02).

The cause of these delays stem from the Ministry of the Interior Czech Republic (Ministerstvo vnitra České republiky) being fully responsible for drafting the implementation law without
sufficient capacity to do so within the two year time period after the GDPR’s adoption, as well as the 2017 parliamentary election and the delays in establishing a stable government (Kadlecová & Weiss, 2021; Řeháková, 2018).

One of the tasks ČSDA undertook regarding its archived data (discussed in Section 4.1.5.4) found ČSDA staff further anonymizing or outright removing several already-archived data sets due to concerns about whether they were or could be made “more” GDPR compliant, because of concerns and interpretations that “personal data” are “everything,” according to ČSDA02, and what “anonymization” meant in practice was still unclear. As well, the 2019 ZZOÚ does allow personal data processing for “a task carried out in the public interest” (§ 5(b)). However, how this “public interest” exemption, which is new to the Czech data protection legal regime, applies to scientific reuse of personal data is still unclear. ČSDA01 identified the likely situation where, for example, a qualitative data set was collected with consent, but the consent did not mention future data archiving, as one scenario where relying on the legal basis of “public interest” could be useful to allow for archiving without that consent. However, … nobody knows how broad it is, how to understand the public interest for this scientific exemption, and even the lawyers are hesitating. … And maybe we could use this exemption for scientific use in public interest of promoting the science, but we are not sure if we can use it (ČSDA01).

The second set of challenges stem from ČSDA’s position as a dependent department of the SOÚ. According to ČSDA02, the SOÚ “was very slow in preparation for the GDPR. So, we had to develop a kind of pressure to move things forward.” While ČSDA was relatively free to determine what it specifically needed to do as a department, there were GDPR requirements it needed to meet that were the SOÚ’s responsibility as the legal entity. That was challenging as well because the SOÚ did not have dedicated data protection expertise outside of ČSDA. One ČSDA staff member (hereafter, “ČSDA Expert”) was considered the data protection expert for
the entire SOÚ, due to their experience addressing personal data privacy concerns with the research data from the SOÚ, and became the point of contact for the entire SOÚ. As ČSDA01 described,

… we didn't have this person from the top level of Academy of Sciences who will be responsible for the data protection and will be able to help us to solve our potential problems with personal data. So, for example, when somebody in the institute plan to collect some data, some sensitive data, some personal data, to do something that he or she expects to be problematic from the point of view of personal data protection, he is communicating with [ČSDA Expert], asking [them] .... They don't want to be this kind of expert, to be responsible person, in fact that they are not responsible. It's only partly formal, partly informal responsibility.

While the SOÚ now has the dedicated expertise, the newly contracted data protection officer, who is a lawyer, only took on this role in 2022, and thus was not available for consult during the GDPR transition period. In addition to ČSDA Expert’s increased burden addressing the entire SOÚ’s data protection concerns, ČSDA staff found efforts to become GDPR compliant inefficient and costly in terms of effort required on top of their day-to-day responsibilities (ČSDA02).

4.1.7 Impacts on ČSDA operations

While ČSDA’s formal changes in response to the GDPR were not extensive, the GDPR itself did have an impact on the archive. In this section, I will discuss whether there were measurable trends in ČSDA usage that may be connected to its GDPR adaptation. From 2011 to 2020, the number of data sets in data archive has grown relatively steadily; according to ČSDA01, ČSDA averages approximately forty to fifty new data sets per year, “influenced by our stronger emphasis on communicating with researchers from other branches of social science” (Fig. 4.2). From 2016 to 2020, the number of data sets published per year did fluctuate, with the lowest number of published data sets since 2014 published in 2018, and the highest in 2019 (Fig. 4.2)
Despite this outlier, ČSDA interviewees did not think that the GDPR had a significant effect on data deposits themselves, in part because of the SOÚ data archiving policy and the contracts that ČSDA has with external data providers (Section 4.1.1). However, in recent years ČSDA staff have begun more active data acquisition efforts, and if the GDPR did have an effect, the interviewees hypothesized that it would have been on individual, independent researchers’ willingness to share their data with an archive:

… the number is quite stable, but maybe, I'm not sure, some potential depositors can be more hesitating about depositing data because of GDPR. But because we are asking new potential depositors, or informing them about our existence and possibility to deposit data in the archive, so at the end, the total number is quite stable (ČSDA01).

That's also because I have already said that we have to ask for the data actively, and search for the data actively. So, the changes are based on us, very much. There are a lot of factors and I think that GDPR was not the major factor in this (ČSDA02).

However, whether this was indeed the case has not been investigated by ČSDA or this dissertation, and cannot be conclusively confirmed.

ČSDA does not report archive usage statistics, so it is not possible to report what, if any, changes may have occurred for use of the data in the archive (either in terms of downloads of data sets or use of the Nesstar analysis platform). According to ČSDA01,

in relation to people in academic sphere, some researchers... I don't think so, that changed much their attitude to data. Our data are anonymous, or at least we present them as anonymous, and I’m sure that the large majority of the data sets really are anonymous. There can be some problematic points in some data sets, maybe. So yeah. If we would have a larger collection of qualitative data, so maybe there will be more problems.

When ČSDA staff meet with researchers, they do often have questions about personal data protection, but the interviewees do not think that the GDPR has had any effect, positive or negative, on ČSDA data usage either. As with what ČSDA01 said above regarding increased data deposits due to the expansion of disciplines that ČSDA reached out to, ČSDA02 similarly attributes the annual growth of registered ČSDA users (Fig. 4.4) to the fact that “we made more contacts, more channels for communication and more contacts with a different institution based on the discussions about GDPR and our workshops and such.” The totals counted here reflect
the number of unique registered users at the end of each calendar year; several (but not all) of the
SOÚ’s Annual Reports these totals were aggregated from (Section 3.3.2.2) report the end-of-year
totals as well as the counts of unique “new” registered users and the number of users who deleted
their accounts.

![Unique Registered Users (ČSDA) 2011-2021](image)

*Figure 4.4 The growth in number of registered ČSDA users (2011-2020) (Sociologický ústav Akademie věd České

While the measurable operational effects of the GDPR are minimal, a change in mindset
at all levels of the organization was reported. As discussed in Section 4.1.5.3, ČSDA01
describes this effect as being “more precise” during the deposit review. Similarly,

Nowadays, documents to every meeting, training as well as contracts or
agreements are discussed automatically from GDPR point of view (ČSDA03).

The responsible persons began to be aware about personal data. And the
regulation... I mean not a regulation as a European regulation, but more the
rules on personal data, because before that, I think they were aware of that, but
they’re not lawyers, so it was not their fault, but the whole society was not
quite aware of personal data processing issues and problems (ČSDA04).

ČSDA02 also reported a change in how ČSDA is viewed by the SOÚ as well as the wider Czech
social science research community: “I think that we profited from the new topic, because now we
are more important with this expertise.”
4.2 Finnish Social Science Data Archive (FSD)

4.2.1 Archive description

FSD operates within Tampere University, in Tampere, Finland. It is funded by the Finnish Ministry of Education and Culture, with additional, project-specific funding from the Finnish government and the European Union. It has been in operation since 1999, and supports researchers, teachers, and students across Finland, as well as multinational researchers whose data or research are relevant to the Finnish context. There are currently twenty-nine employees of the archive, as well as temporary and student employees. Its CoreTrustSeal certification was renewed from 2020 through 2022 (CoreTrustSeal, 2020).

When FSD was established, it was a component of the University of Tampere. On January 1, 2019, the University of Tampere and the Tampere University of Technology merged into the current Tampere University. FSD is a separate institute within Tampere University, though the FSD servers are backed up with the Tampere University ICT data center, and the archived research data are also backed up by the Ministry of Education and Culture (opetus- ja kulttuuriministeriö) national long-term digital preservation service.

FSD is a CESSDA member and is the Finnish CESSDA node. In addition to being the CESSDA node, FSD also participates in several CESSDA infrastructure development projects, including the Social Science & Humanities Open Cloud (SSHOC) project121 and the TRIPLE Project122 (Consortium of European Social Science Data Archives, 2020). FSD does not conduct

---

121 The SSHOC project is a EU project supporting the development of the social sciences and humanities component of the European Open Science Cloud (EOSC), the EU’s federated open access research data support platform. FSD, GESIS, and DANS staff, along with 44 other research institutions and infrastructures across the EU collaborated on the integrated, networked data infrastructures supporting the social sciences and humanities. See https://sshopencloud.eu and https://eosc-portal.eu for more information on both projects.

122 The Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration (TRIPLE) Project aims to develop a multilingual and multicultural social sciences and humanities research and data discovery platform. As with the SSHOC project, the GoTriple platform supports the EOSC as a single access point for research resources for the social sciences and humanities research community. See https://project.gotriple.eu/ for more information on the TRIPLE project.
research, alone or in collaboration with other research institutes or data archives, other than its work on these infrastructure projects (FSD01). Therefore, FSD’s services are focused entirely on the research data that it archives and disseminates through the Aila Data Service.

4.2.1.1 Aila Data Service

While FSD has operated since 1999, the online data portal, the Aila Data Service (hereafter, “Aila”), has only been in use since 2014. Prior to May 2014, data users\textsuperscript{123} could submit a request for multiple data sets through a single data access request. Since Aila came online, each individual data set is requested separately; for this reason, the data use statistics before 2015 cannot be directly compared to those statistics from 1999-2014. Another 2014 change to the FSD data catalogue (the access point to the Aila portal) was the removal of metadata records for data sets archived outside of FSD.

As of the end of 2021, FSD’s archive holds 1,686 data sets, of which 282 are qualitative data sets and 1,403 are quantitative. By access category, 107 data sets are available without Aila registration under a CC BY 4.0 license (Category A), 1,248 are available for research, teaching and study (requiring Aila registration) (Category B), 253 are available for research only (with Aila registration, and for Master’s, PhD/doctoral, and Polytechnic/University of Applied Sciences Master’s theses) (Category C), and 78 are only available with the permission of the data depositor or creator (with Aila registration) (Category D)\textsuperscript{124} (Yhteiskuntatieteenlinen tietoarkisto, n.d.). Data in Finnish, Swedish, and English can be deposited at the archive, and Finnish data can be translated to English upon request. FSD does not have self-archiving capabilities.

\textsuperscript{123} FSD interviewees referred to data providers and data users collectively as “customers” during the interviews. For consistency with the rest of the dissertation, I will collectively refer to “users” unless specifically quoting from the interviews.

\textsuperscript{124} Data set counts based on the Finnish FSD Aila Data Catalogue, as reported on January 1, 2022. These numbers are updated as data sets are processed and published in the catalogue in real time. The Data Catalogue can be found at https://services.fsd.tuni.fi/catalogue/index?lang=en&study_language=fi.
because FSD’s “no non-anonymized data” policy requires that all deposited data must be checked to validate the anonymization work (FSD02). Self-archiving would increase the workload beyond FSD’s current resources.

The Aila data catalogue is available in Finnish and English; there is a discrepancy in number of data sets listed between the two catalogues, as not all data sets are available in both languages. Data in Aila cover various humanities and social sciences disciplines and topics, including education, crime and law enforcement, history, economics, social change, medical and health sciences, and many more. The Finnish components of the European Social Survey (ESS), the International Social Survey Programme (ISSP), and national surveys such as the EVA Surveys on Finnish Values and Attitudes\textsuperscript{125} and the Sociobarometers\textsuperscript{126} are also archived at FSD.

While FSD focuses primarily on the social sciences, the data reflect a broad spectrum of subject areas. The archive ingests and disseminates both qualitative and quantitative data. FSD does not have, and has never had, the infrastructure to make identifiable data available to data users through various restricted data access methods. The reason for this is because the resources to provide some form of restricted access to non-anonymized data were not available when FSD opened (FSD02). To establish some sort of restricted data access infrastructure now would require even more resources, so this is not a priority for FSD. It also only accessions data which the data providers agree to make available for reuse.

As discussed above, FSD only actively archives anonymized data, and the deposit review process includes verifying compliance with FSD’s anonymization guidelines (Finnish Social

\textsuperscript{125} The Surveys on Finnish Values and Attitudes are biennial surveys conducted since 1984 by the Finnish Business and Policy Forum (EVA) which study changes in Finnish citizen views of Finnish society. Further information can be found at https://services.fsd.tuni.fi/catalogue/series/3?lang=en.

\textsuperscript{126} The Sociobarometers are annual surveys conducted in Finland by the Finnish Federation for Social Affairs and Health (SOSTE) evaluating Finnish social services and welfare program changes. Further information about SOSTE can be found at https://www.soste.fi/soste-in-english/.
Science Data Archive, n.d.c). However, there are non-anonymized data (both including direct identifiers and pseudonymized data) at FSD, and there are very specific processes and legal bases in place for these data in accordance with the GDPR (see below). In line with the anonymization policy, FSD restricts data deposits to Finnish, Swedish, and English language data because the anonymization in these languages can be readily verified by FSD staff. If a data provider wants to deposit data in other languages (such as for a multi-national collaborative project with a Finnish-affiliated researcher), the researcher must contact FSD well in advance and strictly follow the data minimization and anonymization guidelines (Finnish Social Science Data Archive, n.d.c; see also FSD01).

The majority of data FSD archives are archived under the “informed consent” legal basis (Art. 6(1)(a) GDPR). However, there are other legal bases for personal data archiving. Article 6(1)(e) allows for processing that serves the public interest. The Finnish Data Protection Act (2018), § 4 expands explicitly on this article:

Personal data may be processed in accordance with point (e) of Article 6(1) of the Data Protection Regulation if: 1) the data describe the position of a person, his or her duties or the performance of these duties in a public sector entity, business and industry, activities of civil society organisations, or other corresponding activities, in so far as the objective of the processing is of public interest and the processing is proportionate to the legitimate aim pursued; … 3) the processing is necessary for scientific or historical research purposes or statistical purposes and it is proportionate to the aim of public interest pursued; or 4) the processing of research material and cultural heritage material containing personal data and the processing of personal data included in their metadata for archiving purposes is necessary and proportionate to the aim of public interest pursued and to the rights of the data subject.

According to FSD01, the GDPR and this section of the Data Protection Act allow FSD to archive data which still retain indirect identifiers under this “public interest” option. However, the researcher is still required to
to inform research participants that the data is archived with personal identifiers, some kind of personal identifiers, it may even not be exact names, but that much indirect identifiers that it's easy to identify those persons, but sometimes it might be even names. But the researcher has to make the privacy notice already in the very beginning, so that it enables the archiving of personal data for reuse at FSD (FSD01).

FSD04 also described a public use exception where the data subjects have not agreed to the archiving:

But if the data doesn't contain anything very... health data or similar, where that would be very strictly protected, or it is very important, then we might be able to accept it. One example about the people who run a position in the national parliament, that data was considered to be publicly important. So, we would archive it, even if that all those people hadn't given their permission for it.

In this situation, data can also be archived under § 27 of the Data Protection Act, which reconciles Article 85 GDPR (this article directs the Member States to reconcile personal data protection and the “right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression”). As FSD01 argues, the GDPR does not define what does it practically mean. But the thing that the law says is that it had to be necessary and proportionate. And also, anonymizing has to be impossible. For instance, if you do that, you destroy the possibilities to use the data and so on. … [In this case, FSD01 advises] to do the privacy notice so that it would be possible to archive to our data archive. And so, for that kind of data, you have to look for if minimization would be possible. And that means that even though you can archive the personal information, you shouldn't have more personal information than is needed to be able to use the data.

The data this applies to at FSD are data collected from media, except social media, or data about which the analysis... and publications that included personal information. And usually that is, for instance, interviews, very well-known experts of some fields or artists. And the interviews may include also referring to the published works of those artists and so on. So, there is this ‘freedom of speech’ exception by which we can archive certain types of data with identifiers. And in addition, after GDPR, there is an option to use some... to have personal data, if the
archiving is a bit of public interest (FSD01).

4.2.1.2 Organizational structure

Organizationally part of Tampere University, FSD is considered a national research center and is the national service provider for CESSDA. Under the Universities Act, Tampere University has administrative control and responsibility for the archive, despite Finnish government funding of the archive. FSD is relatively independent in its operations, including how it has responded to the GDPR. The reason for this is that the university (both the then-University of Tampere and the newly-created Tampere University) does not have enough experience with data archiving considerations (FSD04). In addition to funding from Tampere University, FSD receives funding from the Ministry of Education and Culture, as well as both the Academy of Finland (Suomen Akatemia) and CESSDA to support infrastructure projects.

FSD has a National Advisory Board (Valtakunnallinen neuvottelukunta) appointed by the president of Tampere University. The role of the National Advisory Board is “to evaluate the operations of the unit and its future development, submit motions and proposals related to the development of the unit, and to consolidate the collaborative network of the unit and enhance aware of data repository work” (Finnish Social Science Data Archive, n.d.b). There are 14 members of the advisory board (technically 28, as each member has a deputy member). The members represent research institutes, universities, and other organizations in Finland; there are no representatives from non-Finnish organizations.127

Data protection in Finland is coordinated by the Office of the Data Protection Ombudsman (Tietosuojavaltuutetun toimisto), which is responsible for supervising compliance

---

127 The members of the National Advisory Board represent: Tampere University, the Finnish Institute for Health and Welfare, the Academy of Finland, the National Archives of Finland, Statistics Finland, the CSC – IT Center for Science, the National Library of Finland, the University of Helsinki, the University of Eastern Finland, the University of Jyväskylä, the University of Turku, Åbo Akademi University, Tampere University Library, and the Student Union of Tampere University.
with the GDPR and the 2018 Data Protection Act, investigating complaints, commenting on legislative changes which may affect data subject privacy and rights regarding their personal data, and representing Finland on the European Data Protection Board, among other roles (Tietosuojaavaltuutetun toimisto, n.d.). As Ojanen (2017) states, the Data Protection Ombudsman is heavily involved in the Finnish preventative approach to data protection, instead of a retroactive approach reliant on legal proceedings to reactively address privacy concerns. However, the Data Protection Ombudsman does also issue fines for organizations which do not comply with the data protection requirements, and has mandated that Finnish company websites must not just inform users about cookies, but must also allow users to refuse cookies (Fagerström, 2020).

4.2.1.3 Legal framework for FSD

In addition to the GDPR and the 2018 Data Protection Act, there are four additional laws directly regulating the Finnish Social Science Date Archive’s (FSD) operations. The first law, the 1999 Act on the Openness of Government Activities,128 details how anyone can request information collected by public authorities. Data from a public authority’s personal data filing system may be shared “if the person requesting access has the right to record and use such data according to the legislation on the protection of personal data” (§ 16(2)). At FSD, these government register data without identifiers can be deposited; however, data collected under this Act do not usually have permission for archiving, and researchers who want to archive these data need to obtain this permission from the register data providers (Finnish Social Science Data Archive, n.d.e). This permission is something that FSD staff check for during deposit review.

The second law is the 1961 Copyright Act. There is no “fair use” clause for archiving copyrighted material collected during research under the Finnish Copyright Act, according to FSD01. To that end, in 2015, FSD and Kopiosto, one of seven copyright societies in Finland, signed an agreement “that allows FSD to archive and disseminate for reuse certain digital or digitized works analysed in research, namely works belonging to the fields of rightsholder organisations represented by Kopiosto (for instance, magazine articles, photographs, illustrations and comics)” (Finnish Social Science Data Archive, n.d.e).

The 2019 Act on Information Management in Public Administration established data management and data protection measures aimed at protecting public authority data and promoting openness. § 3 states that this law applies as well to universities governed by the Universities Act, which Tampere University is. As a component of Tampere University, FSD’s data security and data processing operations are governed by this law. The final law, the 2019 Act on the Secondary Use of Health and Social Data supplements the requirements of the GDPR. In particular, it allows for processing of social and health care data for statistical research, scientific research, development and innovation, education, and knowledge management (among other purposes) “even if the data were not originally stored for such a purpose” (§ 2).

4.2.2 Data archiving in Finland

The Finnish Social Science Data Archive (FSD) is considered a national resource center in Finland, and is funded by the Ministry of Education and Culture and Tampere University (Finnish Social Science Data Archive, n.d.a). It is one of 13 data repositories and archives in Finland, and is the only one listed by the Registry of Research Data Repositories as specifically supporting social science data archives (re3data.org, 2022). There is no legal mandate in Finland for researchers to archive their research data or to specify where research data would be archived, though research funders may require it. However, as FSD02 said, FSD is the only archive in Finland for social science data, and “if they don't intend to archive it with us, I don't know if the reviewers would think that researcher doesn't know what they're doing. And I'm not sure how they assess that part of the application then.”

FSD’s primary service is to archive and disseminate Finnish social science research data. Since FSD opened in 1999, it has predominantly archived anonymous data; it does have some data sets with personal data which are archived under the “public interest” and “freedom of speech” exemptions. These data are predominantly data about public figures. It does not have any secure or remote data access infrastructure to support use of non-anonymized data. There is a health data service, the Finnish Social and Health Data Permit Authority (Findata), which does have a remote access system for personal (health and social security) data analysis, in accordance with the Act on the Secondary Use of Health and Social Data. Findata provides access to aggregated data which do not require a permit, as well as individual level data which require a permit and can only be accessed through its secure platform, Kapseli (Findata – Finnish Social and Health Data Permit Authority, 2022). Though this option is available for the health and
social security data, FSD does not have this capability, and according to FSD02, it is cost prohibitive at this time to develop this infrastructure for FSD.

4.2.3 Privacy and the GDPR in Finland

Panic and uncertainty were the primary reactions to the GDPR in Finland, according to several FSD interviewees. FSD03 connected the panic in the research community specifically to a lack of awareness of existing data protection legislation and how the GDPR was discussed in the media:

Basically, people, generally, they weren't aware that we have these. … there are basic security or consent, researchers will think about them in terms of ethics. So, they sort of understood that they were basically applying same sort of principles through different set of norms. But how the GDPR was in media, talked about in media, that's everything is going to be everything's going to stop... But people were panicking that this is the end of research, and how do you handle this? But that was a general reaction, it was for panic.

However, Finland has continuously had a data protection law since the 1987 Personal Data File Act,132 the purpose of which was “to protect the privacy, interests and rights of the person, to ensure the security of the State and to maintain good data file practice” (§ 1). Additionally, since 2004, the Act on the Protection of Privacy in Working Life133 has restricted employee personal data collection and processing to only that which is absolutely necessary, with no exceptions, even if the employee consents (§ 3). There is historical motivation behind this, as in the early 20th Century, records of union membership were used by employers to refuse to hire union members and to fire existing employees (Suuriranta, 1995). Therefore, there were personal data

---


protection laws to protect privacy in Finland for over thirty years by the time the GDPR entered into force.

However, there was also the view that the GDPR was not a significant concern, because Finnish people have a long-standing reputation for valuing their privacy. In 1995, Suviranta described this as “they like to be left alone, without other people meddling in their private matters” (Suviranta, 1995, p. 45). FSD05 similarly said that

People talked about [GDPR] for a while, but now it’s already a normal part of life. I don’t remember that there would have been a big critiques against it. People here respect their privacy and I think they saw [the GDPR] as a good thing, although a bit complicated.

As a right, privacy is included in the Constitution of Finland:134 “Everyone's private life, honour and the sanctity of the home are guaranteed” (§ 10). The same section discusses personal data protection as a component of privacy, noting that the provisions for it are in a different act (the 1999 and then the 2018 Data Protection Act). However, the Constitution also guarantees freedom of expression and right of access to information, something which comes into conflict with that right to privacy. This guarantee supports one of the Finland-specific bases for data processing and archiving when informed consent for archiving has not been granted. Finnish courts, such as the Supreme Administrative Court (Korkein hallinto-oikeus), have only rarely ruled on digital privacy and data protection, which according to Ojanen is because the “Finnish data protection regime is primarily based on preventative methods to ensure compliance by the data protection authorities, instead of retroactive court proceedings” (Ojanen, 2017, p. 86).

Thus, while Finns are perceived as valuing their privacy, there is disagreement as to whether, and to what extent, the general Finnish public were aware of their existing privacy

---

rights. Several pre-GDPR studies highlight these mixed views. For example Sirkkunen (2015) conducted the first Finnish study investigating what Finns thought about their online privacy. He found, as Suviranta (1995) had twenty years earlier, that Finns still valued their privacy, with 68% of survey respondents concerned about how much information was being collected about them online, 76% wanting to know what the data were and what they were used for, and 87% wanting to decide how their data were used. He framed this last point as the respondents being “concerned about losing their right to self-determination” (Sirkkunen, 2015, p. 5). However, he also identified a contradiction between that concern and the behaviors the respondents took to protect their privacy online. Whether respondents had actually read the terms and conditions or end user license agreements depended on the service (63% reported reading the Facebook terms, while 40% or fewer reported the same for Google, Instagram, or Whatsapp (Sirkkunen, 2015, p. 12). Sirkkunen also found that, contrary to the perception of the young “digital natives” as blasé about their online privacy, the younger respondents were more protective of their privacy than middle-aged and elderly respondents; only 29% of 15-24 year-olds felt that online privacy did not matter because they had nothing to hide, while 53% of 65-70 years felt the same. Younger respondents were also more critical of government surveillance of online behavior than the middle-aged and elderly.

Malik et al. (2015) conducted a survey of Facebook users from Finland (163 respondents), the United Kingdom (56), the United States (50), and “other” countries (109) from December 2014-January 2015. The research team was interested in whether and how the users’ concerns and awareness about privacy influenced their behaviors on Facebook. The researchers did not report the country-by-country results of the survey, and it is not apparent whether there were any findings that specifically distinguished the Finnish respondents from the others.
Despite this, the research team reported a positive relationship between a users’ knowledge and awareness of privacy concerns on Facebook and how active they were; thus, a well-informed user would know what to expect, and could adapt their behavior accordingly (Malik et al., 2015, p. 375).

This finding is interesting when compared to two other of Sirkkunen’s findings, that only 13% of Finnish respondents had confidence in Facebook’s personal data protection practices, and that 36% of respondents had stopped using certain online services because of “insufficient” privacy protection (Sirkkunen, 2015, p. 12). Kini and Suomi’s pre-GDPR study of United States and Finnish college students also found that, when comparing between the two groups, Finnish students were “more influenced by the risk of losing privacy and control” of their personal information in the context of location-based advertising on their smartphones (2018, p. 77). This also reflects Sirkkunen’s (2015) finding that younger survey respondents were very concerned about their online privacy. Elsewhere, Finnish media were reporting as early as 2014 that Finnish internet users were already exercising a “right to be forgotten” against at least 3,000 web providers that would become formalized European Union-wide in the GDPR (e.g., Brunila, 2014; Naalisvaara, 2014).

In addition to these studies and media reports about Finnish perspectives and concerns about online privacy and data protection from before the adoption of the GDPR, the 2015 Eurobarometer 83.1 survey (European Commission, Brussels, 2018; see also Appendix F) also asked respondents from all EU Member States about their online behaviors, concerns over personal information, and what they felt were the biggest risks to providing personal information online. In terms of feeling in control of their information online, a year before the GDPR was adopted, the vast majority of Finnish respondents felt they had only partial control (66.44%),
while 14.64% felt in complete control. However, at the same time, more than half felt concerned about this limited control (63.25%), not too far below the European Union average (65.13%).

When asked about whether explicit consent should be obtained before collecting or processing personal data, there were three possible “yes” answers, and multiple responses were possible: “yes, in all cases” (65.94%), “yes, in case of personal information required online” (18.80%), and “yes, in the case of sensitive information whether online or offline” (11.12%). The spread of “yes” responses varies on the EU level, but in all EU Member States, more than half said that consent should be required in all cases. The EU average for consent in all cases was 70.21%.

Thus, despite the reputation that Finns value their privacy, studies of internet behaviors and perspectives before the GDPR, including the EU-wide Eurobarometer 83.1, show that levels of concern in Finland differ between generations. Finnish respondents to the Eurobarometer as well reported their concerns about specific activities at a lower rate than the EU average, though the most serious risks of providing information online, according to the Finnish respondents, were all reported at a higher rate than the EU average.

This has started to change over time, as has been reported in post-GDPR studies in Finland. Both FSD03 and FSD05 described an increased awareness within Finland about the importance of data privacy and also of what rights people had over their data. This was growing even before the GDPR (for example, the growing number of “right to be forgotten” requests noted as early as 2014). FSD03 commented that much of the panic reported was driven by media reports on the GDPR. For instance, news articles highlighted fears that photographs of

---

135 For claims that involve contrasting rates between one of the four countries of interest and the EU excluding the four countries of interest, the smallest sample sizes were 588 (single country) and 14406 (remainder of EU). In this worst-case scenario, the standard deviation of the difference of two percentages is no more than 2.1%, yielding a 95% margin of error of 4.2%. Thus, any difference between two percentages reported here that exceeds 4.2 percentage points is statistically significant.
school children could no longer be taken and shared (Lappinen, 2018). On the other hand, many articles in both general and industry-specific newspapers took optimistic perspectives on the upcoming GDPR. For example, an April 2017 article in the commercial newspaper Kauppalehti assured readers that the GDPR was not a “panic project” and urged them to focus on the benefits the GDPR will bring and not just the amount of work required (Kauppalehti, 2017).

The Special Eurobarometer 487a (2021) conducted in March 2019 (see Appendix G) found that 34.90% of Finnish respondents had both heard of the GDPR and knew what it meant, a slightly lower rate than the EU average (35.85%). Of the data subject rights, Finnish respondents reported knowing about all of the rights at a higher rate than the EU average. For five of the six, more than half of Finnish respondents knew about the rights of access (75.50%), objection to marketing (79.70%), rectification (75.50%), to be forgotten (57.30%), and portability (53.50%), while less than half (41.30%) knew about the right not to be subject to automated decision making. Of these, already by March 2019, 39.30% had already exercised their right to object to direct marketing (only slightly less than those who simply knew about the right (40.40%). By early 2019 as well, the Finnish Data Protection Ombudsman had received more than 3,100 complaints from Finnish customers who had tried to exercise their right of access, but the companies were not prepared to provide those data (STT, 2019). The Special Eurobarometer 487a also found that the percentage of Finnish respondents feeling partially or completely in control of their online information had shifted from the 2015 Eurobarometer 83.1, from 66.44% to 65.56% (partially) and 14.64% to 16.01% (completely). There was a small decrease in concern about the limited control, from 63.25% in 2015 to 61.65% in 2019 (European Commission, Brussels, 2018, 2021).
4.2.4 Timeline of GDPR response

FSD began preparing their response to the GDPR at the beginning of 2017. Later in 2017, FSD formally appointed a data protection officer (DPO). The Finnish GDPR implementation law, the Data Protection Act, was not adopted until November 2018, several months after the full entry into force of the GDPR and did not itself enter into force until January 1, 2019. This delay, acknowledged by several interviews as contributing to the uncertainty around how FSD needed to respond to the GDPR, is the result of Finnish legislative processes that can be time consuming. The law went through a preliminary proposal process called “pantiin mietintö” that was relatively slow, and the Data Protection Act itself was being debated around the same time as two other related laws, the Act on Information Management in Public Administration and the National Archives Act.136

Concurrent with the GDPR response, the University of Tampere, FSD’s previous parent organization, and the Tampere University of Technology were merging into the current Tampere University, FSD’s parent organization. The merger occurred on January 1, 2019, though the complete transition of FSD’s website and online services to the new domain was not complete until November 26, 2020 (Finnish Social Science Data Archive (FSD), 2020). Tampere University is part of the Tampere Universities community, a collaborative partnership with the Tampere University of Applied Sciences. This merger added an additional administrative burden to the on-going GDPR response efforts and the regular, day-to-day operations. Despite the delay in adoption and implementation of the Data Protection Act as well as that organizational restructuring, FSD staff feel that the archive itself was well prepared for the

GDPR to enter into force, with the immediately necessary new and changed policies in place before May 25, 2018.

4.2.5 Priorities in the response

FSD was systematic in its work to become GDPR compliant. FSD staff began the process by mapping out the information and data flows into and out of FSD, and once those were all identified, the focus shifted to determining the best ways to ensure that their processes were legally compliant and to demonstrate it. As FSD01 explained while describing the “Privacy Policy” revisions, “as the data controller for our customer data, we have to indicate and then demonstrate our compliance with GDPR.”

4.2.5.1 Priorities and goals development

The FSD staff actively involved in planning and implementing changes at FSD in response to the GDPR were the director, data protection officer, legal advisor, archive team manager, and IT director. The data protection officer was not formally appointed until fall 2017, but the staff member in that role had been the point of contact for privacy issues for several years already. The legal advisor was hired early in the process; in additional to the legal background, the legal advisor had also been working in FSD’s archive for several years as a data processor. It was the legal advisor who provided most of the assistance and who reviewed and drafted proposed changes during the transition period. The GDPR transition process at FSD was independent of the process at the then-University of Tampere because the University administration did not “have much archiving things other than us. So, they really couldn’t guide us or didn’t want to tell us what to do, because they felt that we knew best how these things would be handled” (FSD04). FSD did not consult the Data Protection Ombudsman either, for similar reasons: “They don’t know really what we do, or they didn’t know anything about us”
FSD’s priorities when responding to the GDPR focused on: the contracts and agreements; the data acquisitions requirements; and, the administrative and operational data. Within these priorities, there were a number of changes and new developments that were made, or still remain to be completed.

4.2.5.2 Processor versus controller

At FSD, there are two types of data under the GDPR’s purview: the Aila research data (specifically the data with personal identifiers archived under the public interest exemption and the relatively new pseudonymized data) and the administrative and operational data (including the data provider and data user Aila registration data). FSD is the processor for all data available in Aila, with the exception of the Finnish components of the ISSP, where “FSD is the controller… since FSD applies and gets the financing for collecting Finnish data. For other international data we are in the normal role: as the processor” (FSD01). FSD’s role as processor for these data directed several changes to their data deposit processes which will be discussed in Section 4.2.5.3. For the administrative and operational data, FSD is the controller, as it both collects and carries out any necessary processing for these data.

4.2.5.3 Regulating information flows

FSD made a number of changes targeting the flow of data into and out of Aila. These involved: the agreements for depositing data; the requirements for data deposit; the terms of use; and, documenting its processes. The first priority was to update the legal contract that authorizes the transfer of data from the data provider to FSD. There are now two agreements that must be signed before data deposit can occur: the “Deposition Agreement” (Arkistointisopimus) (Finnish Social Science Data Archive, 2018f) and the “Agreement on personal data processing for assessing the suitability of research data for archiving” (Sopimus henkilötietojen käsittelystä
arkistointiedellytysten arvioimiseksi) (hereafter, “Processing Agreement”) (Finnish Social Science Data Archive, 2018e). Pre-GDPR, data providers only needed to sign the “Agreement on the deposition of data to the Finnish Social Science Data Archive” (Finnish Social Science Data Archive, 2013a). The pre- and post-GDPR deposit agreements are similar with much of the required information the same. However, there are significant differences in the information provided to the data provider.

First, the 2018 “Deposition Agreement” is more explicit about how and why the data provider’s information will be retained in a register. It is also more explicit about what type of data processing FSD will do on deposited data, while in the 2013 “Deposition Agreement,” the only reference to FSD’s processing is that the agreement established FSD’s right to process data for its operational needs. The 2018 “Deposition Agreement” also requires the data provider to list the types of personal data and the categories of data subjects that are included in the deposited data. This agreement is only signed once the deposited data have been reviewed and approved for deposit by FSD, or at the latest, when the Dissemination Information Package has been created (Finnish Social Science Data Archive, n.d.e). For the few data sets already in the archive that had personal data, FSD did need those data providers to agree to the new agreement, which they could do just by replying to a notification email to say that they accepted it (FSD01). For data that were deposited before the agreements went into effect not yet checked or processed, FSD sent those data providers the “Processing Agreement,” but they did not need to sign or acknowledge the 2018 “Deposition Agreement.”

The “Processing Agreement” is entirely new to FSD and is a direct result of both the GDPR and FSD’s decision to be the processor of archived data. FSD determined that they needed this agreement separate from the “Deposition Agreement” because they had noted over
the previous several years that around 70% of data deposited at FSD were not anonymized even when the data providers said that they were (FSD01 & FSD02). Per this agreement, FSD expects the data provider to anonymize the data before deposit, and then explicitly states that FSD will process all data, whether the data provider indicated the presence of personal data in the “Deposition Agreement.” The data provider must agree to the “Processing Agreement” before data are transferred to and reviewed by FSD. According to FSD01:

… it's easier for us to have, every time data is transferred to FSD, this agreement. And it's handled electronically, in connection with email correspondence, it does not require a separate signature.

Both agreements came into effect in March 2018.

The next set of changes at FSD relate to the requirements for data deposit. These changes are about how FSD handles data anonymization and how informed consent is enforced. Since FSD opened in 1999, the policy has been to actively only archive anonymized data, because FSD has never had the resources or infrastructure to provide secure access to identifiable data (FSD02). While this is starting to change—according to FSD01 in a 2021 follow-up interview, FSD has developed the processes to work with pseudonymized longitudinal data—FSD puts the responsibility for anonymization of the data pre-deposit more firmly on the data provider: “… earlier, we could anonymize the data ourselves. But nowadays, we demand that data should be anonymized beforehand before it's ingested” (FSD02).

After FSD began planning the response to GDPR, it has revised the guidance about the responsibility for anonymization provided on the website several times. In the current version, FSD is clearer that the data provider do the anonymization work before the data are transferred (Finnish Social Science Data Archive, n.d.f). FSD does provide extensive guidance for data providers for this anonymization in Finnish and English (but not Swedish) on the website,
including an anonymization template for researchers and data providers to follow (Finnish Social Science Data Archive, 2019d). As part of the deposit review process, “FSD reviews the anonymisation and makes additional changes if necessary,” which are then approved by the data provider (Finnish Social Science Data Archive, n.d.e). Despite language on the FSD website that indicates that full anonymization work could be requested, FSD01, FSD02, and FSD05 all reported that FSD no longer wants to actively do anonymization work beyond that described in the separate processing agreement. In the deposit review process, FSD is also much stricter as well about reviewing the completeness of the anonymization work done by the data depositor.

The next set of changes were to the Terms and Conditions (ToC). FSD has two separate ToC’s, one for the data in the archive and one for the use of the Aila system. The “General Terms and Conditions for Data Use” (hereafter, “General ToC”) are applicable to both the qualitative and quantitative data (Finnish Social Science Data Archive, 2020a). There were minor changes to the “General ToC” between 2014 (Finnish Social Science Data Archive, 2014a) and 2020, as the updated version has added a scope limitation that states that the 2020 “General ToC” do not apply to data under the CC BY 4.0 license (Finnish Social Science Data Archive, 2020a). Otherwise, the text of the “General ToC” is essentially unchanged. On the other hand, there were more changes to the “General Terms and Conditions for the Use of the Aila Service” (hereafter, “Aila ToC”) (Finnish Social Science Data Archive, 2020d) compared to the 2014 version (Finnish Social Science Data Archive, 2014b). The “Aila ToC” also stated that the CC BY 4.0 data do not require Aila registration to be downloaded through the Aila portal and added a requirement that the data (including the CC BY 4.0 data) must be cited if used in publications or presentations (Finnish Social Science Data Archive, 2020d). Other sections were essentially unchanged, though it is less informative about how the user’s personal data would be
processed; this information is now in the GDPR-compliant “Privacy Policy” (Finnish Social Science Data Archive, n.d.d).

FSD also formally documented their processes, as required by Art. 30 GDPR. For IT processes, for example, “there was a need to make more formal documentation so that we can show everyone how we are doing it. And some processes and technical details have to change quite a lot to meet the new requirements that that GDPR set up on us” (FSD04). In 2016, FSD also began producing an annual Data Balance Sheet (Tietotilinpäätökset), which includes descriptions of the data FSD has, how the data are processed, data breaches, and so forth, separate from the annual reports (Vuosikertomukset). The Data Balance Sheets are internal reports (though publicly accessible), and do not need to be shared with other organizations, such as the Data Protection Ombudsman.

4.2.5.4 Minimizing harm

The next set of changes that FSD made were intended to minimize the potential harm to data subjects. Under the GDPR, FSD had to address a number of issues related to data subject consent for archiving at FSD and the privacy policy. For all data collected before May 25, 2018, FSD05 says that “it is enough if the respondents have been informed about the archiving in some way, not necessarily in the privacy statements.” For all data collected after the GDPR became enforceable, however, FSD now requires that the data subjects have been informed that their data will be archived in the “privacy notices” (Finnish Social Science Data Archive, 2019c) provided by the researchers.

---

137 The Data Balance Sheets are publicly available in Finnish at [https://www.fsd.tuni.fi/ki/tietotilinpäätökset/](https://www.fsd.tuni.fi/ki/tietotilinpäätökset/).
139 FSD interviewees consistently referred to “privacy notices” given to research participants; the “privacy notices” are what are referred to elsewhere as “informed consent” forms or documentation. For consistency in this case report, I will refer to these as “privacy notices.”
If the data are to be archived at FSD, the notice should identify FSD specifically as where the data will be archived, as well as the level of accessibility (open or under some access restriction). FSD requests a copy of the privacy notice as part of the deposit to confirm that this requirement is met (Finnish Social Science Data Archive, 2018a). If the data subjects have not been explicitly informed in the privacy notice that their data will be archived at FSD, then the data cannot be: “For the new data, we demand that everyone should know about archiving to us, before they answer the study or research or whatever it is, that it would be archived to us” (FSD04). However, if the data depositor contacts every data subject to inform them about archiving and obtains their consent to it, FSD will accept this “updated” privacy notice and move forward with data acquisition (FSD05). As discussed in Section 4.2.1.1, there are other legal bases for data archiving, so this requirement does not apply if the data provider uses one of the other bases to support their data archiving efforts.

The next set of changes FSD made targeting harm minimization focus on administrative and operational data: the privacy policy and the administrative and operational databases. As the controller for the data collected and processed for FSD’s operations, FSD must be able to demonstrate its compliance with what the GDPR calls “data protection by design and by default” (Art. 25 GDPR), which it does through its revised “Privacy Policy” (Finnish Social Science Data Archive, 2018b). The new policy was shared with disseminated on May 17, 2018, a week before GDPR went into effect.

The previous “Privacy Policy” (Finnish Social Science Data Archive, 2015c) was generally informative about the purposes for processing and the data subjects’ rights, but the 2015 “Privacy Policy” did not refer to any particular legal requirements, nor did it explain in any detail the purposes for data processing. The 2018 “Privacy Policy” (Finnish Social Science Data
Archive, 2018b), in contrast, clearly identifies FSD as the controller and that FSD users whose data are collected in any form are data subjects with rights under the GDPR. As part of the review of data holdings that FSD conducted at the start of the GDPR transition period, five distinct user groups for whose data FSD was the controller were identified. This delineation of the five different user groups is important because the data subjects’ rights and applied legal bases for processing differ for each group, and are listed separately for each group.

The 2018 “Privacy Policy” was added to every FSD webpage a week before the GDPR went into effect (previously, the policy was accessible through the “About us – Documents” page, but was not otherwise easily accessible through the website as it is now). That website update process itself “was for most part automated of course. But it was because we have to, as the data controller of our customer data, we have to indicate and then demonstrate our compliance with GDPR” (FSD01). It has been minorly revised since 2018.

4.2.5.5 Response to technological risks

The changes FSD made addressing the privacy risks from a technical perspective were made in how FSD addresses data retention periods and in data protection measures, as well as to data protection training. As part of the new “Privacy Policy,” FSD set specific retention periods for the given data. In addition to establishing these retention periods, FSD’s IT staff have also begun to apply automatic deletion processes to more of the databases to reduce the manual work.

---

140 The five user groups are: registered Aila users; users who submitted a use report for FSD’s research methods resources; data depositors with a deposition agreement; contact information from user services and user communication; and web-traffic data from cookies and Google Analytics (Finnish Social Science Data Archive, 2018b). As of December 2021, there is a sixth user group, “Candidate selector data,” which are “responses given by election candidates to candidate selectors/voting advice applications in Finnish Parliamentary, European Parliamentary and Finnish Presidential elections. The data can be reused for research. The media collects candidate selector data for journalistic purposes in compliance with Section 27 of the Finnish Data Protection Act and provides open access to the data with the consent of the candidates” (Finnish Social Science Data Archive, n.d.d).

141 The only direct link to the pre-GDPR privacy policy was found under the “About us” page (https://web.archive.org/web/20180419170439/http://www.fsd.uta.fi/en/organisation/documents/AMS/privacy_notice.html).
required. According to FSD04, this process was already in place for the Aila data download portal, where “account information is automatically deleted if the customer doesn’t use our service for two years.” For one type of personal data, FSD developed a retention period where one had not previously existed: the contact information for researchers whom FSD identified while searching publications for possible new data to bring into the archive. For these data,

… nowadays we must contact the researcher in 30 days, if we want to keep his/her contacts (email address) in our database. … Before GDPR there wasn’t any time limit on our procedures to send the acquisition e-mail. If we don’t send the acquisition e-mail in one month latest, the contacts of the person must be deleted from our database. We also have to say in the acquisition e-mail from where we have got the contacts (source of the personal data) (FSD05).

Finally, FSD01 worked with FSD04 to expand the training on data privacy and security:

… an ongoing process. We have a special training data once a year. … It’s internal, me and the data protection officer, we are the ones to give the training. … There were trainings before [2018 and the GDPR], but now they are longer and more aimed at the privacy matters (FSD04).

For a period of time at the start of the transition process, FSD01 and other FSD staff also found themselves providing both formal and informal training on GDPR’s requirements to researchers and potential data providers. This training increased internal and external awareness of the privacy risks possible in the research data and in the analysis.

4.2.5.6 Other changes

Outside of the changes aligning with my theoretical framework, there was one other change FSD made as a direct result of the GDPR: the early, formal establishment of a data protection officer (DPO). While there had previously been a point of contact for legal and privacy issues at FSD, there was no position of DPO. However, in accordance with Article 37 GDPR, the FSD management team formally appointed them as the DPO in fall 2017. In addition
to the responsibilities laid out in the GDPR, the DPO also meets with all new FSD employees to discuss FSD’s data protection practices and the GDPR, to reinforce that “that data protection is actually a very big role in everyday life” (FSD01). FSD also temporarily hired a separate legal advisor who also had experience working with the Aila data to support the DPO in identifying needed changes, reviewing policy revisions, and serving as an additional advisor to the GDPR response.

4.2.5.7 Changes remaining to be completed

All respondents agree that FSD has completed their transition to make FSD GDPR compliant, although FSD02 thinks that they will need to wait and see how the legal practicalities will play out through the Finnish and European court systems. There are only a few things that the respondents feel remain to be completed. The first is the Article 35 GDPR data protection impact assessment (DPIA). While all the information for the DPIA exists in the Data Balance Sheets (Section 4.2.5.3), “the DPIA is just something... because no one has demanded it from us, we have just a habit every year. Yeah, let's do it. And it's up to me. But there is always something that seems to be more important” (FSD01). FSD04 also feels that improvements to the IT systems are necessary, not to be compliant, but to automate processes like data deletion that are currently manual:

Now, sometimes we have to do some extra manual work to ensure that everything is as it should be. Because some old systems have been created years ago, before there was any knowledge about GDPR. So, the core of them isn’t really made in a way that we would do it now if we could start over.

4.2.6 Challenges at FSD

While FSD was well-prepared for GDPR by May 2018, there were challenges along the way. These challenges primarily focused on: the general workload, uncertainty about what
GDPR meant for data archives, and interactions with FSD’s users. Responding to the GDPR was considered an administrative burden on top of the normal, day-to-day archive operations and the preparations for the Tampere University merger. Both FSD01 and FSD04 highlighted a lack of additional resources, both budgetary and staff hours, which meant that priorities and plans for other projects had change so that FSD staff could focus on the GDPR-related needs.

That increased workload was complicated by confusion about just how GDPR actually applied to FSD. The GDPR is a large and complex regulation, and there were also several months of delay before the national Data Protection Act (2018) was passed. This gap was in part due to the lack of a legal basis specifically for data archiving in the Finnish law, which had to be corrected and accounted for. The legal advisor during the transition, FSD03, described the situation in this way:

… for you to be prepared to apply GDPR you would have to also know how there are part of the GDPR that require national legislation, how they would be implemented, but something really interesting there was that the legal basis that you could base, based on GDPR for archival purposes and everything related to that was actually missing. It was a blank spot, yeah. … As an organization you would be in this state of uncertainty basically. And, especially the archives, it was missing there.

The implications of this at FSD can be seen, for example, in the Data Protection Act’s § 4 expansion on “public interest data” in Article 6(1)(e) GDPR and § 27 reconciliation of the “freedom of expression and information” in Article 85 GDPR, as discussed in Section 4.2.1.1. FSD03 does acknowledge that FSD did at least have an easier time than other organizations, even with all of the confusion, because FSD was very aware of the importance of data protection and privacy even before the GDPR.
Confusion about how GDPR would affect research data on the part of data providers also proved a challenge for FSD staff. At the same time that FSD staff were trying to familiarize themselves with the GDPR and figure out how it may apply, they had to help a lot of researchers and that was really difficult for our staff, for those who do the customer services because it was even difficult for us to learn that thing. And then where we were in the position of trying to help the researchers, that was difficult (FSD01).

According to FSD05, for the next year after May 2018, they “had to advise the researchers on its effects to archiving process. It took some time that the researchers realised that they really have to have privacy statement for example, if they collect and archive data.”

FSD01 attributes much of this data provider confusion to a lack of knowledge and understanding, especially early on, about GDPR’s impact on research by the researchers’ own institutional data protection officers and legal advisers. While FSD also faced uncertainties, both FSD01 and FSD03 had experience with research data, data archiving, privacy concerns, and legal requirements from the pre-GDPR period. The data providers and newly appointed data protection officers, however, according to FSD03, “weren’t aware that we had these [previous laws],” and the general reaction within Finland (in general and in the research community) was “panic” (see also Section 4.2.3). Because FSD was and is seen as well-prepared and knowledgeable by other people and institutions, it is more well-known now than it ever was before, and many people come to FSD with their questions (FSD02). This has decreased over time, as the researchers and their DPOs have become more familiar with the laws, and FSD now re-directs researchers with questions back to their institutional DPOs for most questions.
4.2.7 Impacts on FSD operations

In the time after GDPR’s adoption, FSD made a number of policy changes and became a more well-known resource and advisor for privacy and data protection concerns. What, then, have been the outcomes of this experience on FSD as a research data archive? This section will examine the operational metrics of archive use—data deposits, data set publications, data set downloads, and user interactions—for trends during the GDPR transition period and discuss other outcomes identified through the interviews.

Since 2010, there has been a steady increase in the number of data sets available for reuse at FSD (Figure 4.5). The values from 2015-2021 reflect the total number of data sets through the Aila Data Service. As discussed in Section 3.3.2.2, the Aila Data Service only came into operation in May 2014, and so the values from this year include totals from both systems, and the 2010-2013 values are based only on the previous archiving system. As well, in October 2014, the small number of data set descriptions which had previously been published through FSD’s data catalogue (descriptions only, while the data were archived elsewhere) were removed.

from the Aila Data Catalogue permanently. The annual end-of-year totals have been adjusted so that the values represented in Figure 4.5 reflect only the data sets archived at FSD from 2010-2021.

Looking at the period of time from just before GDPR’s adoption to several years after it became enforceable (2015-2021), the overall trend for total data set publications is linear. However, when we look at the per year data set publication totals for 2015-2021, there is variation (Fig. 4.6). From 2017-2019, the number of data sets published through the Aila Data Service decreased; this was driven by the decrease in published quantitative data sets over the same period. However, at the same time, the number of qualitative data sets in Aila increased steadily from 187 in 2017 to 282 in 2021.

The data set publication numbers from 2017-2019 are an interesting reflection of the workload changes that occurred at FSD. FSD01, FSD03, and FSD04 discussed the need to change FSD’s priorities and workflows during this time to account for the work adapting to the GDPR. While nothing was specifically said about the data curation or data deposit review process, from the annual reports we can see that the processing time for deposited data sets
before they were published in the Aila Data Catalogue fluctuated: in 2017, 55% of deposited data sets were processed in 9 months or less; in 2018, this increased to 62%; and in 2019 it dropped to 50% (Finnish Social Science Data Archive, 2018c, 2019b, 2020b). These fluctuations may be connected to the re-prioritization that occurred.

The interviewees do not think that these fluctuations in the data publications from 2017-2019 were a direct result of the GDPR other than that re-prioritization of tasks. While not reported here, FSD01 also does not see that data provider concerns over GDPR had a significant affect in the rates of data deposit. FSD05 acknowledges an increase in qualitative data deposits over the period, but attributes the change to changing data management and data sharing policies at universities and research funders as the EU in general continues to promote open science and open data initiatives. However, it is possible that for data publication rates, the workload and FSD prioritization changes had some effect on how long the deposit reviews and processing took in the immediate period after FSD began working on the GDPR transition.

The situation is different when considering archive use by data users. From 2017-2019, the total downloads of data from Aila decreased from 2017 to 2019 (Figure 4.7). This decrease was driven by the drop in quantitative data downloads in that period. Qualitative data downloads fluctuated between 392 and 442 over that same three-year span, and there was an overall, though small, increase over the 2015-2016 period for qualitative data downloads. Since 2019, however, data downloads have increased sharply, which FSD interviewees have connected to the COVID-19 pandemic.
Figure 4.7 Qualitative, quantitative, and total data downloads from Aila (2015-2021) (Kuula-Luomi, 2022; University of Tampere, 2021).

Alongside the drop in overall data downloads from FSD from 2017-2019, there was a drop as well in the number of data access requests received for Category B, C, and D data holdings in the same period (Fig. 4.8). There are no corresponding data reflecting the number of approved data access requests, and the Aila Data Downloads data set counts each data download, and it is possible that there may have been multiple data downloads under the same data access request. Therefore, the number of downloads per year for these three categories do not match the number of received data access requests. The drop in data access requests from 2017-2019 came after three years of increasing numbers of requests. In 2014, there were 1,068 requests, representing both the requests submitted before May 2014, when each request could ask for multiple data sets, and after, when the Aila Data Service came into operation, and requests because restricted to a single data set per requests (Finnish Social Science Data Archive, 2015a). A similar trend of a 2015-2017 rise and then 2017-2019 drop can also be seen in the total number of registered Aila users (Fig. 4.9).
When asked about whether they saw an effect on data use from FSD in response to the GDPR, the interviewees did not think that there was a direct connection. However, 2017 to 2019 saw a significant increase in customer service requests that were specifically questions and requests for assistance related to GDPR, privacy, and data protection issues. For example, there were approximately 800 requests about these topics 2017 and approximately 1,600 requests in 2018 (FSD01). As FSD01 further described,
that's because of GDPR. Researchers were very nervous about GDPR. And most of them thought that well, when it's GDPR, now we can't collect or process or analyze any personal data. And that's not the case. So, there was needs, and people didn't know much about GDPR. And they were worried, and a lot of rumors telling that now everything would be really difficult and so on.

It is likely that, along with confusion about whether data collection was still possible under the GDPR, there was confusion also on the part of data users, their institutions, and/or their data protection officers about whether research using data from an archive was still possible, and that this was seen in the drop in download numbers. What motivations or confusion potential data users had about the GDPR, and how their behaviors regarding data reuse were affected by it, is unfortunately outside the scope of this dissertation, but remains an interesting question for future work investigating GDPR’s impact on research.

After 2019, data set publications (Fig. 4.6), data downloads (Fig. 4.7), and data access requests (Fig. 4.8) increased significantly. Indeed, 2021 saw the highest number of deposits, publications, total downloads, and requests ever reported. For example, in summer 2020, FSD01 had already described 2020 as “the most successful year when it comes to the amount of data downloads,” which 2021 well exceeded. The COVID-19 pandemic was considered to be the primary likely cause of this increased usage: “I guess people just can’t just go and interview, it’s easier to try to get ready-to-use data” (FSD01). However, the availability of open access data under the Creative Commons CC BY 4.0 license of interest to journalists and other data users may also play a role. While the increase in Category B data downloads was fairly steady from 2019-2021, there were more than 1,000 more downloads of Category A data in 2021 than in 2020. Both COVID-19 and the CC BY 4.0 data availability were highlighted in the news release reporting on FSD’s 2021 operations (Finnish Social Science Data Archive, 2022b).

There was general agreement that the most significant changes that happened at FSD because of the GDPR were the redesign of the “Deposition Agreement” and the requirement for
the “Processing Agreement” (FSD01, FSD02, FSD03, FSD05), which “added the privacy bureaucracy on our ingest process” (FSD01). After the contracts, the requirement that data subjects be notified specifically about archiving at FSD was considered especially important (FSD02, FSD05) because of the effects it has on whether data can be archived in the first place, as was the temporary post-GDPR-implementation need to remind researchers about and assist researchers with their privacy notices.

There are two other outcomes at FSD, neither of which are easily measurable. The first is that, even though FSD already had experience with data privacy and protection requirements, there is greater awareness across the organization, and not just for the FSD staff who work directly with either the research data or the administrative and organizational data. Even though there was a lot of confusion about the requirements, FSD had “previous experience with the Personal Data Act, so they reacted more rationally” (FSD03), unlike the panic experienced more generally. Lastly, the introduction of procedures to process and archive pseudonymized data will enable FSD to more readily provide access to more data to more users.

4.3 Data Archiving and Networked Services (DANS)

4.3.1 Archive description

In the Netherlands, the Data Archiving and Networked Services (DANS) is a research support service provider within the KNAW. Formally established as a KNAW service provider in 2005, DANS’ origins trace back to the Steinmetz Institute (later renamed the Steinmetz Archive) founded in 1964 (“Steinmetz Archive,” 1989). In 1972, the Steinmetz Archive, which archived Dutch quantitative social science research data, became a part of the KNAW’s Social Science Information and Documentation Centre (SWIDOC) (Social-Wetenschappelijke Informatie- en Documentiecentrum). In 1989, the Steinmetz Archive and the Low Countries
Association for History and Computing (VGI) (Vereniging voor Geschiedenis en Informatica) co-founded the Netherlands Historical Data Archive (NHDA) (Nederlands Historisch Data Archief) to pilot a computing and research support service for historical research (DANS12).

In 1995, the NHDA became an independent KNAW institute, while the Steinmetz Archive remained a part of SWIDOC. The Netherlands Institute for Scientific Information Services (NIWI) (Nederlands Instituut voor Wetenschappelijke Informatiediensten) was established as a KNAW institute in 1997, merging together five KNAW institutes, including NHDA and SWIDOC, into one institute (European Commission, 1996).142 On July 1, 2005, NIWI was disbanded and its services (including research infrastructure, databases, archives, and bibliographies) re-established elsewhere. DANS was established and took over the holdings and services of the NHDA and Steinmetz Archive, as well as the Scientific Statistical Agency NWO (WSA) (Wetenschappelijk Statistisch Agentschap NWO).

DANS has three primary services to support the Dutch research community: EASY, NARCIS,143 and DataverseNL. As the Dutch CESSDA node, DANS also participates in European-wide research and infrastructure projects, including FAIRsFAIR144 and EOSC,145 among many others.

142 The five institutes that made up NIWI in the 1997 merger were: SWIDOC, the Netherlands Bureau for Research Information, the Bureau for the Bibliography of Dutch Studies, the KNAW Library, and the Documentation and Literature Research Department of the Foundation for Educational Research.

143 NARCIS is the national research portal for the Netherlands. It provides the Dutch public a view on on-going and completed Dutch research, including information about the research, the researcher, and the research institutions. It began in 2004 and will likely shut down when the new Open Knowledge Base becomes operational (DANS, 2022c).

144 The FAIRsFAIR project (Fostering FAIR Data Practices in Europe) is a large-scale project supporting the European Open Science Cloud (EOSC). The goal of the project is to develop the knowledge infrastructure to support FAIR (Findability, Accessibility, Interoperability, and Reusability) principles-based research data management and standards. DANS is the project coordinator with 22 European data and research institutions. This project concluded in 2022, and the FAIR-IMPACT project began in June 2022. The focus of FAIR-IMPACT is to implement FAIR-based persistent identifiers, metadata, ontologies, metrics, certifications, and interoperability practices to support the social sciences and humanities, photon and neutron sciences, life sciences, and food and environmental sciences. DANS remains the leader of the project. More information can be found https://fair-impact.eu.

145 The European Open Science Cloud (EOSC) is a European Commission-funded infrastructure that aims to provide access to FAIR, open, multi-disciplinary data for reuse by researchers as well as other data users. The EOSC portal provides access to research support resources as well as to specific data archives (such as FSDs AILA, the GESIS Data Archive, etc.). DANS is a member of and service provider for EOSC. More information can be found https://eosc-portal.eu.
4.3.1.1 EASY

EASY, the Electronic Archiving System, is the DANS online archiving system through which data users can access the former NHDA and Steinmetz Archive holdings, as well as research data and data from Statistics Netherlands (CBS) (Centraal Bureau voor de Statistiek) and the e-Depot voor Dutch Archaeology (e-Depot). Developed at DANS, EASY has been in operation since January 2007 (DANS, 2008) and contains both qualitative and quantitative data, primarily in either English or Dutch, though occasionally data are deposited in other languages. This can be problematic, according to DANS14, because if the data and/or metadata are not in a language DANS is able to work in, “we can also not do any checks. We have no way of keeping up with our standards if we can’t check what’s in the metadata set.”

DANS provides disciplinary specific requirements and guidance for the historical sciences, social and behavioral sciences, language and literature studies, archaeology, and oral histories regarding documentation, metadata schemas, formats, and deposit protocols (DANS, 2021d). Data in EASY are free for users to download, and any data set under 50 GB is free for individual researchers (i.e., not contracted data providers) to deposit. If the data are larger than 50 GB, there are a number of factors that determine the deposit fee, including file format, data type, manual or automated ingest process, storage terms, consultation, etc. (DANS, 2021f).

There are two access categories possible for data deposited in EASY—Open Access and Restricted Access—though there are still a few legacy categories which will be fully phased out. All data deposited at DANS have a usage license applied at deposit. The vast majority of Open Access data (113,117 data sets as of December 31, 2021) do not require the user to register for an EASY account before downloading. These data are licensed under one of 22 possible open licenses, including the Creative Commons licenses and open software licenses. Which license is
applied is at the discretion of the data provider, including the use of the CC0 1.0 public domain license (DANS, 2021e). Data providers can also still choose to deposit under Open Access but require that the user register for an EASY account (2,965 data sets as of December 31, 2021); these data use the DANS license, instead of an open license, and is not as frequently used as before. This category will no longer be available once DANS completes the transition from EASY to thematic data stations (discussed below).

There are 5,560 data sets available (as of December 31, 2021) under the Restricted Access category. Two of these are under a legacy restriction for use only by members of a specific archaeology user group. The remaining Restricted Access data are licensed under the DANS license and are only available to registered EASY users who request and receive permission from the data provider. Data providers using the DANS license can also apply additional restrictions or requirements to these data requests at will. All data sets with personal data are automatically deposited under the Restricted Access category. Finally, there are 74,026 data sets which are archived and catalogued in EASY, but can only be accessed through a Mendeley Data, Dryad, or other institutional data portal outside of DANS (DANS, 2021a).

In total, as of December 31, 2021, EASY contains 195,668 data sets, of which 121,642 are accessible through EASY. In addition to the former-NHDA and -Steinmetz Archive holdings, the data primarily come from Dutch researchers or research institutions in the social sciences, humanities, and life sciences. Archaeology is the largest discipline archived at DANS, comprising 112,267 of the accessible data sets in EASY (111,187 of these are available Open Access to everyone without EASY registration), due to the requirement for digital archaeological data files to be submitted to the e-Depot (see Sections 4.3.1.4 and 4.3.2). There is also a large oral history collection (3,004 data sets), the majority of which are Restricted Access. This
collection covers a variety of topics, including Holocaust experiences, Dutch colonialism, religion, art, and the Yugoslav wars. In general, data sets are deposited individually through the EASY self-deposit system. Since 2017, DANS has contracts to automatically ingest data via secure machine-to-machine ftp or SWORD (Simple Web-service Offering Repository Deposit) for long-term preservation purposes (CoreTrustSeal, 2021); these data do not undergo any curation or processing, including a deposit review.

Beginning in the fall of 2021, DANS began shifting from the EASY platform to a series of topic-based “data stations” built on the Dataverse framework. There will be five data stations when the project is complete, though only four will be visible to data providers and potential data users: Data Station Archaeology, Data Station Social Sciences and Humanities, Data Station Life, Health and Medical Sciences, and Data Station Physical and Technical Sciences. As of December 2021, the DANS webpage (Fig. 4.10) has directed visitors to the data stations, though until June 2022, the user would ultimately land on the EASY platform. As of June 2022, however, the Data Station Archaeology146 is now functional (DANS, 2022e; Hollander, 2021); the other three “visible” data stations still link to the EASY platform.

146 The Data Station Archaeology data can be accessed here: https://archaeology.datastations.nl/dataverse/root.
The fifth data station will be invisible to external audiences and will be for those data from “the customers that do dissemination themselves, like the Mendeley… because Mendeley is disseminating the data themselves” (DANS13). The four publicly visible data stations will be the access points to their respective data and will also serve as a resource for researchers in a more targeted way. For example, as Data Station Archaeology is further developed, in addition to the dedicated portal for archaeological data, it will also provide archaeological thesauri, metadata fields for data deposits tailored to allow for disciplinary preferences, and other resources that the Dutch archaeological research community would find useful, instead of broader and more generic resources. New data deposits into the data stations will only be available under the Open Access (without registration) and Restricted Access categories, because the Dataverse platform does not accommodate the registration requirement. DANS is beginning this process with the archaeological data because these data are the most prevalent in EASY and will use the lessons learned from the process to support the transition for the other data stations (DANS13). Once all five data stations are fully operational, EASY will be phased out.

Figure 4.10 DANS homepage with the links to the thematic data stations, as of December 10, 2021 (DANS, 2021g).
4.3.1.2 DataverseNL

DataverseNL is a service DANS has offered since 2014 in partnership with sixteen universities and research institutions.\textsuperscript{147} DANS hosts the technical infrastructure, while the partnering institutions are responsible for the data in their own Dataverses. Only researchers at the DataverseNL-participating institutions can deposit data into one of these Dataverse. These institutions pay an annual membership fee, plus a storage fee per terabyte of storage per year, to use the service (DataverseNL, n.d.).

As of December 31, 2021, there are 500 individual Dataverses within DataverseNL associated with the sixteen partners. The subject areas represented by the various Dataverses span from astronomy and astrophysics to law, medicine, and the social sciences. The main limitation for depositing data is file size, not file type. According to DANS15, the maximum size of an individual file in a data set is currently 9.3 GB (DataverseNL, n.d.), which DANS15 acknowledges is a limitation that DataverseNL is working to solve. Data providers at the partnering institutions can set data access terms and conditions, including restricting access and requiring potential users to request access, from the dashboard page for the particular data set. When data are added to their Dataverse, the CC0 public domain license is automatically applied; the data provider can then edit the terms of use, select the “No, do not apply CC0 public domain dedication” button, and also choose specific files within the data set to restrict access to (DataverseNL, n.d.).

\textsuperscript{147} The current DataverseNL partnering institutions are: 4TU.ResearchData; the NIOO; Trimbos Instituut; Utrecht University; Vrije Universiteit Amsterdam; Hanzehogeschool Groningen University of Applied Sciences; Universiteit Leiden; Avans Hogeschool; Tilburg University; Protestantse Theologische Universiteit; Maastricht University; Rijksuniversiteit Groningen; UMC Utrecht; Rotterdam University of Applied Sciences; Fontys Hogescholen; and, the Cultural Heritage Agency.
4.3.1.3 Organizational structure

As discussed above, DANS is a research infrastructure service provider of the Royal Netherlands Academy of Arts and Sciences (KNAW). The KNAW itself comprises 11 national research institutes and one research support infrastructure, including DANS.148 The predecessor of the KNAW was founded on May 8, 1808 by King Louis Napoleon as the Royal Institute of Science, Letters and Fine Arts (Koninklijk Instituut van Wetenschappen, Letterkunde en Schoone Kunsten). In October 1851, the Royal Institute was re-established as the Royal Academy of Sciences (Koninklijke Akademie van Wetenschappen). The first institute of the Academy, the Netherlands Institute for Brain Research (Centraal Instituut voor Hersenonderzoek), now known as the Netherlands Institute for Neuroscience (Nederlands Herseninstituut), was established in 1909. Additional institutes and services have been established, disbanded, and merged throughout the following century (Klein et al., 1998).

As the parent organization, KNAW is the legal entity with overall legal responsibility for its institutes’ activities and functions. The directors of the institutes, including the DANS director, are responsible for ensuring that their organizations comply with all applicable legal requirements (Koninklijke Nederlandse Akademie van Wetenschappen, 2020). They are also authorized to sign agreements and contracts for their organizations on behalf of the KNAW. KNAW provides annual funding to DANS, and DANS’ infrastructural policies (such as IT, security, and privacy) must align with the KNAW. KNAW does not dictate DANS’ projects or operations, although DANS leadership would discuss DANS’ responses to and participation in

148 The KNAW institutes: Hubrecht Institute for Developmental Biology and Stem Cell Research; Huygens Institute for Dutch History (Huygens ING); International Institute of Social History (IISH); Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV); Meertens Institute; NIOD Institute for War, Holocaust and Genocide Studies; Netherlands Interdisciplinary Demographic Institute (NIDI); Netherlands Institute of Ecology (NIOO); Netherlands Institute for Neuroscience; Westerdijk Fungal Biodiversity Institute; Data Archiving and Networked Services (DANS); and, Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).
national and international projects, as well as changes to the organizational structure such as the transition to the data stations structure.

The Dutch Research Council (NWO) (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) is DANS’ second parent institution. It is a major research funder within the Netherlands; NWO distributes “public money for science from the Ministry of Education, Culture and Science and from almost all the other government ministries,” as well as from some commercial or civil organizations in support of specific research areas (Dutch Research Council, 2022). The NWO itself has a number of institutes and domains under its direct purview, including nine national institutes managed by the Institutes Organisation of NWO (NWO-I) (Stichting Nederlandse Wetenschappelijk Onderzoek Instituten), three research funding domains that provide targeted funding for the applied and engineering sciences, social science and humanities, and the physical sciences, and an interdisciplinary research funding domain for global development.

NWO is a co-funder of DANS, in partnership with the KNAW. As a funder and the second DANS parent institution, the NWO also does not dictate DANS’ priorities and how it spends that funding outside of specific projects for which NWO dedicates specific funding. The amount of funding from the NWO is somewhat less than what DANS receives from the KNAW (DANS01, DANS10). The NWO does, however, mandate that NWO-funded projects make their results publicly available (Dutch Research Council, 2019). At its founding, DANS had an agreement with the NWO that NWO-funded social science and humanities projects were to set

149 The NWO-I institutes are: Physics of functional complex matter (AMOLF), Advanced Research Center for Nanolithography (ARCNL), Netherlands Institute for Radio Astronomy (ASTRON), Centrum Wiskunde & Informatica (CWI), Dutch Institute for Fundamental Energy Research (DIFFER), National Institute for subatomic physics (Nikhef), Royal Netherlands Institute for Sea Research (NIOZ), Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), and Netherlands Institute for Space Research (SRON). The other NWO domains are: Domain Science (ENW), Domain Social Sciences and Humanities (SSH), Domain Applied and Engineering Sciences, the WOTRO Science for Global Development, DANS, Netherlands eScience Center (NLeSC), Netherlands Initiative for Education Research (NRO), and the Taskforce for Applied Research (NRPO-SIA).
up data contracts for the data to be deposited with DANS once the project was complete (DANS, 2015a; see also DANS12).

In terms of other oversight, DANS has a Scientific Advisory Council, the members of which are proposed by DANS and appointed by the KNAW based on their own expertise (DANS07).\textsuperscript{150} The Scientific Advisory Council advises on DANS’ strategic plans, particularly big changes such as the more intentional targeting of domains outside of the social sciences and humanities (DANS06). DataverseNL also has an advisory board made up of all sixteen DataverseNL member institutions. This advisory board makes decisions about changes to contracts and agreements, what type of persistent identifiers should be included in the DataverseNL records, and so forth.

The Dutch Data Protection Authority (AP) (Autoriteit Persoonsgegevens) is the data protection authority in the Netherlands. Its role is to enforce compliance with the GDPR and the UAVG, by supervising all personal data processors in the Netherlands, whether businesses, the government, schools, or individuals. It provides information for personal data processors, advises on potential new laws, responds to questions about privacy and data protection, and responds to reports of data breaches. The AP also issues fines to organizations which violate the GDPR; the most recent fine was on the Dutch Ministry of Foreign Affairs (Ministerie van Buitenlandse Zaken) for failing to correct known data security risks for personal data in visa applications (Autoriteit Persoonsgegevens, 2022). There is an agreement with the KNAW institutions that only the KNAW, through KNAW’s data protection officer (DPO), will contact the AP (Koninklijke Nederlandse Akademie van Wetenschappen, 2020).\textsuperscript{151} For example, if

\textsuperscript{150} As of 2022, the Scientific Advisory Council consists of six members representing the Cultural Heritage Agency, Vrije Universiteit Amsterdam, the Netherlands Institute for Social Research, CentERdata, UvA-AMC, and Statistics Netherlands.

\textsuperscript{151} This was referenced in several interviews (DANS008, DANS009, DANS011), and in the KNAW’s “De standaardmandateringsregeling KNAW” governing the relationship between the KNAW and the institutes, this mandate for the DPO is: “Aan de (plaatsvervangend) functionaris voor de gegevensbescherming is gemandateerd het melden van een datalek bij
EASY is hacked, DANS’ privacy coordinator will notify the KNAW DPO, who would then notify the AP. There is no formal communication between any institute and the AP.

4.3.1.4 Legal framework for DANS

In addition to the GDPR and the Uitvoeringswet Algemene verordening gegevensbescherming (UAVG), the Dutch GDPR implementation law, the other law which directly regulates DANS’ operations is the Heritage Act (Erfgoedwet).\footnote{Wet van 9 december 2015, houdende bundeling en aanpassing van regels op het terrein van cultureel erfgoed (Erfgoedwet), Stb. 2015, 511. (Heritage Act). Retrieved from \url{https://wetten.overheid.nl/BWBR0037521/2021-08-01}. Translation by the author.} As will be discussed in Section 4.3.2, the Heritage Act requires archaeological excavations in the Netherlands to submit their excavation reports to the Cultural Heritage Agency, and the Dutch Archaeological Quality Standard directs that the digital data files are deposited at DANS’ e-Depot for Dutch Archaeology. As a service provider of the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) (the Royal Netherlands Academy of Arts and Sciences), DANS is also obligated to follow KNAW’s policies and regulations.

4.3.2 Data archiving in the Netherlands

In the Netherlands, DANS is the most well-known of the Dutch data archives among universities and researchers. According to the Registry of Research Data Repositories, there are 69 data repositories and archives in the Netherlands covering all disciplines, 35 of which are classified as “humanities and social sciences” (re3data.org, 2022). Called the “Dutch national centre of expertise and repository for research data” (DANS, 2022a), DANS provides long-term data preservation, data dissemination, research support, and other services to researchers at Dutch universities and research institutions. Although DANS is service provider for the Royal
Netherlands Academy of Arts and Sciences (KNAW), there is no national mandate that research data be deposited at DANS (or at any other specific data archive), with the exception of archaeological data.

Since 2007, DANS and the Cultural Heritage Agency (Rijksdienst voor Cultureel Erfgoed) have collaborated to support the permanent archiving of reports and digital data files from all archaeological excavations in the country. The requirement to document and write a report on all excavations and then submit to the Cultural Heritage Agency is required under Articles 5.4(1) and 5.6(4) the Heritage Act (Erfgoedwet), and the requirement to submit the digital data files to the e-Depot for Dutch Archaeology within the EASY archive at DANS is in accordance with the Dutch Archeology Quality Standard (Kwaliteitsnorm Nederlandse Archeologie) (Stichting Infrastructuur Kwaliteitsborging Bodembeheer, 2018). While DANS is not explicitly identified in either text, it is well understood in the archaeological community that the “e-depot” discussed is DANS, since DANS is the only CoreTrustSeal-certified repository in the Netherlands for archaeological data (DANS05). Other than for these archaeological data

153 Article 5.4(1) – “Een certificaathouder zorgt ervoor dat bij het verrichten van een opgraving de verrichte handelingen en aangetroffen archeologische vondsten worden gedocumenteerd, de vondsten worden geconserveerd en een rapport wordt opgesteld waarin de resultaten van de handelingen zijn beschreven” (A certificate holder shall ensure that when an excavation is carried out, the conducted operations and archaeological finds are documented, the finds are conserved, and a report describing the results of the actions is prepared) (Translated by the author).

154 Beoordelingsrichtlijn: Archeologie – “Alle digitaal aangeleverde projectdocumentatie (waaronder standaardrapporten) wordt, conform protocol 4010 Depotbeheer, duurzaam bewaard voor later archeologisch onderzoek. Dit geschiedt in een e-depot” (All digitally delivered project documentation (including standard reports) is permanently stored for future archaeological research, in accordance with Protocol 4010 Depot Management. This occurs in an e-depot) (Stichting Infrastructuur Kwaliteitsborging Bodembeheer, 2018, p. 20) (Translated by the author).
files (which can result from academic, commercial, or coincidental excavations and finds), no disciplines are explicitly required to share their research data, but organizations like the Universities of the Netherlands (UvN) (formerly the Association of Dutch Universities (VSNU)\(^\text{155}\)) have developed codes of conduct for research that emphasize transparency about data collection, why research and/or data cannot be made available, effective and secure data management, etc. (Koninklijke Nederlandse Akademie van Wetenschappen et al., 2018).

### 4.3.3 Privacy and the GDPR in the Netherlands

There were a number of reactions in the Netherlands to the GDPR reported by the DANS interviewees as well as several post-GDPR studies and general news media coverage. Common reactions (e.g., panic, anxiety, frustration) intersect in interesting ways with Dutch views of privacy, especially pre-GDPR views of privacy. In Europe, the Netherlands has a reputation for being a very open society—known, for example, for its toleration of drugs, legal prostitution, minority rights, and equality (all of which have come under scrutiny in recent decades (see Coates, 2015; Gordijn, 2001; Pruijt, 2013)—unconcerned with surveillance and outside intrusion into the private sphere of the home (Mols & Janssen, 2017; Van Der Horst & Messing, 2006; Vera, 1989), even as it has a well-established tradition of collecting personal information. For over 800 years, the region that is now the Netherlands has had ever evolving requirements for registering personal information (beginning with records of annual Catholic confession as of 1125, through the post-Protestant Reformation baptismal, marriage, church membership, and burial registrations, national identity cards, and so forth) (Ketelaar, 2020). The Netherlands has never had a society filled with CCTV cameras (such as in the United Kingdom) or neighbors

\(^{155}\) At the end of November 2021, VSNU (Vereniging van Universiteiten) announced that it was changing its name from the Association of Dutch Universities to Universities of the Netherlands (UNL) (Universiteiten van Nederland) (Het Hoger Onderwijs Persbureau, 2021; Universiteiten van Nederland, 2021).
reporting on neighbors (such as in East Germany or the Soviet Bloc) (Ketelaar, 2020; Mols & Janssen, 2017). Therefore, while the Constitution of the Netherlands (Grondwet voor het Koninkrijk der Nederlanden)\textsuperscript{156} includes a right to privacy (Art. 10), this history paired with the open society has given the Netherlands the reputation of not being too concerned with privacy \textit{as long as} the norms around that openness are adhered to.

This is exemplified in the title of Van Der Horst and Messing’s (2006) study of the boundary between public and private in a multi-ethnic neighborhood in The Hague: “‘It's not Dutch to close the curtains’: Visual struggles on the threshold between public and private in a multi-ethnic Dutch neighborhood.” Dutch windows, whether unlit during the day or lit at night, with their curtains open, provide a view into the private sphere of the home. However, both Van Der Horst and Messing (2006) and Vera (1989) make clear that there are expectations about why and how an observer should observe the goings on inside: For example,

look but don’t look, even though the open window is a blatant invitation to exchange information. … One was not supposed to turn the head sideways, but rather glance imperceptibly. By no means should one stand squarely in front of a window (Vera, 1989, p. 223).

The idea of actually, actively looking into the homes on the other side of the open curtains is anathema and violating the rules about the windows is treated as a hostile invasion of the private sphere. Thus, the reputation for openness does not mean that Dutch people, or Dutch society writ large, do not care about privacy.

In the aftermath of the 2013 Edward Snowden leaks, Mols and Janssen (2017) evaluated Dutch journalistic and non-journalistic publications (off- and online) responding to the leaks in the subsequent two weeks (June 6-20, 2013). They identified six ways the Dutch public discussions about the leaks were framed: “end justifies the mean” (grounded in trust in the government); “nothing to hide” (grounded in the open culture that has nothing to hide); “privacy paradox” (grounded in the fear of being left out of the digital reality); “empower the user” (grounded in the idea that people should control their own information); “privacy is dead” (grounded in the feeling that there is no hope for a future with privacy); and, “Orwellian dystopia” (grounded in deep mistrust in the government). Of these, the most frequent way the Dutch discussions were framed was as a call to “empower the user,” especially through the use of regulations (22% of the publications), though the “Orwellian dystopia” (20%), “end justifies the means” (18%), and “privacy is dead” (18%) frames were not far behind. Countering the idea that privacy is unimportant in the Netherlands because Dutch people have nothing to hide, that frame was only identified and promoted in 7% of the publications.

In addition to these two studies of Dutch windows and the post-Snowden analysis, the 2015 Eurobarometer 83.1 survey (European Commission, Brussels, 2018; see also Appendix F) asked respondents from all EU Member States about their online behaviors, concerns over personal information, and what they felt were the biggest risks to providing personal information online. Unlike the two windows studies, Eurobarometer 83.1 focused on the digital sphere, not

---

157 Beginning in June 2013, former National Security Agency (NSA) contractor Edward Snowden released more than 1.7 million classified documents that revealed the extent of the (primarily but not exclusively) United States and Western European intelligence gathering and sharing, not just on “known” threats to national security but on ordinary citizens. The documents also revealed the role that telecommunications companies played in granting the intelligence agencies access to their networks and intercepting electronic messages. In February 2014, the Dutch Interior Minister, Ronald Plasterk, publicly acknowledged that he lied to the public about which agency had collected the data about Dutch citizens; instead of the NSA, the Dutch intelligence service Algemene Inlichtingen- en Veiligheidsdienst (AIVD) collected those data and shared them with the NSA (Tartwijk, 2014).
the private sphere of the home, and reflects the opinions of respondents a year before the GDPR was formally adopted. In terms of feeling in control of their information online, the vast majority of Dutch respondents felt they had only partial control (58.58%), while only 8.74% felt in complete control. However, at the same time, less than half felt concerned about this limited control (49.46%).

When asked about whether explicit consent should be obtained before collecting or processing personal data, there were three possible “yes” answers, and multiple responses were possible: “yes, in all cases” (58.23%), “yes, in case of personal information required online” (25.69%), and “yes, in the case of sensitive information whether online or offline” (17.39%). The spread of “yes” responses varies on the EU level, but in all EU Member States, more than half said that consent should be required in all cases. The EU average for consent in all cases was 70.21%, so the Netherlands’ respondents were well below the EU average.158

Before the GDPR, thus, we can see that, even with the reputation for openness, in allowing for the possible view into the private sphere, the Dutch were not blasé about their privacy, particularly regarding digital and information privacy. Even the openness of the windows is changing with the changing demographics of Dutch cities (Van Der Horst & Messing, 2006) and changing technologies (Mols & Janssen, 2017). The DANS interviewees described their thoughts on the “general” Dutch reactions to the GDPR in two main ways: “One was of a group of people who said we should have more control of our data, and one was of a group of people who said, ‘if you have got nothing to hide, what does privacy mean?’” (DANS08). On the one hand, the DANS respondents reported an increased awareness in Dutch

---

158 For claims that involve contrasting rates between one of the four countries of interest and the EU excluding the four countries of interest, the smallest sample sizes were 588 (single country) and 14406 (remainder of EU). In this worst-case scenario, the standard deviation of the difference of two percentages is no more than 2.1%, yielding a 95% margin of error of 4.2%. Thus, any difference between two percentages reported here that exceeds 4.2 percentage points is statistically significant.
society of the GDPR and the rights of data subjects to have more control of their personal data.

For example,

I think it's a really good thing that we have this. It's a really good thing that it raises awareness, not just on the level of policymakers, not just on the level of executive directors of companies, not just on the level of the people who make money for it, but on the level of the citizen walking on the streets, ‘why is that camera there?’ Or, ‘why are you tracking my cell phone?’ It's a basic right, I think. And that's why the GDPR in my eyes is a monster. But it's an important monster, I think (DANS01).

This aligns with what three post-GDPR studies have found.

The Special Eurobarometer 487a (2021), conducted in March 2019 (see Appendix G), found that 60.37% of Dutch respondents had both heard of the GDPR and knew what it meant at a much higher rate than the EU average (35.85%). It also found that more than half of Dutch respondents knew about the rights of access (86.13%), objection to marketing (80.63%), rectification (80.04%), to be forgotten (76.79%), and portability (59.49%), while less than half (42.67%) knew about the right not to be subject to automated decision making. Of these, already by March 2019, more respondents (41.79%) had already exercised their right to object to direct marketing than simply knew about the right; for the rest, most respondents knew of the rights but had not yet exercised them. The Special Eurobarometer 487a also found that the percentage of Dutch respondents feeling partially or completely in control of their online information had increased from the 2015 Eurobarometer 83.1, from 52.58% to 62.13% (partially) and 8.74% to 9.62% (completely). There was a smaller decrease in concern about the limited control, from 49.46% in 2015 to 47.84% in 2019 (European Commission, Brussels, 2018, 2021)

Deloitte’s (2018) survey of eleven countries (including the Netherlands),¹⁵⁹ while not providing a country-by-country break down for all questions in the report, also looked at how

---

¹⁵⁹ The countries surveyed were the United Kingdom, the United States, Australia, Canada, India, Spain, Italy, the Netherlands, Germany, France, and Sweden.
familiar respondents were with the GDPR’s rights of the data subject. Deloitte found similarly high levels of awareness, with the Netherlands and France the most aware of the EU countries of the right to data portability (83% for both, compared with 81% for the surveyed EU countries overall) (Deloitte, 2018, p. 13). The Netherlands also had the highest awareness of the right to be forgotten of all surveyed countries (83%) (Deloitte, 2018, p. 13). The third study, by Strycharz et al. focused exclusively on reactions in the Netherlands and similarly found that, during data collection in 2019, 79% of respondents were aware of the GDPR (2020, p. 414). Familiarity was high as well for the rights of access (66%), rectification (63%), and objection (81%), but comparatively low for the right to be forgotten (32%) and the right to not be subject to automated decision making (16%) (Strycharz et al., 2020, pp. 413–414). If nothing else, Dutch data subjects were overall aware of the rights they had to control their personal data.

The other reaction, that privacy does not mean anything if you have nothing to hide, was also reported. This was highlighted by other interviewees:

Most people in the Netherlands are not aware. They don't even think about it, they don't have a problem with giving up their data and hardly think about it and are not concerned with privacy issues at all (DANS07).

I think, in general, the response was that it was all such a fuss and that you couldn't do anything anymore. There were two takes on it, probably. And then there was people that say, ‘Well, I don't have anything to hide, so, I don't mind’. But I think everybody minds in the end. Yeah, so those were pretty much the two perspectives (DANS14).

Related to this, Strycharz et al. also found that a small percentage of respondents reported negative impacts of the GDPR on them professionally (14%) and personally (13%) (2020, p. 419), such as complicating taking photographs at school events and operating small clubs and organizations; they thought that the GDPR went too far in scope. These same frustrations were expressed anecdotally by DANS04 (regarding sporting photographs of children’s teams) and
DANS03 (regarding collecting personal data for non-DANS groups). Newspapers also frequently reported on the ways the GDPR has limited or prevented common daily life events, such as publishing lists of graduates (Bouma, 2018), the use of dating websites and applications (Techredactie, 2019), or tracing family histories (Schapendonk, 2018). From the perspective of not worrying about privacy when there is nothing to hide, these examples highlight the very aspects of Dutch life that had previously been “open” which are now (actually or presumably) closed because of the GDPR.

There were several specific responses to the GDPR in the wider Netherlands, as reported by the interviewees, which also applied to how they felt about the GDPR’s effects on their work (which will be discussed in later sections). There was a general feeling of panic by organizations, especially smaller ones which did not know whether the GDPR’s requirements would apply to them (see Section 4.3.6 for further discussion of the impact of the legislative delays in the Netherlands), and then struggled with their implementations. For example,

… what in a sense had some impact was that so many people were concerned that they could, outside of DANS, were concerned, they could not do their work properly. And that was going to be too restrictive, and that we didn't know what it would mean for the Netherlands (DANS06).

Panic. I think a lot of people were afraid there would be, it would mean a lot to the way they were working and change a lot, and that they didn’t have a good handle on how to keep up with what was expected of them. And maybe, even much fear of doing something that would certainly be illegal or intrusive. I do think I felt panic. Lots of people did (DANS05).

There was a lot of stress, because people felt that the GDPR (and later, the UAVG) would be too different from the previous Personal Data Protection Act (DANS02). That so much was relying on the decisions the UAVG would make regarding derogations from the GDPR’s restrictions only increased the panic and stress, as evidenced by reports of organizations which were not GDPR compliant by May 2018, whether governmental agencies (e.g., the Tax and Customs
Administration (Belastingdienst) (Belastingdienst, 2018) and the Employee Insurance Agency (Uitvoeringsinstituut Werknemersverzekeringen) (Autoriteit Persoonsgegevens, 2018a, 2018b)), hospitals (e.g., Haga Hospital (HagaZiekenhuis) (Autoriteit Persoonsgegevens, 2019a, 2019c)), or, as will be discussed in depth in this report, DANS itself.

Finally, there was a significant amount of frustration with the GDPR as a bureaucratic imposition. As DANS04 put it, “Well, we laugh about, but the bureaucratic… it's more work, more rules, and it usually does not get a lot of applause from people in the Netherlands.” The sudden bureaucratic impositions on everything from school photographs (DANS04, Strycharz et al., 2020), to trying to put a book on hold at a library (Strycharz et al., 2020), to having to constantly accept cookies for everything were bad enough (interestingly, in 2019, the AP declared that tracking cookies on websites violated the GDPR (Autoriteit Persoonsgegevens, 2019b)). Nearly a quarter of respondents in the Strycharz et al. (2020) study felt that the GDPR was imposed on them without any (or enough) citizen agency into the decision. Considering that the Treaty on European Union states that “Every citizen shall have the right to participate in the democratic life of the Union. Decisions shall be taken as openly and as closely as possible to the citizen” (Art. 10(3)), the top-down development and imposition of the GDPR resulted in apparent burdens that seemed to outweigh the potential benefits, especially when people were relying on other parties for their own protection. With the possibility of changes coming to the UAVG, as well as the potential adoption of the new ePrivacy Regulation, which would replace the Privacy and Electronic Communications Directive 2002/58/EC,\textsuperscript{160} it remains to be seen how people in the Netherlands perceive the benefits and downsides of the GDPR.

4.3.4 Timeline of GDPR response

In the months after the GDPR was adopted, both the KNAW and DANS began preparing their responses. DANS began by reviewing the services and data at DANS, while the KNAW began developing the process register (the GDPR-required database where all KNAW institutes’ personal data processing practices must be registered). KNAW’s DPO began working with the different institutes to complete data protection impact assessments (DPIAs). On May 16, 2018, the UAVG, the Dutch GDPR implementation legislation, was officially signed into law, and the new “KNAW Privacy Statement” (Koninklijke Nederlandse Akademi van Wetenschappen, 2018) entered into force. Just over a week later, the GDPR itself became enforceable.

In mid-June, DANS’ new “Privacy Declaration” (DANS, 2018) entered into force, the first formal change that DANS implemented. At the end of August 2018, the DPIAs for EASY, NARCIS, and DataverseNL (Alfons, 2018c, 2018b, 2018a) were completed, formally establishing the priorities for the changes that DANS needed to make to be GDPR compliant. The GDPR-compliant deposit (DANS, 2019c) and processing agreements (DANS, 2019d) came into use at the end of September 2019. As of January 9, 2020, several additional changes are in effect: the “DANS Licence” (DANS, 2020a) for the Open Access for registered users and Restricted Access data, the ability to use the Creative Commons licenses in addition to the other open licenses, and the GDPR-compliant “EASY General Terms and Conditions of Use” (DANS, 2020b). In spite of these efforts, according to all DANS interviewees, in 2022, DANS is still not fully compliant with the GDPR.

4.3.5 Priorities in the response

In the interviews, many DANS respondents framed their views of DANS’ reaction to and efforts towards GDPR compliance as protecting data subjects’ rights and privacy, whether they
were discussing the research data in EASY or the administrative and operational data about users and staff. For example:

I think, in my sense, the most important implication of the GDPR on DANS was safeguarding the rights of the data subjects. And we essentially needed to change the policy regarding access categories to be able to safeguard the rights of the data subject (DANS02).

So, we ourselves, are being more aware of the data that we process, and also how we should do that in a GDPR compliant manner. So, for instance, when we are collecting information for a workshop, where people have to register, we're now also thinking much more about ‘Okay, we make sure that there's a GDPR you know, that people are given consent that we can use their data for this’ (DANS10).

As will be discussed in the following sections, many of the changes made were made in anticipation of data subjects exercising their rights under the GDPR, whether they were changes to the data acquisition process or deciding who was ultimately responsible for ensuring the data are protected.

4.3.5.1 Priorities and goals development

In the time between 2016 and 2018, DANS did begin preparing for what it expected to need to change, by beginning its own review of its data, processes, and services. However, until the UAVG was signed in late May 2018 and the KNAW understood what it needed to do to be compliant with both the GDPR and the UAVG, DANS was not able to fully set its own priorities for the transition. The priorities and plans for changes ultimately were finalized through the fall 2018 DPIAs:

From that PIA [privacy impact assessment] or DPIA came some issues and that was different from every service that we need to get in order or change or whatever. That's how we work on the necessary changes. As soon as this whole thing, GDPR, et cetera, came into being, that's what I recall at least, we did the PIA quite soon after it started (DANS13).
The priorities that came out of the DPIAs were: addressing potential security concerns, creating a new data acquisitions process, raising DANS staff awareness about privacy and data protection, and settling the question about whether DANS would be processor or controller (Alfons, 2018c, 2018b, 2018a). The DANS staff most involved in the planning, supervising, and implementing GDPR-related changes were the legal advisor (who is also one of the EASY research data managers), the privacy coordinator (part of the administrative department), the management team (including the director, vice directors, and department heads), and the head of the archive, with external support from the KNAW’s data protection officer and security officer.

4.3.5.2 Processor versus controller

The obligations and responsibilities imposed by the GDPR differ significantly between processors and controllers. One of the first decisions DANS needed to make was whether, and for what services, it would be the processor or controller. This decision influenced the later changes to policies and practices throughout the organization. For the research data deposited in EASY, DANS is the processor. This decision continues the pre-GDPR practice where it was the data provider who decided who, and under what conditions, could access data through DANS. However, it was not a simple decision, and according to DANS11, was one area where DANS and KNAW disagreed:

… for instance, that we are a processor. The KNAW still thinks that, ‘Oh, we could have become a controller’, I think, and easier in certain respects. But we, of course, discussed that with the lawyers already at that time that we wanted to do this and that. They just followed that. They had some suggestions, but then we thought, ‘well, we want to do it this way’ (DANS11).

As will be discussed in Section 4.3.6, however, whether DANS should continue as the processor for the archived data is still under discussion.
The data in EASY include data sets with personal data. Before the GDPR, some of these data were available Open Access and/or included data subject personal data in the metadata or file names, such as for the oral histories. While this policy allowing some data sets with personal data to be Open Access has changed (Section 4.3.5.4), the decision to be the processor for these data did not influence the policy change, or vice versa. On the other hand, the question of who was responsible for authorizing access to Restricted Access data did influence the decision to be processor. As DANS14 put it,

DANS doesn't want to be involved in that. And that's also one of the reasons why to choose the data controller or data processor role data, because we don't want to be involved in that. Nevertheless, I mean, we have a certain responsibility, but we don't want to be the person deciding on that. That's something the depositor should take care of.

This choice thus puts much of the GDPR burden on the data provider:

They, of course, need to make sure that their data set complies with certain conditions that we set, which basically comes down to that everything has to be anonymous, and it has to be restricted access. You have to make sure that your files and everything, and file names, are living up to that, so that people can make sure, ‘Okay, this data is safe. It's behind the wall’ (DANS11).

DANS is also the processor for the data in DataverseNL. The partnering institutions are the controllers for the data in the institutional Dataverse. Because the data are archived with the DANS-hosted platform, however, after the GDPR, there must be a processing agreement signed between DANS and each institution because archiving is a form of data processing (Art. 4(2) GDPR).

Outside of these services, DANS is the controller for two other types of data. DANS is the controller for the data produced in many (but not all) of the DANS-led or -involved projects while the projects are on-going (such as the FAIRsFAIR project). When FAIRsFAIR is complete, these data currently stored and shared through the project’s Zenodo community will be
deposited into the DANS data archive. As will be discussed further in Section 4.3.5.7, this dual role of DANS-as-controller and DANS-as-processor for research data at different project stages is one of the reasons why the decision to be the processor for data in the archive is still being debated. Lastly, DANS is also the controller for the administrative and operational data needed to fulfill its role (including the personal data from EASY registrations, events and trainings DANS offers, personal data included in the NARCIS portal, and employee data).

4.3.5.3 Regulating information flows

DANS’ decision to remain the processor of the EASY data directly affects the way the processing, sharing, and archiving of personal data at DANS is regulated. Clarifying the roles and responsibilities of the different parties in ensuring that the privacy of data subjects is protected during these data transfers in a GDPR and UAVG compliant manner is the focus of DANS’s efforts in this area. The information flows of interest here are the sharing of personal data by the data provider to DANS, the processing of personal data (any curation work as well as the archiving itself) by DANS, and the decision about who can access the data and under what conditions access may be granted. In EASY, the changes to control the information flows are seen in the agreements signed between the data provider and DANS, and those agreed to by the data user; the contracts for DataverseNL were revised as well. These changes will be discussed below.

The acquisition process begins with a registered EASY user submitting the “Deposit Form” online (DANS, 2021c). Through this form, the data provider uploads the data files, information about the creator(s), the project description and metadata, chooses the appropriate license, identifies the rightsholder, identifies the type and format of the data, and agrees to the deposit agreement. The changes to the usage licenses will be discussed in Section 4.3.5.4. Much
of the “Deposit Form” is unchanged from earlier versions; however, there is one change that triggers an additional, new agreement, and the automatic application of a specific access license. The second module of the “Deposit Form” now requires the data provider to respond to the question, “Does your data set contain personal data? Yes/No” (Fig. 4.11). If the data provider responds “yes,” the access category is automatically set to Restricted Access, the data can only be licensed under the “DANS Licence,” and the data provider and DANS must conclude the “Processing Agreement” (to be discussed later in this section). This question was added to the deposit form because

… we should know that this data set includes personal data. And that's also something for the depositor, that the depositor has to be aware that when he deposits, the first question he has to answer is, ‘does your data contain personal data?’ Because he will be also responsible, because for EASY, we're only the processor, the depositor will be responsible. But you have to make sure that the depositor knows that and that he also answers the question, formally, that this data set does not contain some personal data (DANS04).

![PERSONAL DATA](image)

*Figure 4.11 Question about presence of personal data in data deposit, added to the DANS deposit form in response to the GDPR (DANS, 2021c).*

The actual agreement to deposit data was more significantly revised. The previous agreement, the “Licence Agreement” (DANS, 2016), identified DANS’ right to archive and disseminate the deposited data sets, that the data provider was the rightsholder, and explicitly indemnified DANS against liability for the data. The new “Deposit Agreement” (Deponeerovereenkomst) (DANS, 2019c) retains the basic structure of the “Licence Agreement,”
with revisions that further align the document with the GDPR, but includes the requirement for the “Processing Agreement” in specific situations:

If the Data set contains personal data, the Depositor and the Depositary will conclude a processing agreement, except in the case of bibliographical data which exclusively refer to personal data that are necessary for the accountability of the Data set, such as its creator, rights holders and citations (hereinafter: “Bibliographical Data”) (§3.4) (DANS, 2019c, p. 3)

The “Processing Agreement” (DANS, 2019d) is required because under Article 30 GDPR, all personal data processing is to be documented in a process register; DANS documents this through these agreements.

Additional changes to the “Deposit Agreement” and DANS’ EASY operations reflecting GDPR compliance relate to the metadata, the license applied to data sets with personal data, explicit obligations for the Restricted Access data users, and limitations on personal data in the metadata or file names related to the research subjects (§4.4). Under the “Licence Agreement,” nothing was specifically said about the content of or inclusion of personal data in the metadata record. While the new restrictions on personal data in metadata records only applies to new data deposits, DANS is currently addressing how this can be applied to the data already in EASY; in particular, with the oral history collections, which by their very nature include the data subject’s personal data (i.e., name) in file names and the metadata record.

The five access categories listed in the “Licence Agreement” (Open Access, Open Access for Registered Users, Restricted Access, and Other Access) have been changed in accordance with DANS’ interpretation of the GDPR requirements. This change will be discussed further in Section 4.3.5.4. The aspect of this change relevant to this discussion is that, per the “Deposit Agreement,” “If the files in the Data set, or parts of them, contain personal data within the meaning of the GDPR, the Restricted Access category will be applied exclusively and the files
will not be made available directly to third parties” (DANS, 2019c). The “Licence Agreement” did not automatically mandate the Restricted Access category for these data. Under the new policy, transfers of all new data sets with personal data to data users (including oral histories and any re-categorized data), and how the data can be reused, are controlled by the data provider. Beyond facilitating communication between the data provider and the potential data user, per the “Deposit Agreement,” DANS “cannot be held responsible for the Depositor’s decision whether or not to make the Data set available, nor for any conditions under which this is done” (DANS, 2019c).

The “Processing agreement sustained archiving” (hereafter, “Processing Agreement”) (Verwerkersovereenkomst duurzame archivering) (DANS, 2019d) must be concluded between DANS and the data provider for all data sets with personal data. It does not apply to personal data only included in the metadata or in citations, such as the “creator” field or the names of the research team. If the presence of personal data in the data set is indicated by the “Yes” response in the “Deposit Form,” then DANS will contact the data provider to conclude this agreement. Similarly, if the answer is “No” but personal data are found during the deposit review, then processing will stop, and DANS will contact the data provider about the “Processing Agreement.” Once the agreement is concluded, it is registered with the KNAW’s process register (DANS04).

The purpose of the “Processing Agreement” is to establish the specifics of what the personal data are, who the data subjects are, who at DANS will receive and work with the data, and what processing will be done. The “Processing Agreement” specifies that the data provider-as-Controller has determined “which data are necessary and will ensure that the Personal Data in question are correct, sufficient and not excessive in accordance with Article 5 of the GDPR”
DANS is not responsible for ensuring which legal bases for processing apply; the “Processing Agreement” firmly establishes this as the Controller’s obligation. The “Processing Agreement” also sets the security expectations and data leak notifications, and again indemnifies DANS against liability (outside of gross negligence). This is important because by adopting the processing role, while DANS is still required by the GDPR to protect data subject privacy and process the data lawfully, it is not DANS’ responsibility to prove that everything is done in accordance with the GDPR.

DANS is responsible for processing in accordance with the “instructions of the Controller and under the explicit (final) responsibility of the Controller” (DANS, 2019c). If the data provider and DANS cannot agree on the processing plan, then the data must be fully anonymized or they cannot be deposited at DANS. The “Processing Agreement” also applies to pseudonymized data; these data must be treated as if they were a data set with personal data as long as the data provider retains the key connecting the data subject to their data. Several DANS and KNAW respondents (DANS05, DANS06, DANS07, DANS08, DANS09, and DANS11) consider the creation and implementation of the “Processing Agreement” to be one of the most significant changes that DANS made in response to the GDPR.

EASY was not the only service for which DANS determined that a processing agreement was necessary; DataverseNL also now requires one to be concluded with the partnering institutions along with the initial “Collaboration Agreement” (DANS, 2022b). The reason for this is because “archiving” is a type of processing under GDPR’s purview.

The data are at our site. We don’t do the curation, but the data are stored in the environment we use for storing data. So that’s why… ’It's because the data are stored in our infrastructure, therefore we need such arrangements (DANS15).
The DataverseNL processing agreement (Verwerkersovereenkomst DataverseNL) (hereafter, the “Processing Agreement DataverseNL” (DANS, 2022d)\textsuperscript{161}) is similar the EASY “Processing Agreement” but has several important differences which reflect the different missions of the two services.

For example, the “Processing Agreement DataverseNL” does not including a reference to retaining personal data for processing in the public interest or for research purposes, while the “Processing Agreement” does. The two other significant differences relate to DANS’ financial responsibilities and liability. First, both agreements discuss the potential for an independent IT system audit that a controller (data provider) can request at any time to evaluate DANS’ data security measures. For DataverseNL, if the audit finds that DANS failed to comply with the “Processing Agreement DataverseNL,” the GDPR, or any other applicable laws, DANS is responsible for the cost (otherwise, the controller remains responsible) (DANS, 2022d), while the controller is responsible regarding EASY. Second, for DataverseNL, however, DANS is explicitly liable for all damages attributed to some failure or shortcoming, whether there was intent or gross negligence involved or not (DANS, 2022d). All signed “Processing Agreements” and the “Processing Agreement DataverseNLs,” along with all other DANS personal data processing, are recorded with the KNAW Process Register, in accordance with Art. 30 GDPR.

The other set of policies which DANS revised relating to regulating the information flows into and out of DANS are the terms and conditions applicable to DANS data users. Before the GDPR, the General Conditions of Use (hereafter, “General Conditions”) (Algemene gebruiksvoorwaarden) (DANS, 2015b) were the primary guidance for users of all non-CC0

\textsuperscript{161}Minor changes have been made to the text over time since the first Verwerkersovereenkomst DataverseNL was put into operation. I have reviewed several iterations of this agreement and will refer from here on out to the most recent version, as the changes in the text do not substantially alter the aspects of interest in this dissertation.
(public use waiver) data. It applied only to users who downloaded these data from DANS, and its main purpose was to direct what users could and should do with the data (DANS, 2015). After the GDPR, the “EASY General terms and conditions of use” (hereafter, “EASY Terms”) (Algemene gebruiksvoorwaarden EASY) (DANS, 2020b) came into effect. However, these were not a direct replacement. Unlike the “General Conditions,” the “EASY Terms” apply to all EASY account holders, including data providers, and explicitly define the user’s responsibility for the accuracy of their own personal data in their EASY account (DANS, 2020b). It also describes: the four access categories still used in EASY; which licenses apply to which access categories; how the data providers personal data are processed; and, what security measures are used.

The “General Conditions” and “EASY Terms” are thus very different in scope and application; in fact, many of the components of the General Conditions are instead found in the “DANS Licence” (DANS-licentie) (DANS, 2020a), though modified to reflect the GDPR. Unlike the “General Conditions,” the “DANS Licence” explicitly states that it only applies to the Open Access for registered users and Restricted Access data, not just “any data” available in EASY. The Open Access licenses cover the remaining data not covered by the “DANS Licence.” Beyond the obligations and restrictions for data users in the “DANS Licence,” DANS does not dictate what users do with the data they have or are given access to; that is the responsibility of the data provider. The licenses will be discussed further in Section 4.3.5.4.

4.3.5.4 Minimizing harm

Many of the changes DANS made were intended to minimize the harms that may result from misuse of personal data. These range from reconsidering how personal data can be accessed, revising and expanding the privacy policy, and ensuring that consent is given before
personal data are collected and processed in EASY or for the administrative and operational data. As introduced in the previous section, an important change evidenced in the “Deposit Agreement” and “EASY Terms” is the reduction of the five pre-GDPR access categories (Open Access, Open Access for registered users, Restricted Access ‘archaeology’ group, Restricted Access, and Other Access) (DANS, 2016) to two (Open Access and Restricted Access) (DANS, 2019c, 2020b).

Two of the previous categories (Open Access for registered users and Restricted Access ‘archaeology’ group) remain legacy categories in EASY and will be phased out in the full transition to the data stations. Doing so requires that data providers agree to adopt a new or different license for their already-deposited data sets; DANS is currently negotiating this with the data providers. Many have agreed to do so via email since the new categories came into effect on January 9, 2020, often by changing to one of the CC-BY licenses (DANS05). For example, in May 2019, there were 9,082 data sets in the Restricted Access ‘archaeology’ group; as of December 31, 2021, there are only two. Similarly, the Open Access for registered users contained 29,591 data sets in May 2019, while by the end of 2021, there are only 2,965. In this latter category, there are still some new data sets being deposited into the category, many from on-going archaeological investigations, but overall, this category’s use is decreasing.

This change occurred in conjunction with the new license scheme, which also came into effect on January 9, 2020. According to DANS06, the GDPR and the need to review DANS practices and processes provided a good opportunity for us to revise that. To make sure that we could comply with GDPR, but also to provide a more clear service to people who wanted to deposit with us. And that was also, for us, the moment to implement the Creative Commons family of licensing … we noticed that they were useful, and people seem to like them and seem to understand them, which is, of course, good, because then we don't have to explain our own alternatives.
Previously, the Open Access category meant that all data were only licensed under the CC0 (public use waiver) and was the only category into which personal data could not be deposited. According to DANS02, the purpose of the Open Access category which did not require EASY registration was to be able to use and disseminate data in the public domain. Open Access under the new scheme has expanded the possible open licenses to 22, including CC0 and the other Creative Commons and open software licenses. This expansion in possible licenses was motivated in large part by the data providers themselves:

… several former licenses were already asked about by depositors for many years. So, we would often get questioned why we didn't support CC-BY, for example. And some of the old access categories, they were very specific to our archive. And it's also limited us sometimes from making the data in our archive accessible via international portals. … And we always strove to make as much data as possible available on some kind of open access category. So, moving to these formal licenses was only the logical, natural choice to do (DANS05).

The Open Access licenses also support machine-to-machine access to the data supporting the various EU projects, which had been a challenge when the non-CC0 Open Access data required an EASY account login. The challenge DANS and the EU project teams faced was for computer, machine to machine interfaces to enable access to our content. That access needed to have a direct handle, direct link with no interference from any access restrictions, like the necessity to login. So, that could be a big problem in the past, and now with formal licenses, it's much easier (DANS05).

For those increasingly uncommon data sets deposited in the Open Access for registered users category, as well as the Restricted Access category, the “DANS Licence” (2020a) applies, whether there are personal data in the data set or not. Because the “DANS Licence” applies to the data sets with personal data, it sets out the user’s obligations for maintaining and prioritizing the privacy of the personal data (DANS, 2020a).
With the new access categories and license scheme, DANS now automatically restricts the accessibility of data sets with personal data to the Restricted Access category. Before this policy change, as long as the data subjects gave their informed consent to have their personal data made available Open Access, DANS could make the data available in this way (DANS05). Now, even if the data subjects consent to it, DANS will only make these data available with the Restricted Access category and “DANS Licence” in place. This is one area where DANS05 argues that the GDPR requirements (or at least, how DANS has chosen to interpret and implement them) goes too far, by restricting the reusability of certain data:

I sometimes wonder indeed if it's really necessary to have every data set with personal data set to restrict access, even if everybody involved with it agreed to open access, so, yes. … Because I like having things directly accessible by setting up open access. Yeah. Sometimes, I've seen examples where people just weren't allowed to, because there was some personal data in it, and everybody would have a gate to open access, and... then I think GDPR might go a bit too far there.

The data provider can also include additional restrictions or requirements the user must meet to be granted access to the Restricted Access data. Putting all the personal data under one category and one license is also more supportive of responding to data subjects’ invocation of their rights under GDPR, such as erasure, by enabling DANS to identify just who has already downloaded the data:

… you're quite forced into one particular access category with a corresponding license and yeah. The main reason for that was to be able to have a data subject say 'I want to have my permission revoked’, and therefore to accommodate and delete personal data we have to know where the data is, essentially. We can only do that if we know what users downloaded the data (DANS02).

DANS revised the “Privacy Declaration” (Privacyverklaring) (DANS, 2018) only after the “KNAW Privacy Statement” (Privacyverklaring (KNAW)) (Koninklijke Nederlandse Akademie van Wetenschappen, 2018) entered into force. It replaced the previous DANS Privacy
Regulation (hereafter, “Privacy Regulation”) (Privacyreglement DANS) (DANS, 2009). The text of the “Privacy Declaration” draws heavily from the “KNAW Privacy Statement,” which applies to all KNAW institutes.

The 2018 “Privacy Declaration” is more detailed than the “Privacy Regulation,” in large part because it replicates much of the text of the “KNAW Privacy Statement” while also tailoring it to fit DANS’ needs and operations. It is focused almost entirely on the processing DANS does with the administrative and operational data used for EASY, NARCIS (the national Dutch research informational portal), and DataverseNL. For each of these services, the “Privacy Declaration” gives the legal bases for processing, data subject rights, and which personal data are processed and why (DANS, 2018). Personal data in the research data sets in EASY are referenced only under the description of EASY; here, the “Privacy Declaration” reiterates that DANS is the processor for the EASY data and directs the data subject to contact the controller with concerns or objections.

The 2009 “Privacy Regulation” was the only document to state that, for data sets containing the Directive’s Special Categories of personal data (i.e., about religion, race, criminal background, health, etc. (Article 8(1) DPD)), the data provider was required to [demonstrate] that the parties concerned have given their consent for the archiving of their data and making these available by DANS for statistical and scientific purposes. If the licensor cannot demonstrate this consent, the data set shall only be accessible under very strict conditions, which, in accordance with Article 23 of the Personal Data Protection Act (WBP – Wet Bescherming Persoonsgegevens), set further requirements to the nature of the scientific research or the statistical processing, for the benefit of which use of the data set is considered necessary (DANS, 2009).

This is the only specific requirement regarding consent for archiving in any pre- or post-GDPR policy (such as, the “Privacy Declaration,” depositor instructions, etc.). For example, the “Deposit instructions for social and behavioral sciences” briefly discussed anonymization
expectations and who to contact if the data provider wants to deposit “certain privacy-sensitive information within the meaning of the Dutch Personal Data Protecting Act” (DANS, 2012), but nothing about providing evidence or proof that the data subjects consented to the archiving specifically. After the GDPR,

DANS does not ask from depositors any statements or documents regarding their data set on how it is GDPR compliant. We do have a clause in the deposit contract stating that a data set should be compiled in accordance with the GDPR. Since we are not the controller, we do not need that kind of documentation (DANS11).

While proof of consent for archiving is not required, when the data provider signs the “Deposit Agreement,” among everything else that they agree to, they are confirming “that the Dataset does not contain any data or other elements which, in isolation or upon disclosure outside the context of scientific research, are inconsistent with the Dutch Penal Code or other relevant national or international legislation” (DANS, 2019c).

For the administrative and operational data, DANS updated their processes to ensure that the data subjects (i.e., EASY account holders and DANS training participants) consent to the processing of their personal data. After the new policies were implemented in 2020, for example, DANS added a required acknowledgement that “Yes, I agree and understand the EASY General Terms and Conditions of Use and DANS privacy policy” (DANS, 2021b). Previous versions of the “EASY account registration form” only asked about the “General Conditions” (e.g., DANS, 2019b), even though the GDPR-compliant “Privacy Declaration” was already in effect. Similarly, for events, including trainings and workshops, DANS asks for the attendees’ consent for the processing when they register. According to DANS10, the increased focus on the importance of consent has meant that
we’re much more aware that we just should ask participants for their permission, because then it's fine. So, we have usually now every workshop has a little tag where we say, ‘we are processing your data for this workshop, Please agree’. So, I see that has changed in a lot of the projects.

4.3.5.5 Response to technological risks

DANS itself only made one change specifically addressing technological risks to privacy: DANS staff revised their data protection and data management training to be GDPR-compliant. The other changes reflecting this dimension of my theoretical framework were imposed or recommended by external parties. The GDPR itself was a response to and reflection of the advancements in data transfer, data linking, data analysis, and decryption capabilities since 1995. With many widely used products and services such as Microsoft Office 365, Google Cloud, and Dropbox, as well as online survey and other data collection tools, based outside of the EU, the Dutch research community has remained concerned about the security and GDPR compliance of these products and services. The SURF community’s Dropbox alternative, the SURFdrive, is one such attempt to address these concerns:

[it] is safe for personal data which is not very sensitive. So, it complies with the GDPR for normal personal data. And if you want to save special or sensitive personal data, you have to encrypt it or do something with it. And that resolves one of the problems, and it's accessible for all researchers (DANS08).

The EU projects and collaborations with CESSDA have also had to address concerns about the use of US-based servers, and whether the Google Cloud services could be used for these multi-national projects (DANS10). While the solutions differ across the various projects and needs, DANS staff are more aware of and mindful about just where data, personal and otherwise, are

---

162 SURF is the IT cooperative organization for academic and research institutions in the Netherlands. Both KNAW and NWO are considered members of SURF. The SURF community works together to develop solutions and best practices for IT services for education and research, provide IT services, and collaborate and innovate new solutions for open access, data storage and protection, and more. For example, SURF provides the SURFdrive for its members, a Dutch-based secure storage platform replacing Dropbox and Google Drive, which have privacy and data protection concerns under the GDPR, and the SURFfilesender, a secure file transfer system.
being collected, stored, and transferred. According to DANS08, the KNAW has no specific policy restricting the use of these products and services, though the “KNAW Privacy Statement” (2018) says that KNAW and its institutes must comply with the GDPR, UAVG, and any other relevant laws. Thus, the services and products that the institutes use must also be compliant.

There was also training for DANS staff about the GDPR and the data processing principles, including a quick reference guide explaining the data processing principles (Kraaikamp, 2020a) and exemplars on where GDPR concerns may arise with administrative and operational data (Kraaikamp, 2020b). According to DANS03, this training and the resulting guides and materials are frequently referenced as reminders for how to apply GDPR to DANS staff work.

4.3.5.6 Other changes

In addition to the changes which reflect concerns about the three dimensions of informational privacy, there is an additional change DANS made in response to the GDPR that is important to discuss: a change to DANS staff. DANS itself does not need a data protection officer (DPO) of its own; the KNAW as the legal parent organization has appointed a DPO who serves in that role for the KNAW and for all institutes and service providers. Prior to 2018, DANS already had a legal advisor, whose role was as the intermediary between the data archive and a “real” lawyer (who could also be the KNAW DPO). Neither the previous legal advisor nor the current one has a legal background; what is important for their role at DANS is their experience with the archive and the research data. The role of the legal advisor focuses specifically on the legal questions and considerations regarding the research data at DANS, and in many ways, that focus did not change after the GDPR (DANS11).
There was, however, a new position established at DANS and all the other KNAW institutes: privacy coordinator. In 2018, KNAW directed all institute directors to appoint a privacy coordinator (separate from the legal advisor role) (Koninklijke Nederlandse Akadanie van Wetenschappen, 2019). The privacy coordinator is someone who must “have an overview of which processes there are in this institute which involves personal data. And the privacy coordinator is also the one person who can put these processes into our central registration utility” (DANS08). At DANS, this position is the counterpart of the legal advisor, and in addition to coordinating the GDPR response with the legal advisor, director, and other DANS leadership, is also responsible for addressing the personal data concerns for all data not in EASY. The privacy coordinator is also the point of contact for and intermediary between DANS and the KNAW Chief Privacy Officer regarding the administrative and organizational data, but also GDPR compliance in general (DANS09).

4.3.5.7 Changes remaining to be completed

As of January 2022, DANS still does not consider itself to be fully GDPR compliant, though it is more compliant than it was in May 2018. As outlined in Section 4.3.4, DANS did not begin implementing changes until June 2018 when the “Privacy Declaration” was published, and it would be over a year before the first GDPR-compliant agreements came into use. What remains, according to the interviewees, are: reconsidering whether to be the EASY processor or controller, completing the KNAW process register, and addressing the backlog of EASY data requiring re-categorization and re-licensing.

The decision to be the processor for the EASY research data was a continuation of the pre-GDPR practice, although the explicit roles of processor and controller did not come into effect until the GDPR. As the processor, DANS is not responsible for deciding how and to
whom data sets with personal data are disseminated beyond the post-GDPR policy that all new
data deposits with personal data are automatically under Restricted Access. However, as the
processor, DANS’ mission as a data archive supporting long-term data preservation and reuse
has its limitations. For example, data sets with personal data without a clear, current rightsholder
(where, for example, DANS can no longer find and contact the data provider), “DANS is
thinking about being a controller for some data sets. But the situation will be a bit different in the
new version of EASY. And I don't really know what timescale that will be implemented”
(DANS02). Also, the controller

has still the right to remove the data, which is very contradictive with our
mission of long-term accessibility or long-term preservation. With long term,
we basically mean, until technical things fail, forever is actually what it means.
Of course, the GDPR allows the controller to retract data, and to have certain
data erased if that comes up in the situation. Of course, we do have to adhere to
these rights, but if we say, ‘we have a certain mission to do this and this’, then
we have more possibilities in terms of the long-term preservation, so to speak
(DANS11).

No decision has been made yet to change the status for the research data, because DANS is
waiting for a decision about possible UAVG revisions. According to DANS11, “if the proposed
changes are implemented, it might be more suitable for DANS to be controller, due to the legal
bases we could then apply. This would match our mission in providing long-term preservation.”
If DANS does change this role, then many of the policies, practices, and agreements discussed in
the previous sections will also be revised.

Another outstanding obligation relates to the “KNAW Process Register.” One of
KNAW’s obligations as the legal parent organization was to implement a process register
(verwerkingsregister) to document the following for all personal data processing: name and
contact information for the controller and the processor, purpose(s) of the processing, categories
of data subjects and personal data, categories of personal data recipients, retention periods, and
data security measures (Art. 30(1) and 30(2) GDPR). KNAW began first by identifying the KNAW and institute management personal data processes. It only began focusing on the research data-related processes in 2019 (DANS08). There is one central “KNAW Process Register” for the KNAW and all its institutes and service providers. For DANS, all “Processing Agreements” and “Processing Agreements DatverseNL,” as well as all processing for the other services and the administrative and operational data must be registered there. According to DANS11, this is still an on-going effort to register everything in the database:

The same thing goes for maintaining a register, the KNAW has a central register where all processing you do. We are still working on filling that in in the correct way and making sure that everything is also aligned within the institutions. This is a whole process in itself. So, I would say we are very well on our way.

The work is slow but fairly steady to identify all of the different processes and places within the organization (both the physical organization and the webpages) where personal data are collected, stored, and processed.

As discussed in Section 4.3.5.4, the changes in the access categories and usage licenses have resulted in a significant reduction in legacy access category use as data providers agree to re-categorize and re-license their data sets. This process is still not complete, and there are still data sets with personal data and metadata records with data subject personal data which pre-date the GDPR accessible in both Open Access categories still in use. The backlog of data sets is diminishing, but “the focus for now and for 2020 was implementing changes for the newer data sets. And the second step will be to change all the old data sets to go to the new policy” (DANS02). When DANS began developing the new access categories, the immediate priority was the post-GDPR data deposits (DANS02, DANS04, DANS11), and identifying all non-GDPR compliant data sets in EASY beyond the oral histories is not easy or fast. DANS archive
staff do look for them, but they are also notified by data users who find what could be considered “personal data” during their reuse, or something in a new deposit that may draw attention to a previously-deposited, related data set (DANS11). No matter how DANS identifies the data in EASY, it is “a very difficult and long-term project” (DANS11), not just identifying the data, but then to contact the data provider and determine what they are willing to do.

4.3.6 Challenges at DANS

DANS faced a number of challenges complying with the GDPR requirements. These challenges can be classified into three broad categories: related to the GDPR Implementation Act (UAVG), related to the relationship with KNAW, and related to compliance. A common concern running through all of the interviews was the issue of timeliness. At the time of their 2021 interviews, all DANS staff agreed that DANS was still not fully compliant with the GDPR, and that in May 2018, DANS was very aware that it would not be by the date the GDPR entered into force. This opinion remains as of mid-2022. The reasons behind this can be traced through many of the specific challenges discussed below.

One of the major challenges that hindered DANS’ ability to be fully compliant with the GDPR was how long it took for the UAVG to be become law in the Netherlands. The Dutch Parliament began discussing the replacement to the 2000 Personal Data Protection Act not long after the GDPR itself was announced in 2012. In 2016, the draft of the UAVG was made available for public consultation from December 9, 2016 to January 20, 2017.¹⁶³ There were 111 total responses, including one from the then-director of DANS. After the public consultation, the Autoriteit Persoonsgegevens¹⁶⁴ (AP) and the Council for the Judiciary (Raad voor de

¹⁶³ Of the 111 submitted comments, 67 are publicly available at https://www.internetconsultatie.nl/uitvoeringswetavg/reacties/datum. The remaining 44 have not been made public because the submitters did not consent to public viewing.
¹⁶⁴ The Dutch Data Protection Authority.
rechtspraak) provided their feedback, and the revised draft was sent to the House of Representatives (Tweede Kamer der Staten-Generaal) in December 2017. After the final vote to approve it on March 13, 2018, it went to the Senate (Eerste Kamer der Staten-Generaal), which can only approve or reject legislation, not propose revisions (Breitbarth, 2019; Eerste Kamer de Staten-Generaal, n.d.). The UAVG was adopted by the Senate on May 15, 2018, and signed by King Willem-Alexander the next day, nine days before the GDPR entered into force. The UAVG entered into force on May 25, 2018, the same day as the GDPR.

During the period of public and official consultation, a number of issues were raised which had to be addressed and clarified before the UAVG was voted on, including the burdens that the GDPR would place on small- and medium-sized organizations, the independence and powers of the AP, whether the requirements for a DPO and process register applied to the small- and medium-sized organizations, the age of consent for personal data processing, and how the GDPR and the UAVG would apply to journalistic, historical, statistical, and scientific processing (Breitbarth, 2019). The UAVG in the end is relatively policy neutral, in that it does not take advantage of most of the opportunities the GDPR left for the Member States to derogate from the requirements. Two derogations in particular, however, were used to allow for the processing of special categories in the public interest (Art. 24(b) UAVG) and to allow for data archiving (Art. 45(1) UAVG). Only once these derogations were confirmed as part of the UAVG could data archives like DANS fully understand what they could and could not do:

At DANS and at many other organizations, of course, in 2016 and 2017 even, there was still a lot of waiting on the implementation law and on how to handle this GDPR with your organization. Also, the KNAW was-- I think, almost all organizations were a bit-- They know they were overwhelmed by the new regulation, but they didn't do that much to prepare already or they couldn't do anything yet because they are dependent on specific implementation decisions there. Held up by the implementation law, and that came quite late (DANS11).
This delay was especially important for DANS because the initial versions of the draft UAVG did not provide derogations for scientific research data (DANS02, DANS10, DANS12). As mentioned above, DANS was one of several Dutch research and scientific institutions to submit comments during the public consultation period. The concerns DANS raised in this comment include: that DANS was not a formal, official archive according to the Archives Act, and thus the EASY research data were not covered by the official archival documents component of the draft; that the UAVG did not explicitly clarify that scientific, historical, and statistical research were in the public interest; and that the data subjects’ rights to data portability (Art. 20 GDPR) and to object to personal data processing (Art. 21 GDPR) would endanger the ability to verify and replicate research results (Doorn, 2017). The feeling of many was that they were not paying too much attention to make exceptions for research data. So, I know that DANS also initiated sort of lobbying work there, which must have been before it came into effect, because there was this development of the Dutch law (DANS10).

The time it took to confirm that the law would allow DANS to continue to function as a data archive and service provider significantly contributed to DANS’s failure to be fully, or even partly, compliant by May 25, 2018.

The delay in adopting the UAVG also complicated DANS’ response because DANS then had to wait for the KNAW, as DANS’ parent organization, to determine what it needed to do. For DANS to develop GDPR- and UAVG-compliant policies, first, the UAVG needed to codify the derogations to the GDPR, and then the KNAW needed to decide what both laws meant for KNAW. Only then could the KNAW determine what the institutions and service providers needed to do, at a minimum. From the KNAW perspective, this experience was especially

---

challenging because of how vastly different the institutes and service providers are, on top of
continuing with the day-to-day operations, and with only limited funding to support the transition
work, it was neither easy nor fast (DANS08, DANS09). Even then, the institutes and service
providers’ priorities came out of the data protection impact assessments (DPIAs); DANS’ DPIAs
were completed relatively early in fall 2018. While DANS had already begun reviewing its data,
processes, and security measures not long after the GDPR was adopted, nothing could be
permanently changed or adopted until the KNAW had done their part:

What's interesting is that, actually, we were quite a few steps ahead of the
KNAW. We were looking for answers, but they hadn’t really picked it up yet
for themselves. … We had often thought we were ahead of—Well, ahead of
them in some ways in the details, I guess. That made it quite hard because we
also felt we wanted some guidance or some steering from them, because that
makes it's, of course, easier, as long as that guidance is also in the direction we
wanted to go (DANS11).

The DPO also changed, which according to DANS04, meant that:

… in the last three years, there were I think, probably two restarts of the whole
process. … And that's why I think it was a fully time consuming, because it
changed. … In the GDPR you have a data protection officer. And that person
also changed. So, with the new data protection officer, they started over, and so
it came with the same questions. And when you said, ‘well, we've discussed it
with the previous officer’. And then they said, ‘Well, no, I want to start again
over. Because I want to be sure that everything is alright’. So, there was a lot
of repetition.

The final set of challenges that the interviewees identified related to compliance. These
challenges are on-going, and not simply what DANS dealt with between 2016 and 2020 when
the final agreements came into use. There are two compliance-based challenges I will discuss
below: understanding the requirements and DANS and EASY user compliance. Understanding
not just what the GDPR and UAVG said in the text, but how DANS’ day-to-day operations
would be affected, was highlighted as a particular challenge by several respondents:
Comprehension? Yeah, it's even with these little images, and when you've sat through and understand it, it's too complex to easily remember. You need to be able to reference it all the time. Even for me … When I have a doubt, I have to go back and check these things, I can't remember specifically. And I think that makes it hard to like day-by-day be like, ‘yeah, this is GDPR and it's okay’ (DANS03).

Because there are all kinds of interpretations of GDPR. I heard colleagues saying things like, ‘I don't ask our legal advisers anymore because if you ask five legal advisers, you will get stuck with six opinions.’ One thing in all kinds of projects that I learned is that GDPR is something that we all have questions about that legal advice come with different answers. It's not always that crystal clear, so you need to implement it, but on the other end, you need to accomplish your mission (DANS13).

Even though some people find the GDPR itself to be a very readable law as legal texts go (e.g., DANS12), and the GDPR requires that privacy policies be in “clear and plain language” (Art. 12(1) GDPR, see also: Becher & Benoliel, 2021; Spagnuelo et al., 2020), there are many possible interpretations of the requirements. Often, legal advisors and DPOs do not have experience with both research data and legal requirements for protecting privacy. For example, at DANS, the legal advisor does have the archives experience but not the legal background, while the KNAW DPO and Security Officer have the legal and data protection backgrounds, respectively, but no experience with research data or archives.

Not only were DANS staff concerned about their own compliance in their work, they also had to be concerned about whether DANS data providers and data users were compliant. For example, even after adding the question about the presence of personal data to the “Deposit Form” (Section 4.3.5.3), DANS staff must still review all deposited data sets for the presence of personal data, whether the data provider responded “yes” or “no,” and the review process has become much stricter. This is not an indication of possible nefarious intent on the part of the data providers so much as it reflects their continued confusion over how GDPR affects what they can do with their data:
Well, it's much more difficult for researchers to do their research with data. With privacy aspects, of course, it's harder to get the data, to get access to them. So, that's one of the main problems of GDPR from the standpoint of a researcher. … This is, from a standpoint of a researcher, this is really a challenge. Yeah. And that's one of the problems for researchers and universities also that the privacy officers and legal experts always want to be on the safe side. So, they are not of much help to researchers, who want to make use of the data set (DANS07).

Data providers may also not be sure about whether their data are or include personal data as GDPR defines them. They are also often not aware that, by retaining the key to their anonymized data set, the data are actually pseudonymized and thus under GDPR’s purview as personal data (DANS02, DANS13, DANS15).

4.3.7 Impacts on DANS operations

Since the GDPR entered into force, DANS implemented a number of policy and practice changes which affected the ways in which data are brought into DANS and how users are able to access them. Have there been observable trends in DANS usage aligning with their response to the GDPR? This section will examine the operational metrics of archive use—data deposits and data set downloads—and discuss other outcomes identified through the interviews. From 2010 to 2016, there was a relatively steady increase in the number of data sets available in EASY (Fig. 4.12) and in the number of data sets published per year in EASY (Fig. 4.13). Beginning in 2017, DANS has negotiated contracts with various data providers to securely ingest and archive data through secure, automated machine-to-machine transfers. This automated data ingesting process accounts for the sharp increase in the total number of data sets (particularly the archaeological data), although not all data providers are able to use it.
The two most-represented discipline areas—archaeology and the humanities (excluding archaeology)—saw corresponding sharp increases in total numbers of data sets after 2017 (Fig. 4.14 and 4.15,\textsuperscript{166} respectively). As previously discussed (section 4.3.2), all archaeological excavations in the Netherlands are required to deposit their digital data files at DANS, and these data represent the majority of data available in EASY. The implementation of the automated

\textsuperscript{166} This chart reflects the discipline areas with the most data sets in DANS EASY and does not include some of the smaller discipline areas with far fewer data sets. The totals in this chart do not equal the total data sets in EASY.
data ingest process supported the deposit of over 60,000 archaeological data sets after 2016, more than twice what had been deposited from 2010-2016; these data are not curated or processed by DANS staff, and can be considered “published” when they are deposited. A similarly sharp increase in data set publications can be seen with humanities data from 2017-2019 (an increase of over 3,000 data sets in three years, compared with approximately 2,100 from 2010-2016), though it has slowed down since 2019. The other main disciplinary areas (behavioral science, life sciences and medicine, and social sciences) saw a slight increase in the number of data sets published per year (Fig. 4.15), but not to the same degree as archaeology and humanities.

![End-of-Year Total Number of Data sets (Archaeology) 2010-2020](image)

*Figure 4.14 Annual end-of-year total archaeological data sets in DANS EASY (2010-2020) (Doorn, 2022).*
From the interviewees’ perspective, the GDPR itself did not have much of an effect on the number of data set publications at DANS. With the number of published data sets in the tens of thousands per year since 2017 as a result of the new automated ingest capabilities, it is difficult to identify what, if any, effect the GDPR may have had on data depositing itself in that time period. As DANS13 described it,

… we are more strict at the door. That's, I think, the most important issue, but I don't think we have many, many less data sets because of GDPR. I don't believe that, no. Maybe, that, in the head of the researchers, if I would check with the research, that they would tell me, ‘I'm depositing less data’. That would be something that I don't know, of course, but I didn't see it at DANS in the numbers.

Since the data providers had the responsibilities, if not the title, of the controller before the GDPR came into effect, the requirements after the GDPR are not significantly different from what the data providers were responsible for before. The most significant impact on data coming into EASY from 2016 on is the automated deposit system.

One metric where we can see the effects of GDPR-related changes at DANS is in how the data are made accessible in EASY. Before the GDPR entered into force, the vast majority of
data were available Open Access, as long as the user was registered with EASY (Fig. 4.16, see also Section 4.3.5.4). Doorn has attributed this pre-GDPR growth in Open Access accessibility, even with the registration requirement, to a greater “sign of acceptance of open science principles, at least among researchers depositing their data in the DANS archive” (Doorn, 2020, p. 46). In 2018, the data in that category were still the majority of accessible data, though there are more data archived at DANS but accessed through Mendeley or Dryad. In January 2020, however, DANS implemented its new access categories and began the process of re-categorizing the data into the two main categories which will carry over to the Data Stations: Open Access and Restricted Access.

![Figure 4.16 Biannual Availability of DANS EASY Data sets by Access Category (2012-2018) (DANS, 2019a; Doorn, 2022).](image)

In terms of use of the data from EASY as measured by number of downloads per years, after period of growth from 2010-2013, the downloads per year remained fairly stable through 2017 (Fig. 4.17). There is one exception not visible in Fig. 4.17: in 2014, as reported in Doorn (2020, 2022), one user downloaded every Open Access archaeological data set available at the
time, approximately 15,000 individual data sets.\textsuperscript{167} Beginning in 2018, the total annual downloads has increased again. Use of the archaeological data has dropped slightly from its 2016 peak, but the archaeological data remain the most downloaded overall. Use of the other humanities data and the social sciences data, as well as the behavioral sciences data, has fluctuated since 2015, while use of the life sciences data downloads has steadily increased since 2010.

![Total Annual Downloads (DANS) 2010-2019](image)

\textit{Figure 4.17 Total annual data downloads from DANS EASY (2010-2019). Of note, in 2014, approximately 15,000 data sets were downloaded by a single user, a value not included in this chart (Doorn, 2020, 2022).}

As with the rate of data publications, DANS interviewees do not feel that the GDPR had much effect on the later reuse of the data in any significant way. Several \textit{did} identify the requirement for data provider permission to access the Restricted Access data as a potential barrier to reuse in general. For example,

\begin{quote}
Maybe it's like, we live in a kind of a click-world where you want things in the moment that you come across them. And in the case that something isn't just open access completely, the barriers to having to sign up for something, and to sign up for your login. And then further, if the data set isn't completely open access, that maybe the data manager needs to get into contact with the data owner. And that there is this kind of back and forth, and it's a couple of days
\end{quote}

\textsuperscript{167} The exact number of downloaded data sets has not been reported in greater detail than “about 15,000 data sets in total” (Doorn, 2020, p. 47).
later and stuff. I think that if it's some time-pressed researcher, by the time they get the alert that the data set can be downloaded, it’s just gone into some kind of spam box or something (DANS03).

However, the need for data provider permission for these data already existed before the GDPR, so it is unlikely that the fluctuations in data downloads after 2016 can be attributed to this, and, from 2018 to 2021, the number of overall Restricted Access data sets has decreased as DANS negotiates with data providers to adopt the new access categories and licenses.

Overall, the main impacts of the GDPR at DANS were experienced in the policies and practices internal to the archive, the agreements and the data access categories and licenses in particular. While data providers and data users also encountered these changes in their interactions with EASY, and DANS did receive questions from both sides related to confusion about the GDPR, the changes do not appear to have significantly affected the use of DANS as an archive. What changes we can see—the sharp increase in data deposits after 2017—are attributable to the adoption of large scale, machine-to-machine data depositing for certain data providers who have contracts with DANS to do so. The on-going transition from EASY to the Data Stations is also not likely to significantly affect DANS use, as the EASY platform will still be available while the data are migrated to the new services. Instead, the biggest change outside of the policies and practices (for example the “Processing Agreement”) is the greater appreciation and awareness of the importance of data protection and privacy. DANS staff were already well-versed in the importance of these two ideas when it came to the research data, but there is a much greater focus on the other personal data they interact with on a daily basis.
4.4 GESIS — Leibniz Institute for the Social Sciences (GESIS)

4.4.1 Archive description

The GESIS — Leibniz Institute for the Social Sciences is the largest social science research infrastructure in Europe and one of the oldest in the world. The predecessor of the modern GESIS data archive, the Central Archive for Empirical Social Research (Zentralarchiv für Empirische Sozialforschung) (ZA), was established in 1960 as an archive for social science researchers at the University of Cologne (Mauer, 2012). The ZA was a founding member of the Council of European Social Science Data Archives (which became the present-day Consortium of European Social Science Data Archives (CESSDA) in 2013) in 1976 along with the Steinmetz Archive, the UK Data Archive, the Danish Data Archive, the Belgian Archives for Social Sciences, the Norwegian Social Science Data Services, and the Data Archive for Social Sciences in Italy (Consortium of European Social Science Data Archives, n.d.a).

The German Social Science Infrastructure Services (Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen) was founded in 1986, comprising three legally independent social science institutes: the ZA in Cologne, the Information Center for Social Sciences (InformationsZentrum Sozialwissenschaften (IZ)) in Bonn, and the Center for Surveys, Methods, and Analysis (Zentrum für Umfragen, Methoden und Analysen (ZUMA)) in Mannheim. The IZ was initially founded in 1969 by the Association of Social Science Institutes (ASI), and provided information about current and completed social science research at partnering German-speaking research institutes and universities (Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, 2000). ZUMA was founded in 1974 by the German Research Foundation (Deutsche Forschungsgemeinschaft), a federal- and state-funded research funder. ZUMA’s role was to provide research methods support services for survey-based research,
though it supported other methods and did not restrict its services only to social science researchers (“ZUMA - ZENTRUM FÜR UMFRAGEN, METHODEN Und ANALYSEN (Center for Surveys, Methods, and Analyses),” 1984). By 2007, ZA, IZ, and ZUMA were all members of the Leibniz Association, one of several large associations supporting research in Germany.

GESIS was formally created in 2007 when ZA, IZ, and ZUMA merged into a single legal entity, though the three different locations were maintained; the headquarters of GESIS was in Mannheim. The full infrastructure consolidation took approximately five more years, especially the IT infrastructure consolidation and standardization (GESIS01). GESIS was renamed the GESIS-Leibniz-Institut für Sozialwissenschaft the following November 2008. The Leibniz Association was founded in 1990 as part of German reunification and originally comprised 46 so-called “Blue List” research institutes identified beginning in 1977 as eligible for joint federal and state funding in West Germany. There were eighty-one institutions by 1992; there are now 97 Leibniz Institutes spanning all discipline areas.

In November 2011, the Bonn and Cologne GESIS locations merged into the Cologne location while the headquarters remains in Mannheim. The Cologne location is in North Rhine-Westphalia, while Mannheim is in Baden-Württemberg; this dual-location is important because of legal differences between the two states that GESIS must operate between (Sections 4.4.1.5 and 4.4.6). 2013 and 2014 saw the launch of several new GESIS services and research areas, including the Secure Data Center for on-site, secure access to identifiable data and the Computation Social Science (CSS) department in 2013, and the self-archiving platform SowiDataNet|datorium in 2014. GESIS received the CoreTrustSeal in September 2017 (CoreTrustSeal, 2017).
In 2018 when the GDPR entered into force, GESIS was organized into five research departments in addition to the IT, Administration, and Knowledge Transfer departments. The research departments were: Survey Design and Methodology (SDM) (survey research support, including expertise in the European Social Survey (ESS)); Monitoring Society and Social Change (DBG) (German components of international surveys, gender equality research, host of German Microdata Lab, national-level surveys); CSS (developed research approaches to incorporate digital behavioral data into more “traditional survey data in order to improve the analyses of a broad range of socio-cultural phenomena” (GESIS Leibniz Institut für Sozialwissenschaften, 2018b)); Knowledge Technologies for the Social Sciences (WTS) (provided digital research resources to social science researchers, including information portals and research data collection and archiving tools); and, the Data Archive for the Social Sciences (DAS) (research data archiving for long-term preservation and reuse, tools and standards development for metadata, DOI registration through da|ra,168 the Secure Data Center, and geospatial and digital behavioral data linking). In addition to these departments, GESIS also runs five Research Data Centers (RDC) supporting international and national surveys GESIS is involved with can be accessed: RDC ALLBUS, RDC German Microdata Lab, RDC International Survey Programs, RDC PIAAC, and RDC Elections. GESIS staff from other departments involved in these studies are also part of their respective RDCs, and data from several of the RDC surveys are later archived in the Data Archive.

In July 2021, GESIS restructured its operations, including the DAS, which was split into two separate departments: Data Services for the Social Sciences (DSS) and Survey Data

---

168 da|ra is the DOI registration service for German social science and economics data. GESIS collaborates with DataCite and the ZBW-Leibniz Information Centre for Economics to offer this service. All data archived at GESIS (through both the Data Archive and SowiDataNet(datorium) receive a DOI through da|ra. For more information see https://www.da-ra.de/.
Curation (SDC).\textsuperscript{169} This division was made because, by that time, the DAS had become very large and difficult to manage, especially in comparison with other GESIS departments (GESIS09, GESIS10). There were also two fairly distinct sets of operations within the DAS: one focused on the data acquisitions and archiving, and one focused more on the various GESIS-involved surveys (Section 4.4.1.3). The work of the two new departments is still somewhat interconnected, but there are now two dedicated management structures with more manageable department sizes. The DSS is responsible for the data archive operations, the Secure Data Center, the da\textit{ra} service, and research on and support for research data management, preservation, and metadata. The SDC “processes and prepares social science research data across a diverse range of data types to support research, especially in comparative perspectives across space and time” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.o), with a focus on the international and national surveys that GESIS conducts or collaborates on. The SDC is also responsible for data linking and georeferencing to add context to traditional survey.

\textsuperscript{169} An additional 2021 organizational change was the renaming of the Monitoring Society and Social Change (DBG) department to Data and Research on Society (DRS).
As a leading social science research infrastructure in Germany, GESIS’ services support all stages of a research project (see Fig. 4.18). Many of these services may come at a cost. For example, for researchers at the very beginning of the research process, GESIS provides consultations on survey development, pre-testing, obtaining samples, and using the GESIS panel. Many of these consulting services, but not all, have a cost; e.g., €100+VAT to draw a telephone sample or €400+VAT for “complex” samples (GESIS Leibniz Institut für Sozialwissenschaften, n.d.d). Other phases are supported through the work of the DSS, SDC, and the RDCs.

4.4.1.1 Archiving services

The Data Services for the Social Sciences department (DSS) is comprised of three teams: Archiving, Data Acquisitions and Access, and Metadata Standards and Interoperability. Data providers have three paths (or “service packages”) to archive their data at GESIS, either with or without a service contract, which determine which data archive the data are archived in. All data deposited without a service contract are deposited into the SowiDataNet|datorium repository,170 GESIS’ self-archiving system implemented in 2014. This service package is called “Archiving BASIS” and offers data storage for a minimum of ten years, bit-stream preservation, DOI registration through da|ra, a form for structured metadata, and access to these data through the SowiDataNet|datorium archive without cost (GESIS Leibniz Institut für Sozialwissenschaften, n.d.i). The data provider also provides the data set description and chooses the access method for their data; no matter the access method, the data are available without cost. Research institutions without their own archiving infrastructure can sign a contract with SowiDataNet|datorium to publish up to ten studies per year, with added fees for additional

---

170 Archiving BASIS data can be searched for and accessed either through the SowiDataNet|datorium webpage at https://data.gesis.org/sharing/#/Search or through the main GESIS search page, https://search.gesis.org by filtering the data source to “GESIS-SowiDataNet | datorium.”
studies (GESIS Leibniz Institut für Sozialwissenschaften, 2019a). This contract enables affiliating data sets with specific institute collections, and the data deposit will be reviewed by both a GESIS data curator and a curator from the home institution.

There are four access methods available for the Archiving BASIS data in SowiDataNet|datorium: Free access (without registration), Free access (with registration), Restricted Access, and Embargo. GESIS account registration is required for all SowiDataNet|datorium data providers and for data users to access the Free access (with registration) and Restricted Access data; the Restricted Access data also require the data provider’s permission. The access request is sent to the data curator at GESIS, who then contacts the data providers. Once permission is granted, the curator then makes the data available to the user (GESIS Leibniz Institut für Sozialwissenschaften, n.d.i). The data under Embargo can be embargoed for up to two years and then automatically published, though the metadata will be available immediately.

GESIS staff review all Archiving BASIS data for the following (referred to as “Ingest Check BASIS”): match to collection policy; legal basis for collection; documentation; information on the research methods; completeness; match between metadata/documentation and the data set; whether data protection, privacy, and copyright are accounted for; and, reusability (GESIS Leibniz Institut für Sozialwissenschaften, 2022a). If personal data are found during this deposit review, GESIS staff will not make any changes to the data; they will be returned to the depositor for changes (GESIS09). If Free access (without registration) is selected, data providers can select Creative Commons or Open Data Commons licenses as the license for their data; Creative Commons licenses cannot be applied to the other two options because of conflicts between the licenses and the access category (GESIS Leibniz Institut für Sozialwissenschaften,
SowiDataNet|datorium can be accessed either through the SowiDataNet|datorium web portal\(^{171}\) or through the regular GESIS webpage search feature.

The other path for data deposit is with a service contract signed between GESIS and the data provider. These data are archived in the GESIS Data Archive. There are two types of service packages for the Data Archive—“Archiving PLUS” and “Archiving PREMIUM”—which differ in terms of cost, associated services, and intensity of ingest checking. Once either contract is agreed upon and signed, the data are transferred to GESIS in one of three ways: via secure Cryptshare upload server, a download server, or through the postal service on a data storage medium (GESIS Leibniz Institut für Sozialwissenschaften, n.d.g). Both contracts transfer the “rights of use” (which include the right to make the data available for reuse) from the data provider to GESIS; these “rights of use” are how the data are licensed for reuse. All PLUS- and PREMIUM-contracted data are available for registered GESIS users under one of four access classes (these differ from the SowiDataNet|datorium data access classes): 0, A, B, and C.

Category 0 data are available for all users free of charge. Category A data are also free of charge, but restricted to academic research and teaching purposes.\(^{172}\) Category B data are restricted to academic research and teaching purposes, although publications of results require GESIS permission. This category has not been used since 2010, when two data sets were deposited, though from 2010 to 2018, there were thousands of downloads of Category B data per year (GESIS - Data Archive for the Social Sciences, 2019). Category C data are restricted to academic research and teaching purposes, but also require data provider permission. This

---

\(^{171}\) [https://data.gesis.org/sharing/#!Search](https://data.gesis.org/sharing/#!Search)

\(^{172}\) In the data set catalog record, the data user is required to select from a list of usage purposes before the download can start: “not specified; for final thesis of the study programme (e.g. Bachelor/Master thesis); for research with a commercial mission); for further education and qualification; for scientific research (incl. doctorate); in the course of my studies; in a course as a lecturer.” Selecting a response is required for all downloads, even for Category 0 data, even though usage purpose is not restricted for those data.
permission must be in writing and requires information about the data user and what they intend to do with the data to be shared with the data provider. The request is sent to the data curator, who consults with the data provider; once permission is granted, the curator then makes the data available to the data user (GESIS Leibniz Institut für Sozialwissenschaften, 2018d). Both Category B and C data may come at a cost to the user; whether there is a usage fee is data set dependent (GESIS Leibniz Institut für Sozialwissenschaften, 2019b). Data sets can be recategorized at will, which makes the annual numbers of data sets per category inconsistent depending on whether the counts are based on publication year or record of access category change.

In addition to these possible data usage fees, there are fees for data providers who want to deposit their data into the Data Archive. The Archiving PLUS package costs €1560.75+VAT (GESIS Leibniz Institut für Sozialwissenschaften, 2021a). Once the contract is signed and data are transferred to GESIS, the Archiving team review the deposit. The “Ingest Check PLUS” deposit review includes all the review steps of “Ingest Check BASIS” as well as additional checks for format, including possible format migration; checking the questionnaire or code schema against the variables; and, checking variable and value labels (GESIS Leibniz Institut für Sozialwissenschaften, 2022a). Archiving PLUS also includes long-term preservation and GESIS-provided study level documentation (including bilingual German and English). Archiving PREMIUM is available at a variable price based on the services chosen (GESIS Leibniz Institut für Sozialwissenschaften, 2021b). The “Ingest Check PREMIUM” includes all components of the “Ingest Check BASIS” and “PLUS” as well as: completeness and agreement between data and survey instrument, correcting variable and value labels, and a systematic check for missing values (GESIS Leibniz Institut für Sozialwissenschaften, 2022a). Additional
archiving PREMIUM services include: checking that the data match the survey instrument routing; checking for “substantive” consistency; check for correctness and response quality; and, study and variable level documentation (GESIS Leibniz Institut für Sozialwissenschaften, 2021b, n.d.a). Changes to the deposit review process at GESIS coincided with the new data archiving contracts, and reflect the results of a GESIS study of the data curation work times (Perry & Netscher, 2022).

DSS also operates the Secure Data Center, located on-site in Cologne. The Secure Data Center closed in 2020 due to COVID-19 restrictions, and underwent IT infrastructure upgrades before reopening in late spring 2022 in accordance with COVID-19 policies. The Secure Data Center allows users to access disclosive data to which specific access requirements and restrictions have been applied, as well as survey data which have been geo-referenced (GESIS Leibniz Institut für Sozialwissenschaften, 2022b). For example, the German wave of a multinational, longitudinal survey of children of immigrants, due to the size of the population and risk of re-identification, is only available on-site in the Secure Data Center (Kalter et al., 2021). Secure Data Center users must request access to the data and sign a data use agreement with GESIS. All Secure Data Center work undergo two output checks for potential disclosiveness. If the data are from a specific RDC, one output check will be completed by the RDC and one by Secure Data Center staff; otherwise, both output checks are completed by Secure Data Center staff (GESIS17).

The Data Acquisitions and Access team does some “active” data acquisition. If they are made aware of a research project or data set, they will contact the research team about giving the data to GESIS (GESIS10). Otherwise, there are two main paths for data acquisition. In the first, the data provider contacts GESIS while writing their grant proposal, because the grant funder
requires an estimate for archiving costs. In these discussions, the Acquisitions team will also provide feedback on potential problems or considerations, including the research informed consent documentation, that the researchers should keep in mind if they want to archive their data with GESIS at the end of the project. These research teams do often return to GESIS after the project to deposit their data (GESIS10). The second path is for data providers to come to GESIS to archive their data towards the end of the research process. When the archiving contract is signed, the data are shared with GESIS and reviewed (GESIS10, GESIS12). No matter which archiving package is chosen or path to acquisition is followed, all data archived at GESIS are reviewed by the Archiving team.

GESIS does not restrict data from being archived based on discipline; the only disciplinary requirement is that the data are relevant for answering social science questions\textsuperscript{173} (GESIS Leibniz Institut für Sozialwissenschaften, 2013). To that end, the data in the Data Archive include data sets from political science, economics, criminal justice, genomics, life sciences, theology, and demography, among many others. The self-archived data in SowiDataNet|datorium likewise cover multiple disciplines. The data are primarily quantitative surveys, due to the historical split between qualitative and quantitative research in Germany (see Section 4.4.2). GESIS has a preferred file formats list (GESIS Leibniz Institut für Sozialwissenschaften, n.d.), and as discussed above, during ingest checking, GESIS may migrate the file format if necessary. There is a growing collection of digital behavioral data in both the Data Archive and SowiDataNet|datorium.

For data to be archived at GESIS, the data must be anonymized. There can be no direct identifiers and they must meet “factual anonymization” as defined by the Bundesstatistikgesetz

\textsuperscript{173} GESIS’ collection policy states that “GESIS archiviert Forschungsdaten, die für die Beantwortung sozialwissenschaftlich relevanter Fragestellungen geeignet sind” (GESIS Leibniz Institut für Sozialwissenschaften, 2013).
As discussed in Section 4.4.1.5, factually anonymized data are those data for which re-identification would require a “disproportionally large effort in terms of time, cost, and labor” (§16 Abs 6(1) BStatG). If the research data cannot be anonymized to that extent while remaining usable, usage restrictions can be imposed, such as requiring a usage contract or use only in the Secure Data Center (GESIS Leibniz Institut für Sozialwissenschaften, n.d.j). During the deposit review process, if personal data are found in the data going to the Data Archive, the Archiving team may make minor changes, but if major changes are required to meet the anonymization standards, the data are returned to the data provider (GESIS02, GESIS04). Depositing data that can only be accessed through the Secure Data Center is also a very expensive endeavor, for both the data provider and GESIS (GESIS07, GESIS10).

As of December 31, 2021, GESIS contains a total of 6,694 data sets: 6,360 in the Data Archive and 334 in SowiDataNet|datorium (GESIS Leibniz Institut für Sozialwissenschaften, n.d.n; Zenk-Möltgen, 2022). In the Data Archive, there were 1,047 data sets in Category 0, 4,047 in Category A, 239 in Category B, and 1,027 in Category C (Zenk-Möltgen, 2022). In SowiDataNet|datorium, there were 159 data sets in the Free Access (without registration) class, 114 in the Free Access (with registration) class, and 57 under Restricted Access. The embargo period for SowiDataNet|datorium lasts for up to two years, and from the available data, it is not

175 § 16(6), BStatG (2016): “Für die Durchführung wissenschaftlicher Vorhaben dürfen das Statistische Bundesamt und die statistischen Ämter der Länder Hochschulen oder sonstigen Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung 1. Einzelangaben übermitteln, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können (faktisch anonymisiert Einzelangaben), 2. innerhalb speziell abgesicherter Bereiche de Statistischen Bundesamtes und der statistischen Ämter der Länder Zugang zu formal anonymisierten Einzelangaben gewähren, wenn wirksame Vorkehrungen zur Wahrung der Geheimhaltung getroffen werden.” (For the conduct of scientific projects, the Federal Statistical Office and the states’ statistical offices may permit universities or other institutions tasked with independent scientific research to: 1. Pass on personal data if the personal data can only be matched with a disproportionately large expenditure of time, cost, and manpower (factually anonymized personal data), 2. permit access to formally anonymized personal data within specially secured spaces of the Federal Statistical Office and the state statistical offices, if effective precautions are taken to maintain privacy). Translation by the author.
possible to tell what data were still under embargo before 2022; however, four data sets were published under embargo in 2021 (in that the metadata were published, and the data sets will be published into the chosen access class at the end of the embargo period) (GESIS Leibniz Institut für Sozialwissenschaften, n.d.n).

4.4.1.2 Georeferencing and data linking

One of the teams in the SDC, the Survey Data Augmentation team, work to add value and enrich survey data. A relatively recent GESIS service is linking survey data with geospatial and social media data. The value of these linked data is that the “data from other sources can be used to contextualize survey data and … [enable] researchers to answer novel research questions or to test the robustness of findings that are based exclusively on self-reported data from surveys” (Beuthner et al., 2021, p. 1). Beginning in 2015, the then-DAS and DBG departments collaborated on a project to develop the necessary conceptualizations and infrastructure to use geospatial data to add context to an enrich survey data (see Bensmann et al., 2020; Jünger, 2019; Schweers et al., 2016). Georeferencing or geocoding (this work at GESIS is referred to as both within the interviews) is done for data that GESIS is involved in the collection for, such as the ALLBUS,¹⁷⁶ as well as for external data sets. The process for geocoding is to take the address or location information of the data subjects and convert that to a geocoordinate or other spatial identifier (Bensmann et al., 2020; Schweers et al., 2016). Because there is an increased disclosure risk with these data, they are only available for use in the Secure Data Center.

In addition to this georeferencing work, in 2016, GESIS began working with what they refer to as “digital behavioral data” (and “digital trace data”) from social network applications,

---
¹⁷⁶ ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, or the German General Social Survey (GGSS)) is a biennial survey measuring social and political attitudes in Germany conducted since 1980. Data from this survey are available through the Data Archive; currently, data from the 1980-2018 waves have been published.
such as Twitter or Facebook. These data “can be used to measure social behavior, the formation of attitudes and opinions and interactions of users and digital platforms by making their dynamics measurable” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.h). The Data Archive’s already published data sets are Twitter and Facebook posts from the last three German federal elections, Twitter corpora about COVID-19 and annotated tweets from 2013-2020, sexist tweets, and Wikipedia data about international politicians, multilingual historical focal points, and platform data (GESIS Leibniz Institut für Sozialwissenschaften, n.d.c). There are also nine digital behavioral data sets from external data providers in SowiDataNet.datorium. Just as with the georeferencing work, the Survey Data Augmentation team will link survey data to these digital behavioral data (Breuer et al., 2020, 2021; Jünger & Breuer, 2021; Stier et al., 2020). Currently, the digital behavioral data sets are aggregated and anonymized so that they can be published open access (GESIS14), but access to survey data which have been linked to them may be restricted to the Secure Data Center or other access restrictions. This is an on-going discussion between the staff working with these data and the Secure Data Center.

4.4.1.3 External collaborations

In addition to the data archiving and research support services described in the previous sections, GESIS is very involved in international collaborations related to research in the social sciences. It is a member of two European Research Infrastructure Consortiums (ERICs): the Consortium of European Social Science Data Archives (CESSDA) and the European Social Survey (ESS). As discussed in Section 4.4.1, GESIS’ predecessor ZA was a founding member of CESSDA in 1976, and GESIS remains the German national service provider. GESIS staff have provided CESSDA’s data management training, developed tools for researchers to
harmonize survey variables, and collaborated on European projects including the Social Science & Humanities Open Cloud (SSHOC) and DataverseEU.

The European Social Survey is a biennial cross-national survey based at City, University of London. It has been an ERIC since 2013, though the survey has been conducted since 2002. GESIS is a member of the Core Scientific Team and is responsible for conducting the German component. The ESS team at GESIS is responsible for translating the English language questionnaire and other documentation into German, as well as the data collection, pre-deposit data preparation, and data curation (GESIS Leibniz Institut für Sozialwissenschaften, n.d.q). These external collaborations and other GESIS-led or involved research and infrastructures (both national and international) are supported through the five Research Data Centers. With these RDCs, “GESIS partially participates in data collection on permanently takes over the tasks of data processing, archiving and delivery” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.l). For many of the national surveys associated with the RDCs, GESIS contracts with a field work agency to actually conduct the survey with GESIS’ supervision. In addition to federal funding for its operations, GESIS also receives funding from the Ministry of Education and Research (Bundesministerium für Bildung und Forschung (BMBF)). While there is no specific dedicated funding for either CESSDA or ESS, a portion of the operational funding is allocated to supporting both of these collaborations.

4.4.1.4 Organizational structure

As discussed above, GESIS is an independent research infrastructure in Germany, even though it is funded at the federal and state levels (from both North Rhine-Westphalia and Baden-Württemberg). While the BMBF also provides funding dedicated to specific projects, GESIS is able to determine its operational and strategic goals without additional government interference,
as guaranteed by Article 5(3) of the German constitution, the Basic Law. GESIS (and its predecessor institutes) has been a long-standing member of the Leibniz Association. The purpose of the Leibniz Association is to bring together research institutes across the research disciplinary landscape for interdisciplinary research and knowledge sharing. Every seven years, the Leibniz Association assesses each member institute in the following areas: overall concept (e.g., research outputs, publications, public perception, competitiveness, etc.); strategic planning; controlling and quality management (e.g., funding, IT infrastructure, management, good scientific practice, data management, internal performance, etc.); human resources (including training, doctoral candidates, and equal opportunity efforts); cooperation and collaboration; and, an evaluation of the institution’s subdivisions (Leibniz Gemeinschaft, 2018). If an institute is found to be severely deficient, the Leibniz Association Senate can recommend that federal and state-level funding be discontinued, and the institute can be removed from the Leibniz Association. In this capacity, the Leibniz Association is GESIS’ primary external oversight; however, it does not have supervisory control.

GESIS has three advisory boards acting as more direct oversight of its operations and plans. They are the Board of Trustees (Kuratorium), the User Advisory Board (Nutzerbeirat) (UAB), and the Scientific Advisory Board (Wissenschaftlicher Beirat) (SAB). The Kuratorium supervises the President and generally approves any financial and other business of GESIS beyond normal day-to-day operations. … so, the big projects, future orientated, new perspectives, conceptualizations of new goals and objectives, and so on. All this, and of course, the budget (GESIS16).

Additionally, the Kuratorium approves the appointment and removal of the GESIS president, long-term development plans, and organizational changes (such as the creation of two departments out of the DAS in 2021), among other responsibilities (GESIS-Kuratorium, 2014). The Kuratorium is split into two groups, one with full voting rights, and one with a primarily
advisory role. The first group includes one representative each of the BMBF, the Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg, and the Ministerium für Kultur and Wissenschaft des Landes Nordrhein-Westfalen. All three current ministerial representatives have experience as researchers as well as the policy aspect of research data management (GESIS16). There are also four representatives appointed based on their scientific and research experience, two of whom must be from non-German institutions, and one representative from each university with a joint-appointment cooperation agreement in place with GESIS.177 The second group consists of the President of GESIS, the head of GESIS’ administration, a GESIS staff representative, and the UAB and SAB chairpersons.

The UAB currently consists of ten members who are appointed to four-year terms by the General Meeting of Members (Mitgliederversammlung), representatives of German universities and professional organizations (GESIS-Kuratorium, 2021). The UAB members are social science researchers and professors at ten German universities.178 The role of the UAB is to advise GESIS on user considerations related to proposed service developments and changes. The chair of UAB also has an advisory-only role in the Kuratorium. The SAB currently consists of eleven members appointed to four-year terms by the Kuratorium (GESIS-Kuratorium, 2016). The current SAB members are affiliated with German, Dutch, British, Swiss, Swedish, and American universities and social science institutions.179 The SAB assists the Kuratorium in advising the President of GESIS, in accordance with the Leibniz Association Senate’s

177 Currently, these are: RWTH Aachen University, Universität zu Köln, Universität Mannheim, and Heinrich Heine Universität Düsseldorf.
178 Currently, these are: Universität Potsdam, Ruprecht-Karls-Universität Heidelberg, Universität Konstanz, Universität Bielefeld, Katholische Universität Eichstätt-Ingolstadt, Universität Bamberg, Eberhard Karls Universität Tübingen, Leibniz-Institut für Bildungsverläufe e.V. an der Universität Bamberg, Universität Trier, and Justus-Liebig-Universität Gießen.
179 Currently, these are: University of Essex, Leibniz-Institut für Bildungsverläufe, Universität Salzburg, Annenberg School for Communication University of Pennsylvania, Universität Passau, Statistics Sweden, FORS, Hochschule für Politik München an der Technischen Universität München, Nivel (Utrecht), and Justus-Liebig Universität Gießen.
recommendations. The chair of the SAB is an advisory-only member of the Kuratorium. The SAB and the UAB conduct at least one audit of GESIS between the Leibniz Association’s seven-year evaluation.

There are three different data protection authorities that GESIS is directly responsible to: the Federal Commissioner for Data Protection and Freedom of Information (Bundesbeauftragte für den Datenschutz und die Informationsfreiheit) (BfDI) and the state-level authorities: the Landesbeauftragte für Datenschutz und Informationsfreiheit Nordrhein-Westfalen (LDI-NRW) and Landesbeauftragte für Datenschutz und Informationsfreiheit Baden-Württemberg (LfDI-BW). The three authorities provide information and guidance on data protection for the institutions under their purview as well as the general public. All three can also issue fines for violations of the GDPR. For example, in 2021, the LfDI-BW issued a €300,000 fine against the VfB Stuttgart football club for an Accountability Principle (Art. 5(2) GDPR) violation after personal data transfers to a third party between 2016 and 2018 (Der Landesbeauftragte für den Datenschutz und die Informationsfreiheit, 2021). The LDI-NRW issued a €200 fine and ordered a private citizen to remove YouTube video compilations of dashcam videos that showed identifiable license plate numbers, a violation of Articles 5 and 6 GDPR (Sascha LKW Fahrnünftig, 2020).

4.4.1.5 Legal frameworks for GESIS

In addition to the GDPR and the Bundesdatenschutzgesetz (BDSG), there are a number of additional laws and guidelines that govern GESIS’ operations. These are: the Basic Law; the state-level data protection laws of North Rhine-Westphalia (DSG NRW) and Baden-Württemberg

---

180 This fine is listed on several GDPR fine trackers, including the DS-GVO Portal (https://www.dsgvo-portal.de/), though the LDI-NRW press release about it is no longer available online (including through the Wayback Machine). However, the recipient of the fine has posted a publicly available YouTube video discussing the fine and his perspective on it.
(LSDG); the Federal Statistics Act (Bundesstatistikgesetz) (BStatG); and, the Federal Registration Act (Bundesmeldegesetz) (BMG). These will be discussed below. The first of these is Article 5(3) of the Basic Law, which states that “arts and sciences, research and teaching shall be free.” This freedom means that, while GESIS is funded by both federal and state funders and is considered a public entity, it has the freedom to conduct research without government intrusion (Starck, 2006). For this reason, when GESIS was working through their adaptations to the GDPR, there was no external oversight or direction from government funders.

The Basic Law’s rights to dignity (Art. 1) and personality (Art. 2(1)) have been the foundation for German data protection laws since 1977 (Section 2.3.3). However, the responsibility for protection does not just lie with the Federal Commissioner for Data Protection and Freedom of Information (Bundesbeauftragte für den Datenschutz und die Informationsfreiheit) (BfDI), the national German data protection authority, and the BDSG, which regulates federal entities (including the federally-funded GESIS). Each German state is responsible for all other data protection at the state level, for both the state-funded public entities (including the state-funded GESIS) and private entities. GESIS is headquartered in Baden-Württemberg, but there is a second location in North Rhine-Westphalia; therefore it is responsible to both the LfDI-BW and the LDI-NRW as well. Each state has its own data

---


182 Grundgesetz für die Bundesrepublik Deutschland, Artikel 5 Abs. 3: „Kunst und Wissenschaft, Forschung und Lehre sind frei. Die Freiheit der Lehre entbindet nicht von der Treue zur Verfassung.“ (Arts and sciences, research and teaching shall be free. The freedom of teaching shall not release any person from allegiance to the constitution). Translation by the German Bundestag.
protection law adapting and supplementing the GDPR; GESIS must negotiate compliance with both of these laws (the LSDG\textsuperscript{183} and DSG NRW\textsuperscript{184}, respectively).

Baden-Württemberg’s 2018 LSDG does not have many significant differences from the BDSG that would affect GESIS operations. The DSG NRW, however, does differ more distinctly from the federal law; two differences are significant for GESIS. The first is that the DSG NRW defines anonymization (§ 4) in line with the definition of “factual anonymization” from the BStatG (discussed below), while the LDSG does not revise or supplement any of the GDPR definitions; as a reminder, neither the GDPR nor BDSG clearly define anonymization. The second change says that a data protection impact assessment (DPIA) does not need to be completed if one had already been completed before the GDPR, and the processing has not changed.\textsuperscript{185} This differs from the BDSG, which only identifies the obligations for completing one. The LDSG does not discuss DPIAs in any form; therefore, only what is in the GDPR applies. This is relevant for GESIS because, as will be apparent by its absence in the following case report, GESIS did not complete a DPIA because “it needs to be conducted if you set up a new service. And, since [the DPO] is supervising our activities for 15(?) years now, there was no need to go through this procedure after May 2018” (GESIS04).\textsuperscript{186} This reasoning is grounded

\begin{footnotesize}
\begin{enumerate}
\item[185] § 24(1), DSG NRW: “Eine Datenschutz-Folgenabschätzung nach Artikel 35 Absatz 1 Satz 1 der Verordnung (EU) 2016/679 soll nicht durchgeführt werden, soweit diese für eine Verarbeitung, die im Wesentlichen unverändert übernommen wird, bereits von der fachlich zuständigen obersten Landesbehörde oder von einer durch diese ermächtigten öffentlichen Stelle durchgeführt wurde” (A data protection impact assessment pursuant to Art. 35(1)(1) GDPR does not need to be carried out for a processing operation that is unchanged, provided that it has already been carried out by the highest competent state authority or by a public entity authorized by it). Translation by the author.
\item[186] This was a written response to a follow-up question to the original interview.
\end{enumerate}
\end{footnotesize}
very specifically in the DSG NRW. As well, due to the nature of the external research GESIS conducts across Germany (Section 4.4.1.3), the laws of the other states (and interpretations of those laws) may also be points of contention during that research (Section 4.4.6).

Since the 1987 BStatG, the concept of “factually anonymized data” (faktisch anonymisierte Einzelangaben) has applied to all personal data in Germany (see Section 4.4.1.1 for the full definition). This conceptualization of anonymization has been in use since the 1987 BStatG, is retained in the most recent 2016 BStatG, and is the definition of anonymization that the DSG NRW includes but the LDSG does not. The challenge of what are “anonymized data” will be discussed in depth in Section 4.4.6.

GESIS conducts or collaborates on national and international surveys (see Section 4.4.1.3) which rely on large samples from across the sixteen federal states. These samples are drawn from municipal-level population registers and include names, addresses, gender, nationality, marital status, and other personal information. This is possible because of the 2013 Federal Registration Act (BMG), which allows for information about a group of people in the municipal register to be provided for matters of public interest, including participation in research such as the German General Social Survey (ALLBUS) or the German wave of the European Values Survey (EVS).

### 4.4.2 Data archiving in Germany

GESIS is the largest social science research infrastructure in Germany and in Europe in terms of people and involvement in various projects and services in addition to its data archives.

---

Its stated goal is to be “the most important infrastructure institution for the social sciences in Germany and one of the leading infrastructure institutions for the social sciences in the world” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.e). Of the 474 German institutions listed in the Registry of Research Data Repositories, 86 focus on the social sciences (re3data.org, 2022). As a research infrastructure, GESIS supports researchers at all phases of research, not just the data archiving stage. There is no legal mandate for researchers to deposit their data at GESIS, though GESIS is contracted to collaborate with and archive data from certain federally-funded projects. GESIS10 describes the research data landscape in Germany as “very scattered,” with “many research data centers at different institutes.”

GESIS only archives quantitative data (any qualitative data at GESIS are part of otherwise quantitative data sets). The University of Bremen hosts the qualitative social science data archive Qualiservice. The reasons for this strict division and “job sharing” are historically and methodologically motivated (GESIS02, GESIS04). After heavy censorship and disruption under the Nazi regime, the late 19th and early 20th Century developments in qualitative research for the humanities and social sciences were replaced after 1945 by quantitative methodologies (Mruck & Mey, 2000). This practice continued through the late 1960s, until qualitative methodologies were reintroduced to the West German research community, though there were few places for the resulting data to be archived long-term. From 2003-2005, GESIS and the then-named Archive for Life Course Research at the University of Bremen collaborated on a feasibility study for establishing a data archive for qualitative social science data, resulting in the establishment of Qualiservice (Medjedović, 2011). Since 2018, Qualiservice has been funded by the German Research Foundation to serve as a national qualitative social science research data service.
4.4.3 Privacy and the GDPR in Germany

Reactions to the GDPR in Germany, as reported by the GESIS interviewees, were rather delayed, but when they came, panic, frustration, and confusion were common. This delay was in part because of a long-held perspective that the GDPR was not going to be different from the 2009 Bundesdatenschutzgesetz (BDSG). A common description the interviewees gave of the data protection and privacy landscape in Germany was that, even before the 2016 adoption of the GDPR, it was very strict (e.g., GESIS02, GESIS10, GESIS14). Since the end of World War II, Germany has had a reputation within Europe of being very protective of individual privacy, and the right to privacy is grounded in the first two articles of the German Basic Law. The right to personality (Art. 2(1) Basic Law) has evolved from 1800s Continental conceptualizations of civility and honor, prioritizing respect and reputation as requiring legal protections (Whitman, 2000). While neither “privacy” nor “data protection” are explicitly named in the Basic Law, “there is an implicit acknowledgment of privacy found in provisions aiming to protect individuals’ personality, integrity and autonomy and to establish a separation between society and the state to preserve a free society” (Eichenhofer & Gusy, 2017, p. 101).

The specific German focus on personal data protection as separate from but related to privacy is connected to the actions of the Nazi regime during World War II, which were supported by the availability of population registers with religious identifiers and the computing technology to sort and identify the Jewish population in conquered territories (Section 2.3.1) (Bhaimia, 2018; Black, 2001; Ketelaar, 2020; Seltzer & Anderson, 2001). The experiences under the East German totalitarian regime—in particular, the operations of the Stasi and its

---

188 This prioritization is also seen in the German Criminal Code (Strafgesetzbuch), where Articles 185-200 delineate the legal penalties and procedures for insult, malicious gossip, defamation, gossip and defamation of political figures, defiling the memory of the dead, etc. Strafgesetzbuch (StGB) v. 13.11.1998, BGBl. I S. 3322. Retrieved from http://www.gesetze-im-internet.de/stgb/BINR001270871.html. Translation by Prof. Dr Michael Bohlander.
informants—were also very influential, as even now, German concerns about potential data misuse are aimed more at government institutions than private businesses. DeSimone describes this situation thus: “Since the 1980s Germany has developed a subtle and complex data protection jurisprudence originally designed to protect individual data subjects from rights abuses by market actors. However, the rights aggressor has increasingly been German and European law enforcement authorities” (DeSimone, 2010, p. 291). This perspective is shared by many interviewees. For example:

So, I guess that it does have to do with our history, partly, knowing that data could be misused for political reasons and misused to hurt large parts of the population (GESIS13).

I mean, I'm German. We do have quite some history on authoritarian and totalitarian regimes misusing data to a vast extent. In my view, I studied history. So, at least for me, personally, it has always been a major issue (GESIS04).

The potential for government abuses of personal data before German reunification were not just concerns in East Germany. The first data protection law in the world, the 1970 Hessisches Datenschutz, was a response to the Hessian government’s establishment of a “centralized personal data databank,” the Hessischen Zentrale für Datenverarbeitung (Hessian Center for Data Processing). The German constitutional right to data protection (framed as “a right to informational self-determination”) was established by the Federal Constitutional Court (Bundesverfassungsgericht) in 1983 when it ruled on constitutional complaints about the 1983 census.189 Prior to the 1983 census, West Germans outside of government and academia were not very aware of the West German data protection laws of the previous decade (Frohman, 2012). However, in 1983, the “public sentiment that the census was an unjust and unnecessary

state invasion of privacy led civic groups to file a constitutional complaint challenging the law and to lobby the government to scrap its plans entirely” (DeSimone, 2010, p. 293). The Bundesverfassungsgericht ruling on the complaints established the “right to informational self-determination.” The ruling also made clear that limitations on this right were possible in cases of significant public interest (such as the conduct of the census (a “statistical purpose”)), for which purpose limitations cannot be imposed, though “the collection and processing of such information must be subject to limitations within the relevant information system” (Abs. 3, BVerfG 65, 1).190 This German constitutional right was influential in later iterations of the Bundesdatenschutzgesetz as well as the later European Union-wide Directive and GDPR. This right has been argued to be the foundation of the GDPR’s “Right to be Forgotten” (Art. 17) (Kodde, 2016), though with limitations in recognition of the conflict between freedom of expression (including public interest), itself a right in the Basic Law (Art. 5), and the rights of informational self-determination and personality.

In Germany, thus, despite the reputation that “data protection and privacy are deeply, culturally embedded values” (GESIS11), there remains a tension between these rights and claims of legitimate needs for the data. This tension is exemplified by the 1983 protests against the census, but also more recently, in the public reaction to the German government’s “lack” of response, specifically by Chancellor Angela Merkel and her Ministers, to the 2013 revelations from the Snowden leaks about the NSA’s surveillance program discussed in the previous report. In particular, the German public and opposition parties saw that, “instead of taking initiative to uphold these norms [of constitutional rights, privacy, and data protection], the government was hesitant and instead defended the practices of intelligence agencies and the legitimacy of mass

190 Abs. 3, BVerfG 65, 1. Translation by the Bundesverfassungsgericht.
surveillance” (Schulze, 2015, p. 197). The last decade has witnessed increasing German skepticism about public authorities’ claims to act in the public interest, not helped by the Snowden revelations and lack of consequences to them, and a 2017 survey of German internet users found that “mistrust” was a vital component for that increased cynicism (Lutz et al., 2020; see also Steiger et al., 2017).

The German stereotype of highly valuing privacy does not, however, always translate into privacy-ensuring behaviors. On the one hand, GESIS02 described the changes that have occurred in people’s willingness to share personal information (e.g., addresses, phone numbers, etc.) of and with colleagues or classmates. In the 1980s, for example, this information was shared without much consideration for the potential risk, including to sharing other people’s information with third parties. However, since that time,

I have a little bit more, let's say, restrictive interpretation of privacy with regard to the data of others. I would not share data of others without asking them, this has become a kind of, I think, shared cultural technique right now (GESIS02).

On the other hand, even though several studies both before and after May 2018 have identified high German concerns for data privacy in comparison with other countries, those same surveys have not always identified significantly different rates of, for example, changing privacy settings on social media (European Commission, Brussels, 2018, 2021; Kozyreva et al., 2021; Krasnova & Veltri, 2011; Lutz et al., 2020).

A 2008-2009 survey of primarily university-aged American and German social network users found that German data disclosure behaviors were driven more by privacy concerns (influenced by the amount of control the respondents felt they had) than by enjoyment of or trust in the platforms themselves (Krasnova & Veltri, 2011). While the sample size of this survey was small (138 German and 193 American), this result is not inconsistent with the German
component of the 2015 Eurobarometer 83.1 survey (European Commission, Brussels, 2018; see also Appendix F). In terms of feeling in control of their information online, less than half of German respondents felt they had any control at all, either partial control (39.88%) or complete control (4.96%). At the same time, 68.86% felt concerned about this limited control. When asked about whether explicit consent should be obtained before collecting or processing personal data, of the three possible “yes” answers (where multiple responses were possible), the German responses were: “yes, in all cases” (66.50%), “yes, in case of personal information required online” (19.38%), and “yes, in the case of sensitive information whether online or offline” (19.57%). The spread of “yes” responses varies on the EU level, but in all EU Member States, more than half said that consent should be required in all cases (the EU average for consent in all cases was 70.21%).

Despite these higher-than-average reported concerns about privacy, the several decades of state and federal level data protection laws, and highly visible reports of privacy breaches, the reaction to the GDPR was decidedly mixed and very late. GESIS13 found it “strange for me to see that so many reacted so late, after the 25th. Wasn't there a one or two-year period where people could still adapt to the new regulations?” The GESIS respondents connected much of the confusion and panic about what was necessary to two different problems. The first was a lack of awareness that there already were strong legal protections in place that people and organizations should have been abiding by: “There were huge discussions. And I think, in general, I often had the impression people weren’t aware that we already have a law which is quite strict, and which is quite modern” (GESIS12). Separately, the deeply engrained understanding that the GDPR

---

191 For claims that involve contrasting rates between one of the four countries of interest and the EU excluding the four countries of interest, the smallest sample sizes were 588 (single country) and 14406 (remainder of EU). In this worst-case scenario, the standard deviation of the difference of two percentages is no more than 2.1%, yielding a 95% margin of error of 4.2%. Thus, any difference between two percentages reported here that exceeds 4.2 percentage points is statistically significant.
would not be different from the pre-GDPR BDSG (GESIS02, GESIS10) was very strong (discussed in greater detail in Section 4.4.6). It was not until early 2018 that it was clear that the GDPR and the 2018 BDSG would actually be different, and that organizations and individuals would be affected by them to a greater degree than anticipated. As GESIS02 described,

... when the GDPR was published in 2016, everyone told us, all the experts in Germany told us, ‘this law is very much built on the model of the German Data Protection Law’. So, no one really cared. Yeah, they all told us, ‘it's the same, it's basically the same’. And then it turned out very early in 2018, that this is not the case that there are significant differences between the old German Data Protection Law and the GDPR that was at this time two years old.

There was a lot of panic from individuals as well as small businesses about whether they could still maintain personal websites or how to determine what GDPR requirements applied to them when they themselves did not have the legal expertise to sort out the increased bureaucracy (GESIS10, GESIS15).

Before spring 2018, there was relatively little discussion in the media about the GDPR or what it would mean, but as May 25, 2018 approached, this changed. Newspaper headlines were often very pessimistic, even when the articles themselves were positive. The May 10 headline from the Schwarzwälder Bote, “Für uns kein großes Problem’: EU-Datenschutz-Grundverordnung: Ultimatum läuft endgültig ab,” refers to the May 25 entry into force as an “ultimatum,” with the same implication as in English of penalties for failure to comply (Fritsche, 2018). However, the article itself is a very positive report of the city of Schramberg’s preparedness and ability to protect its residents’ data. Other articles focused primarily on the potential fines for non-compliance. For example, the January 9 headline from the Süddeutsche Zeitung, “Vor diesen neuen Datenschutz-Regeln zittern Unternehmen,” that described businesses as “trembling” (zittern) from the new requirements. The article further called the fines “draconian” (drakonischen Geldstrafen) (Martin-Jung, 2018). Even after the GDPR entered into
force, articles about the GDPR tended to focus more on the penalties than on the data protection, even at the industry level (e.g., the weekly real estate newspaper Hillemacher, 2018).

In hindsight, the late-stage panic in the wider Germany was seen as somewhat of an overreaction:

Well, it did have an immediate impact because of course everyone, probably not only Germany but in Europe, was afraid that the world would change dramatically and that they would wake up one day. ... I still remember that vividly and everyone thought the morning we wake up, the world will have changed, and that of course didn't happen (GESIS13; also GESIS02, GESIS11).

In the end, there was an increased awareness about the importance of data protection and the legal framework for it (GESIS06b). People have become more aware of their rights, especially related to access, portability, and erasure (GESIS01, GESIS15). This is also corroborated by the March 2019 Special Eurobarometer 487a (2021) (see Appendix G), which found that 41.61% of German respondents had both heard of the GDPR and knew what it meant, higher than the EU average (35.85%). It also found that more than half of German respondents knew about the rights of access (72.46%), objection to marketing (68.88%), rectification (67.15%), to be forgotten (68.88%), and portability (57.20%), while less than half (42.99%) knew about the right not to be subject to automated decision making. The Special Eurobarometer 487a also found that the percentage of German respondents feeling partially or completely in control of their online information had increased from the 2015 Eurobarometer 83.1, from 39.88% to 44.91% (partially) and 4.96% to 5.97% (completely). There was a small decrease in concern about the limited control, from 68.86% in 2015 to 59.33% in 2019 (European Commission, Brussels, 2018, 2021).

Opposition to collection and processing of sensitive personal data, including religious affiliation, sexual orientation, and political views, by non-government entities (such as for content personalization) remains very high in Germany (Kozyreva et al., 2021). This 2019
survey found that less than 30% of German respondents considered it “acceptable” to process religious, political, or sexuality information to personalize content, compared with approximately 40% and 50% of British and American respondents, respectively. Of the three countries, German respondents found the collection and processing of sensitive data on online platforms consistently less acceptable, even while the level of concern over privacy issues was very high for all three (82% of German and American respondents and 81% of British). Deloitte (2018) also found that Germans (along with the Dutch and French) were very skeptical that organizations implementing the GDPR took it seriously.

After the GDPR came into effect, thus, German awareness of privacy risks and data protection has increased, and people are generally more skeptical about the intentions behind personal data collection (see also Deloitte, 2018). Bauer et al.’s (2021) 2018 longitudinal German panel survey found that, against the research team’s expectations, respondents’ trust in data collectors after the GDPR entered into force did not increase even as their knowledge of the GDPR did over the three waves of the survey (16-23 April 2018, 24 July-2 August 2018, and 29 October-7 November 2018). However, this increased awareness has not necessarily resulted in increased privacy protection behaviors. Lutz et al.’s 2017 survey of German internet users found that, while the prevalence of a resignation that these behaviors are ultimately useless was situational when it came to respondent cynicism about privacy in online spaces, it was still “capable of preventing users to act on their concerns” and implementing privacy protecting measures at all (2020, p. 1182). Kozyreva et al. (2021) similarly found that despite the aforementioned high level of concern about privacy online (82%), actual implementation of privacy protecting measures fell well short.
4.4.4 Timeline of GDPR response

In 2017, after the GDPR was adopted, the GESIS data protection officer (DPO) developed a roadmap for GESIS to follow to become compliant with the GDPR. However, at the same time, the general opinion at GESIS (as elsewhere in Germany) was that the GDPR was not very different from the 2009 Bundesdatenschutzgesetz (BDSG), and it took a while for the reality that the GDPR and 2018 BDSG would be different to set in (GESIS02, GESIS12). Preparations to follow that roadmap began in early 2017, beginning with a review of GESIS workflows and processes, access policies, software compliance, and starting the changes to documentation (GESIS01, GESIS02, GESIS12). In 2018, when the new German implementation law was announced, the German legal community began raising concerns about possible issues stemming from the differences between the GDPR and the 2009 BDSG. In January 2018, the Leibniz Association held a meeting with the DPOs from most of the institutes to introduce their own expectations about the GDPR.

The redesigned informed consent forms for the GESIS-involved survey projects were implemented to support the March 2018 wave of ALLBUS; this informed consent form could then be adopted and adapted for other GESIS-involved projects. Just prior to the GDPR entering into force, the GDPR-compliant “Privacy Policy” (2018c) went into effect on May 23, and the new “Usage Regulations” (GESIS Leibniz Institut für Sozialwissenschaften, 2018d) for the Data Archive came into effect on May 25 alongside the GDPR. According to GESIS06a, GESIS was compliant by this deadline, though work continued; others at GESIS agree that by summer 2018, GESIS was compliant, though improvements could still be made (GESIS04, GESIS11, GESIS12). In 2019, an internal data protection working group was established. At the end of 2019, the registration database for the DBK (Data Archive) data catalogue and
SowiDataNet|datorium was hacked, affecting approximately 75,000 GESIS users (GESIS Leibniz Institut für Sozialwissenschaften, 2019c). After the hack, additional changes were made to the IT security system and data protection policies.

4.4.5 Priorities in the response

In the interviews, many GESIS respondents framed their discussion about GESIS’ response to the GDPR in terms of how to demonstrate compliance with the requirements. This was connected to demonstrating to data subjects (whether they were GESIS users or data subjects represented in a research data set) and external parties that GESIS was able and would handle the data in a GDPR compliant manner. For example,

… the question immediately came up, ‘Are you GDPR compliant?’ But then, we were able to tell people in the summer of 2018, ‘Yes, we are. We think that the way we are dealing with data is very much the way that we dealt with your data before. And we're not just like, we're not selling it, we're not just putting it on a website without registration. We're not just having it someplace that's easily accessible by third parties’ (GESIS04).

… for the data management after the survey has run, questions are less common. But we would find questions like, ‘how do you treat the personal related data? What kind of encryption methods do you use?’ etc., etc. These are the kinds of questions that we also get from smaller data centers (GESIS01).

They also described demonstrating GDPR compliance through the strict enforcement of informed consent for archiving, because of the importance of transparency with research participants about how their data will be handled. For example,

… earlier, we would have accepted data where it didn't exclude the option to share. And now we rather want to see that they are very open on how they share things. So, it became much more transparent, which we like. But we also feel that there's a lot of hesitation to be transparent, because people think that it encourages… discourages people from taking part. But our feeling is that it actually helps people if you're more transparent, and they know where the data is and where it goes (GESIS10).
4.4.5.1 Priorities and goals development

GESIS’ obligation as a publicly funded research infrastructure was to be compliant with the GDPR and the German laws (GESIS04, GESIS14). However, while GESIS staff did on occasion discuss GDPR-related concerns with other institutions (such as German Data Forum (RatSWD\textsuperscript{192}) committees, other Leibniz Association members, etc.), no external parties other than the DPO directed GESIS’ plans or actions. The DPO provided a 12-point checklist in early 2017 that detailed what the DPO had determined GESIS would need to do to be compliant with the GDPR (GESIS01, GESIS06a). The DPO’s assertion that GESIS was compliant just before May 25, 2018 is based on the final component of that checklist being published (the “Privacy Policy” on May 23, 2018. The DPO contacted GESIS periodically for an update on the progress, and also reviewed proposed changes to contracts, regulations, and other guidance for GESIS and for GESIS’ users (GESIS01, GESIS02, GESIS06a, GESIS13). In addition to the DPO, staff from the then-Data Archive for the Social Science department, the Survey Design and Methodology department, and the Monitoring Society and Social Change department, were involved in planning and implementing GDPR-related changes.

The priorities for the GDPR transition were based on that checklist (also referred to as a “roadmap” by the interviewees). Those priorities were: assessing the IT infrastructure; reviewing and documenting personal data processing across GESIS’ various services; establishing a process register; updating contracts, deposit requirements, and usage regulations; and, establishing new processes for responding to data subject requests for access to or deletion of their data. For the GESIS-involved projects, the informed consent documentation given to

\textsuperscript{192} The German Data Forum (Rat für Sozial- und WirtschaftsDaten (RatSWD)) is an advisory board to the German federal and state government governments focusing on issues related to research data infrastructures for the social, behavioral, and economic sciences. The GESIS RDCs are accredited RatSWD data centres, and several GESIS staff members represent GESIS on various committees.
potential survey respondents was a specific priority due to the timing of the 2018 ALLBUS wave, which would begin before and conclude after the GDPR entered into force (GESIS002, GESIS015).

4.4.5.2 Processor versus controller

The GESIS operations where GDPR’s processor or controller role is influential relate to the research data once published in the Data Archive and SowiDataNet|datorium, the administrative and operational data, and the data from the GESIS-involved projects. For the research data, GESIS is the controller. For this reason, when personal data are found during ingest checking, the data are returned to the data provider (GESIS01, GESIS04), even when the data are deposited under the Archiving PLUS or PREMIUM service packages. According to GESIS02,

For external data, generally, we check the data. And if there is an issue with anonymization in the data, we ask the primary investigator to revise the data set and resubmit. We are not so keen on changing the data ourselves, because we don't want to get the responsibility for mistakes, of course, that could be done at this process.

GESIS is also the controller for the administrative and operational data, including those personal data collected from GESIS account registrations, event participation, job applications, and web analytics (GESIS Leibniz Institut für Sozialwissenschaften, 2021c).

The data from the various projects GESIS is involved in (e.g., ALLBUS, EVS, ESS, ISSP, etc.) are trickier. Depending on the survey, GESIS staff in the RDCs may be responsible

---

193 §5 para. 3 of both the Archiving PLUS and PREMIUM service package contracts states that: “As a matter of principle, the Client shall make the data available to GESIS in anonymized form. In the event of data protection concerns, GESIS reserves the right to suggest and implement further suitable anonymization measures. In this case, GESIS and the Client shall themselves be responsible for the respective processing (within the meaning of Article 4 (7) of the GDPR, where relevant)” (GESIS Leibniz Institut für Sozialwissenschaften, 2021a, 2021b). Despite this, GESIS interviewees indicated that GESIS Data Archive staff would at most possibly make very small changes to data sets in response to anonymization concerns, and if any more than that were required, the data are returned directly to the data provider to correct or enhance the anonymization.
for designing and conducting the whole or Germany-specific portions of the survey, merging multinational responses into a single data set, or preparing the data to be sent to the Data Archive. Whether GESIS is the controller or processor for these data before they are archived (if they are archived at GESIS) depends on the contract between GESIS and their survey partners and collaborators. The RDC staff, who may also be DSS or SDC staff, may find themselves acting first as GESIS-as-processor for data in their projects, then handing the data off to their colleagues who will be the GESIS-as-controller. There is one survey in particular, the European Social Survey (ESS), where what GESIS’ role with the survey would be was a point of contention with the ESS ERIC which had determined that the member countries would be the processors for the resulting data. GESIS13 described the situation with ESS as,

GESIS receives money from the federal government of Germany to conduct the European Social Survey in Germany. … My interpretation always was, we also contract out the fieldwork to a survey agency, we oversee the fieldwork, that we are data controller. We're also the only ones in possession of the addresses. … This was a very long and heated debate we had because along with the idea of controller and processor, there was a subtext that the controller is at the same time the owner of the data. Again, I would disagree and say we still own the German data. … The whole discussion only arose because of the new regulation.

For these surveys, GESIS often works with external fieldwork agencies, and the contracts between the two (or more) parties must clarify which is the processor or controller for liability purposes (GESIS04). However, once the roles are defined, whether GESIS RDC teams are processor or controller has not caused significant problems for the organization.

4.4.5.3 Regulating information flows

The majority of changes GESIS implemented in response to the GDPR were aimed at controlling the informational flows into, out of, and within GESIS. These were changes applied to the acquisition process, to using data from the Data Archive, and to the GESIS-involved
surveys. The only changes to the data acquisition process are evident in the new data deposit agreements, the “Service Package” contracts implemented in 2021 (GESIS Leibniz Institut für Sozialwissenschaften, 2021a, 2021b). The first of these changes is a new requirement that a copy of the informed consent form is included with the data deposit. The data curator reviews the informed consent form as part of the ingest check to ensure that the GDPR-required components are included and that there are no clauses about the survey responses only being used in the project they were collected for; the requirement is also stated in the “Preparation of data transfer” information for data providers (GESIS Leibniz Institut für Sozialwissenschaften, n.d.j). If there are problems with the informed consent form, the form clearly does not allow for archiving or reuse, or if the data provider does not provide the form, the data deposit is rejected (GESIS04, GESIS07, GESIS10, GESIS12). As a result, according to GESIS12,

We are now much stricter with checking and asking, and we really ask people, ‘Please show us your informed consent forms, and if people are not able to show them and prove that they stick to the law, we reject data. And we really check the formulations. Is it, from a legal point of view, possible to hand over data to third parties, to us, to potential reusers? That, as a consequence, we reject more data than before, much more data than before. Because this is one of the most... one of the hardest and most common faults people do when they collect data. … If you make that mistake, your data is dead.

This rejection policy applies even to anonymized data made available in Category 0 and thus, the GDPR does not apply, because GESIS is now strictly applying this policy to the informed consent forms for all deposits. While the information is provided to data subjects to ensure that they are aware of their rights and what will happen to their data, this requirement was also discussed at GESIS in terms of whether the data can be brought into the Data Archive at all.

In addition to the informed consent review, the new data service package agreements also incorporated new deposit review processes that align with each type of data service package. The new ingest checks for the three different service packages do not differ in their focus on data
protection and personal data privacy issues; for all data deposits, the curator checks for informed consent, the scale of regional units, unauthorized use of copyrighted material, size of special populations, etc. (GESIS Leibniz Institut für Sozialwissenschaften, 2022a). The final change to the acquisition process after the GDPR relates to the data provider’s choice of access category. If the data provider wants to impose a stricter access class than GESIS thinks is necessary, DSS staff will try to negotiate a different access class. For example, if the data provider chooses Category C only because they want control over who uses the data but not due to any particularly risky characteristics in the data, DSS staff will recommend Category 0 or A instead (GESIS04). However, this change is only documented in the “Usage Regulations.”

The significant change for data users was to the “Usage Regulations.” The post-GDPR “Usage Regulations” (2018d) is a significant expansion on the previous “Usage Regulations” (2007), with significant changes to two sections. The “General access conditions” section added a new explicit statement that the data user can only store and use the data for the length of the project, and that the personal data which GESIS collects will be processed in accordance with the GESIS “Privacy Policy” (2021c, discussed in Section 4.4.5.4). This section also states that “the data depositor determines access restrictions in consultation with GESIS” (GESIS Leibniz Institut für Sozialwissenschaften, 2018d), which aligns with GESIS04’s acknowledgement above that GESIS staff will recommend a different access class if they feel that the chosen access class is too excessive. The pre-GDPR “Usage Regulation” (2007) said only that the data provider determined those restrictions.

The “Request and access to material” section has enough new rights and obligations compared with the 2007 “Usage Regulation” that it could almost be considered a new section on its own (GESIS Leibniz Institut für Sozialwissenschaften, 2018d). It identifies how the Data
Archive data can be accessed, the rights granted by access class, the data user’s data protection obligations, as well as the process for the data user to request permission to further disseminate data (something expressly forbidden in the 2007 “Usage Regulations”). There are three additional sections entirely new to the 2018 “Usage Regulations” which prohibit any actions which may result in re-identification or individual-level data merging, lay out what a data breach might be, and state that GESIS was not liable to the data user or for the results of the analysis. The Usage Regulations had to be updated so that “for people who demand data from us, we had to make sure that we're pointing to the right articles, or that the legal basis of our working is not the old version of the National Data Protection Law anymore” (GESIS04). There were no changes to the “SowiDataNet|datorium Terms of Use” (GESIS Leibniz Institut für Sozialwissenschaften, n.d.p).

For the GESIS-involved projects, there were two main changes that were made: a documentation change and a change in the relationships with the fieldwork agencies. According to GESIS02 and GESIS15, revising the informed consent forms used for the German national (and German part of the international) surveys was the priority leading into 2018 because the 2018 wave of ALLBUS would be conducted from April to September 2018. More than half of that data collection period had to be GDPR-compliant, and “you cannot start with a consent form that is different from the one you use at the end of the field work, so we had to be ready with the consent form in March 2018” (GESIS02). The changes included updating the data protection regulation GESIS and the fieldwork agency would abide by, and also what rights the survey respondent have regarding their data (GESIS Leibniz Institut für Sozialwissenschaften, 2018a). Once the new consent form was adopted, it could then be adapted to the other surveys. The revised informed consent and information sheets was one of the biggest impacts of GDPR on the
GESIS-involved projects, because of their roles in maintaining the projects’ relationships with the data subjects (GESIS02).

The other main change for all of the GESIS-involved projects was an increase in discussions and coordination with the fieldwork agencies. Prior to the GDPR, GESIS and the agencies did discuss what information should be included in the informed consent forms or the data protection letters given to survey respondents. After the GDPR, these discussions intensified and had a greater focus on making sure all phases of the data collection process were GDPR compliant (GESIS05). The fieldwork agencies and GESIS also had to develop processes for addressing potential survey respondent “Right to be Forgotten” requests for their data in the surveys to be deleted (GESIS13) and negotiate the processor versus controller roles during the fieldwork (GESIS01) more clearly, whether the roles are clearly split or with a mix of responsibilities (GESIS04).

The final change at GESIS related to GESIS’ information flows was to establish a “process register” (Verfahrensverzeichnis), as required by Article 30 GDPR. One of the impacts of GDPR on GESIS in general was an overall increase in staff workload, to which the need to identify and document these processes was a significant contributor. As GESIS01 described,

… when the GDPR came up, then there were much more work to do. And I think the other development and the documentation of certain processes were new to us. … when we started, we had to establish some Verfahrensverzeichnis, all the processing that are necessary to document when you're dealing with personal information.

Identifying and documenting the information flows and data processing is still an on-going effort (GESIS11, GESIS12).
4.4.5.4 Minimizing harm

Three changes at GESIS were intended to minimize harms from data misuse by emphasizing that data subjects had control over their personal data. These are the increased focus on informed consent to archive the data, changes to administrative and operational data, and the creation of a privacy policy. After the GDPR, data providers must include their informed consent forms with their data when depositing data at GESIS (Section 4.4.5.3). If the consent form limits data use to the original research, then the data cannot be archived and published at GESIS. While many GESIS interviewees discussed the much stricter approach to this informed consent from the perspective of how informed consent affected whether a data set could be published at GESIS, they also discussed if from the data subject’s perspective. For example,

… in some cases, we had to refuse data because in the information, it said, like, ‘okay, your data will only be used within the project’, which for a participant, most likely means that, ‘okay, my data will be deleted after the project ends’ (GESIS04).

And there’s always this one point, you find in many, many informed consent forms the formulation, which says, ‘We solely use this data for our own research. We don't give the data to any third parties.’ Which is the end of the line. And then they come with their data to us and say, ‘here's the data’. And when we check this informed consent, and then we say, ‘Stop. You said to your respondents you don't give the data to any other third party. We are a third party, and we give it to other third parties’ (GESIS12).

GESIS09 connected the broader impact of the GDPR on giving data subjects greater control over their own data back to his own view of his work:

… in the case of survey data, again, I mean, I have to rely on the willingness of people to take part in surveys also in the future. And I think it's just fair, from an ethical standpoint. So, it's not a legal standpoint, but from a research ethical standpoint, to tell people what is happening with their information. And, also to help them to understand that it's not something that is really bad for them, because insecurity is happening because people do not know really what is happening with their data.
The primary areas of concern for the administrative and operational data were the GESIS account registration data, participation records for GESIS events, addressing the Right to be Forgotten requests, and GESIS usage statistics. Just as with the research data, informed consent is the primary legal basis for processing these data (GESIS Leibniz Institut für Sozialwissenschaften, 2021c; also GESIS01, GESIS08), though as noted in the “Privacy Policy” (discussed further below), the additional bases of fulfilling contractual obligations (Art. 6(1)(b) GDPR) and legitimate interest (Art. 6(1)(f) GDPR) may also apply. When registering for a GESIS account, the personal data (name, country, discipline area, and email address) collected have not changed. However, where the terms of use acceptance check box on the “Registration Form” previously included an agreement that the user’s personal data would be processed (GESIS Leibniz Institut für Sozialwissenschaften, 2015a) (Fig. 4.19), the post-GDPR registration form specifically states that the personal data may be shared with certain Category C data providers, so that the account holders know what could happen with their personal data (GESIS Leibniz Institut für Sozialwissenschaften, n.d.k) (Fig. 4.20).

Figure 4.19 The mandatory data use and consent agreement in the pre-GDPR GESIS account registration.

Figure 4.20 The mandatory data use and consent agreement in the post-GDPR GESIS account registration.
Another area where GESIS needed to address internal policies and processes was about what to do when a GESIS account holder requested their data be deleted under Article 17 GDPR. The “Privacy Policy” (2021c) tells GESIS users that they have the right to revoke their consent to processing and to have their data erased. A workflow for meeting these requests, which come in several times a month (GESIS03, GESIS12), was developed, but there are a number of ongoing issues that GESIS is still trying to resolve (see Section 4.4.5.7). The final change related to the administrative and operational data at GESIS is to the collection and reporting of GESIS usage statistics. Data providers and funders often request or require usage reports about data downloads or data reuse purposes. However, after the 2019 hack of the GESIS and SowiDataNet|datorium account database (GESIS Leibniz Institut für Sozialwissenschaften, 2019c; Section 4.4.6), GESIS no longer processes and reports these usage statistics because they are based on user account information and other web tracking data. When GESIS staff responsible for this contracted reporting pushed back against this policy, GDPR protections and obligations to data subjects were cited as why these data could no longer be reported (GESIS12).

The final change intended to minimize harm is the creation of the “Privacy Policy” mentioned above (2018c, revised in 2021c). Before the GDPR, there was no specific, separate privacy or data protection policy, though various iterations of the “Legal Imprint” dating back at least to 2010 did include a notification to the user about being able to request data deletion and how web tracking data were used, and that certain GESIS services required personal data (GESIS Leibniz Institut für Sozialwissenschaften, 2010, 2015b). In response to the GDPR, GESIS realized that a separate, more informative policy was necessary; therefore, the “data protection” section of the “Legal Imprint” was extracted, and the new “Privacy Policy” (2018c) went into effect on May 23, 2018. The new policy explicitly identifies the data subject’s GDPR
rights to their data and provides easily accessible links to unsubscribe from GESIS emails.

Minor revisions to the “Privacy Policy” have been made since May 2018, with the most recent iteration in effect since July 2021 (GESIS Leibniz Institut für Sozialwissenschaften, 2021c). With this most recent update, a data protection information sheet (adopted on May 30, 2018) was publicly linked on the “Privacy Policy” website (GESIS Leibniz Institut für Sozialwissenschaften, 2018e); previously, this document was not publicly available. This information sheet details in-depth the legal bases for processing, the data subject’s rights, and the scope and duration of the processing. According to GESIS03, in conjunction with the new “Privacy Policy,” the “reality of what we’re actually doing and what it says in the privacy statement of what we're doing is now probably aligning for the first time” across the various GESIS services and platforms.

4.4.5.5 Response to technological risks

The primary changes made as a response to technological challenges to privacy focused on data protection measures and practices. At GESIS, the changes came in two phases: directly in response to the GDPR, and after the December 2019 hack, the second of which will be discussed in Section 4.4.6. In terms of actual technical data protection measures, there were no new technical measures put in place between 2016 and 2018, because the measures were deemed compliant before 2018 (GESIS01). However, what GESIS did do was formalize and document GESIS’ data protection practices and revise the internally- and externally-provided data protection training materials. According to GESIS01, at the beginning of the process, “the writing down was quite easy.” As the transition process continued, the different needs of the administrative and operational data and the research data limited the extent to which the data protection practices could be fully harmonized across the organization, though GESIS is
harmonizing them as much as possible across the various GESIS services. The content of training materials was revised relatively early (late 2016) to reflect the GDPR; this also rapidly increased the DAS staff’s knowledge of GDPR requirements early on. As GESIS10 described,

… we had to inform ourselves, we had to know much earlier about the law and what will change, because we have these questions from our [training] participants. And back then we still did a lot of RDM [research data management] training. And we had a lot of these questions. So even if we didn't know the answer, all the time, we would look them up afterwards and give some people the feedback later on. So, we kind of knew the law very early.

New rules were also put in place for working with administrative data, including best practices for working while on public transport (such as when commuting between the Cologne and Mannheim locations), sharing employee or applicant information, and using encrypted communication tools for sharing data with external entities (such as the Kuratorium) (GESIS01, GESIS07, GESIS08, GESIS13). GESIS began using a job application platform for hiring purposes where staff can be invited to review an application, instead of forwarding the information over email (GESIS07). These new rules were intended to reinforce compliance with GDPR and to encourage GESIS staff to always be thinking about privacy and data protection.

GESIS IT staff also reviewed the third-party applications that GESIS staff used (such as Dropbox, Zoom, etc.) which were located outside of the European Union (GESIS01, GESIS12). As part of this review, GESIS collected the services’ data protection policies and discussed how to abide by them with the DPO. There are also recommended European Union- or Germany-based services for GESIS staff to use as an alternative, such as the Nextcloud file hosting implementation in place of Dropbox or Google Drive (GESIS01), but there is no formal list of restricted applications.
4.4.5.6 Other changes

There is one additional change GESIS adopted in response to the GDPR: the establishment of a data protection working group. The working group was established in 2019 and comprises two people per department. GESIS does have an external DPO to consult for a fee; during the GDPR transition period (and after 2018), consultations with the DPO increased. Thus, in 2019, GESIS “founded the data protection working group to resolve issues that we can. I’m not part of the group, but that group can resolve on their own and we don’t need the lawyer, don’t need to pay for [the DPO]” (GESIS10). While this group is developing internal data protection expertise, if GESIS staff bring a question to the working group that the group could not answer, then the DPO is brought in (GESIS10, GESIS13). The task of the working group is:

Essentially what we do is either bring our own topics that we feel are important to be discussed in the group, where people in my team or in my department contact me and say we've got this and this issue we would like you to bring it to the group, and then report back how you would handle it. ... And then there was and there still is a need to standardize different processes associated with data protection within the institution, and this is another important and ongoing task of the team (GESIS05).

4.4.5.7 Changes remaining to be completed

In general, the GESIS interviewees consider GESIS to be compliant with the GDPR and other legal obligations, although there are changes that remain unresolved. These changes were framed more in terms of “maintaining” GDPR compliance over time, rather than a matter of “if we do not do this, we are not compliant.” To that end, there are several areas they identified where work is still required which will be discussed below: harmonizing policies and practices, revising data access categories, and hiring a data protection coordinator. As noted in Section 4.4.5.3, a priority for the GESIS-involved projects was developing a GDPR compliant consent form in time for the 2018 ALLBUS wave. This consent form could then be adapted by the other
GESIS projects. The actual harmonizing of this form and its companion information sheets is still on-going. For the Data Archive processes, GESIS12 sees a need to standardize and streamline the ingest and data curation processes for the Data Archive. Data providers have several modes available to share their data with GESIS: Cryptshare transfer, use of a data provider’s secure download server, or by mail. Until the data curators actually look at the data, they do not know whether the data are sufficiently anonymized. After ingest, for both the GESIS-involved projects transferring their data to the Data Archive and external projects, GESIS12 thinks that the curation and ingest checking processes still require harmonization across the various data sources, because “the reality [of the work] is much more dirty.”

There are several unresolved concerns about how to address data subject “Right to be Forgotten” requests. The first of these is establishing a centralized database for all GESIS services where a GESIS user can readily be identified and all information about them deleted across the organization. While there is now a centralized GESIS account database, GESIS has many different services who are not interconnected. So it might be that the person who uses services our department is responsible for comes to us and ask me ‘please delete my personal data’ and then we say ‘yes, no problem. We do that. But did you use any other services within GESIS?’ ‘Well I don’t know’. Normally users, they come, they used 10 years ago with a special service from department X. They don't remember it. … But yeah, this is a real effort, if you really want to fulfill this (GESIS12).

Often these requests do not go to some overarching GESIS-wide contact point but to the specific service, which is a much easier request than when you are saying, okay, I want to be deleted from all the databases of GESIS, which is which, which starts with a problem like, ‘oh, how many databases does GESIS have in the first place? And who can we even talk to regarding them?’ (GESIS03).

To mitigate this issue, GESIS is currently working on a single sign-on system to connect a single user with the different services in a more centralized manner.
Another issue GESIS needs to address is what to do if the data subject in a research data set contacts GESIS about deleting their data already in an anonymized data set. If the data are pseudonymized, then there is a process in place, but “if we can’t identify individual [in] a data set anymore, we can’t delete your data quite simply” (GESIS12). The BDSG (2018) does limit this right
to the extent that these rights [including also the right to access, and restrict or object to processing] are likely to render impossible or seriously impair the achievement of the research or statistical purposes, and such limits are necessary for the fulfillment of the research or statistical purposes (§27(1)).

However, GESIS12 and GESIS15 feel a process should be in place for how to handle this situation if or when it occurs. And then, once these processes are established,

one of the key points on GDPR is not just the doing of the thing, to clean up, to pay attention to data, but the documentation of the doing of the thing, so that not only do you store the data in the correct place, and separate identifiers or whatever else might be called for. But you actually document that activity. GESIS is not as strong on that as we could be (GESIS11).

The Data Archive access categories are also currently being revised. The current access categories (0, A, B, C) have evolved over time since the 1960s, due to the need to convince data providers to share their data by giving them greater control over who could use the data for what purpose. The Category C data requiring the usage contract, for example, was established for that reason, but

at the same time, this contract was also used to make the data more restrictive, for more sensitive data. So, it kind of has a double purpose. And that's why you cannot say all the data that is given out with a contract is sensitive data, some people still use it to track the data use (GESIS10).

However, between the licenses for the data and the outdated access categories (e.g., Category B is no longer available for depositors even though it is still listed in the various contracts and regulations discussed throughout Section 4.4.5). New categories which incorporate the usage
licenses have been developed and are currently being reviewed by external legal advisors; what they will be is not currently publicly available (GESIS12).

The final outstanding change is a staffing need. GESIS has an external data protection officer and an internal legal advisor, in addition to the new data protection working group. However, as GESIS staff were working through their requirements, they realized that they needed someone to specifically specialize on the data protection issues inherent in the research data:

the head of the IT department and he has the contact with the external data protection officer, because before GDPR was mostly the IT department that was dealing with these issues ... and then it turned out that we lacked somebody that also had an eye on the research side, this is where the coordinator will come in (GESIS04).

This would be especially useful for the GESIS-involved projects, because they are very important and very complex, because the field work is done in many different countries, which means many different legislations. Mostly we do the data management, the central data management here in Germany, and then we need to be sure that what we are doing is legal (GESIS12).

Hiring for this position is on-going.

4.4.6 Challenges at GESIS

Going into the GDPR transition period, many interviewees did not think much would need to change. However, not only were there the changes discussed in the previous sections, but there were a number of challenges GESIS faced while adapting to the GDPR, and in the years since. These challenges can be classified into three categories: challenges stemming from the data protection legal regime, challenges for GESIS operations, and the response to the 2019 hack of GESIS’ registration database. There are a number of challenges that the interviewees connected back to the GDPR and the German data protection regime. They are: the mindset that
the GDPR was just a replica of the 2009 Bundesdatenschutzgesetz (BDSG), confusion about how the GDPR defined anonymization, confusion about how the GDPR applied to GESIS, and navigating different state-level data protection laws and data protection authorities. The idea that there would be no significant changes in Germany (much less at GESIS) because of the GDPR stems from how strict the previous German data protection law had already been and the widespread assumption that the GDPR was just the German law applied on a European Union-wide scale (GESIS02, GESIS10, GESIS14). This assumption was based on initial German legal advice about the GDPR:

… when the GDPR was published in 2016, everyone told us, all the experts in Germany told us, ‘this law is very much built on the model of the German Data Protection Law’. So no one really cared. Yeah, no one looked at it, they all told us, ‘it's the same, it's basically the same’. And then it turned out very early in 2018, that this is not the case that there are significant differences between the old German Data Protection Law and the GDPR, that was at this time two years old (GESIS02).

In many ways, this perspective at GESIS mirrors the relative lack of public concern about the GDPR in Germany until 2018 (as discussed in Section 4.4.3), and is why GESIS staff felt unprepared once they realized just how much the GDPR would end up changing (or, at least, having to reconsider at greater detail than expected) (GESIS02, GESIS04, GESIS12).

From there, the revised definitions (or absence thereof) from the Directive to the GDPR were also a problem. Uncertainty about how the terms were defined and could be interpreted for GESIS’ practical operations slowed the transition work, and those definitions are one aspect of the GDPR that need further clarification (GESIS04, GESIS07). This is particularly complicated when it comes to anonymization, because “the GDPR is not very explicit in what kind of anonymization you actually want to do, [and] true anonymization doesn’t really exist” (GESIS03). The Directive never defined just what “anonymization” meant beyond the
Preamble’s statement that the Directive’s principles “shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable (Para. 26). It was not until the GDPR’s Recital 26 that a definition was actually provided (Article 29 Data Protection Working Party, 2014). Therefore, the Member States did not themselves have a definition to transpose into their post-Directive data protection laws.

However, since the 1987 revision of the BStatG, the concept of “factual anonymization,” where data are considered “anonymized” if re-identification requires extraordinary levels of time and effort, has been applied to the data archived at GESIS (see Section 4.4.1.5). It is a “looser” conceptualization than the GDPRs description of anonymous data as those “rendered anonymous in such a manner that the data subject is not or no longer identifiable” (Preamble para. 26, GDPR). Initially, there was little concern at GESIS and other research institutions about their factually anonymized data. However, while the 2009 BDSG defined “rendering anonymous” in accordance with BStatG definition,194 the 2018 BDSG (the GDPR implementation law) does not include a definition for “anonymous” even though several sections refer to data being rendered anonymous. Because of this change between the pre- and post-GDPR BDSG, it was not clear whether factual anonymization was still allowed. If the answer was “no,” it would be significantly disruptive for research institutions like GESIS, for whom a large percentage of archived data were made available without restriction because they were considered factually anonymized.

194 § 3(6), BDSG (2009): “Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmmbaren natürlichen Person zugeordnet werden können” (“Rendering anonymous” means the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual). Translation provided by the Council of Europe.
According to GESIS02, a representative from the federal data protection authority, the BfDI, gave a presentation to a RatSWD committee about the implications of the GDPR on research data infrastructures, and then he told us that the concept of factual anonymization which is the basis for all data access that we provided in the research data centers is not part of the GDPR anymore, it doesn't exist. ... And so, the guy from the data protection office said, ‘you cannot treat your data as factual anonymized data’. So, in this moment, if he would be right, this would mean that we had to shut down our download facilities and to provide the data only with usage contract, and with always secure data facilities. So, for GESIS, this is a problem, because we have roughly 70,000 downloads a year, and we would not administer 70,000 usage contracts a year without a steep increase in resources. And we were a little bit shocked.

The data this would have affected were the so-called “scientific use files” which were factually anonymized and made available to registered GESIS users with any sort of contract regardless of geographic location; they would instead be considered “personal data” under the GDPR’s purview. Had factually anonymized data no longer been allowed, non-European Union data users (approximately 75% of annual data users) would no longer be allowed to access these data, because of the difficulty inherent in sanctioning personal data misuse outside of the EU (GESIS02, GESIS012). This question was only resolved in July 2018 when the RatSWD re-affirmed anonymized data as those where “re-identification may reasonably be deemed unlikely, taking into account the current (technological) conditions” based on a re-identification risk assessment (Rat für Sozial- und Wirtschaftsdaten (RatSWD), 2018), in line with the 2016 BStatG.

In addition to this confusion about whether these data were actually personal data, there was broader confusion about just what GDPR meant for GESIS. Tied closely to the early assumptions that GDPR would not change anything at GESIS, there was also the challenge of figuring out what this law targeted more at commercial data use meant for them (GESIS05).
GESIS04 described that period as: “People were really uncertain and a bit unnerved, because of GDPR, because it took quite some time for everybody to get their heads around what actually things are meaning in this regulation.” The effect of this uncertainty, on the broader scale as well as at GESIS, is that,

We still are uncertain in some areas. The uncertainty has the problems that you tend to be more strict than you probably have to be because you don't want to violate the law and you don't want to maybe face some of the fines because the issue that the GDPR did change in relation to, or in contrast to the old German law, was that the fines that could be called for when violating the regulation are quite high. For us, it would be, of course, a trauma to have to pay, I don't know, €1 million (GESIS13).

The final challenge GESIS faced connected to the legal regime relates to the state-by-state responsibility for data protection in Germany. While the BfDI is responsible for enforcing the GDPR and the BDSG, each state has its own data protection authority, and its own data protection law applicable specifically to public entities in that particular state, of which GESIS is one for both North Rhine-Westphalia and Baden-Württemberg (see Section 4.4.1.5). GESIS thus is directly responsible for abiding by both data protection laws, in addition to the Bundesdatenschutzgesetz, the GDPR, and other laws. Doing so requires GESIS and its DPO to navigate which state law to follow in a given situation. This can lead to situations such as,

if someone comes and says, ‘Okay, this factual anonymization doesn't exist anymore,’ I can say ‘it still exists, it’s in the North Rhine-Westphalian law’. And then they could say ‘okay, but your association GESIS, it's based in Mannheim, Baden-Württemberg, so maybe the law is not relevant for you’. And I can say ‘yeah, but I'm based with my service in Cologne. So, it is’. So, these kinds of discussions are possible (GESIS02).

In addition to differences between those two laws discussed at the start of this case report, GESIS must also contend with the laws and interpretations in the other federal states. This is because GESIS conducts or supports many Germany-wide surveys. A specific problem has already arisen for these surveys based on differing interpretations between the North Rhine-
Westphalian law and that of another state. GESIS was contacted by another state’s government representative in response to a potential survey respondent who had decided not to take part in a national survey who was upset about receiving a reminder letter about the survey. GESIS was directed to cease sending reminder letters because the second letter was viewed as a possible GDPR violation for not having a proper legal basis. GESIS13 finds this directive to be potentially disastrous for research, because

[many] respondents only respond after a first or second reminder. Our samples would be very much biased towards those people who are well organized, have a lot of time on their hands, maybe have more fun than others in participating in surveys. ... Also, that the interpretation that no response is automatically a conscious negation of wanting to take part is also, of course, questionable...

The issues of different state-level laws and interpretations are something the Datenschutzkonferenz, consisting of the state and federal data protection authorities, aims to address.195 When there are conflicts between two state-level laws and consensus cannot be reached,

you have to balance are you located in that area where that data protection of authority is competent? And then, you should think about where you follow that opinion of that competent data protection authority, or would you say, ‘No, I like the opposite opinion of another data protection authority, and for that reason, I don't follow the competent data protection authority’. At the end of the day, I'm quite sure, if you have different opinion between German data protection authorities in one question and you decide to follow the opposite opinion of another authority, the competent authority wouldn't be able to fine you... if they say, ‘Oh, I have that opinion and you get a fine of 1 million or whatever’. It will not work. In those cases, we can be happy that there is maybe a different opinion about it (GESIS06b).

195 From §3 of the Beschluss der DSK vom 05.09.2018: “Zur Erreichung gemeinsamer Positionen strebt die Konferenz Einvernehmen an” (Datenschutzkonferenz, 2018). (In order to achieve common positions, the Conference strives for consensus) (Translation by the author).
While in terms of “general” GESIS operations, the North Rhine-Westphalian and Baden-Württemberg laws have not caused significant GESIS-wide problems, for the GESIS-involved surveys, problems do arise periodically.

The second set of challenges were experienced from the operational perspective. The first of these stems from a lack of resources and expertise. The work at GESIS to address the GDPR requirements was complicated because of the increased workload that come from needing to review and document all processes, revise or formalize policies, or demonstrating compliance to external parties (GESIS01, GESIS05). Had the question about “factual anonymization” not been resolved in 2018, the lack of financial and manpower resources necessary to change the existing infrastructure would have been problematic for GESIS’ future operation as a research institute (GESIS02). GESIS also continues to struggle with a lack of both legal data protection and archive expertise in one person, although the former DPO (who turned the role over to the external DPO in 2008) still works in the Data Archive. This is the motivation for GESIS’ on-going search for a data protection coordinator to focus on the research data concerns.

GESIS had challenges as well enforcing compliance by GESIS staff. This is an on-going challenge involved in maintaining compliance over time. Data protection training, for example, is an annual requirement, but

we realize that not everyone knows these things, and that is sad, even at GESIS, we don't manage to tell the younger people who join that they need an informed consent, for example, like that still happens that it's missing, or that it's wrong, or that there's no awareness for it. So, even in an organization where it's so central, not everyone knows (GESIS10).

During discussions about how the regulations should be interpreted, GESIS01 noted that these discussions could be “sometimes quite heavy and controversial” because staff, especially those
involved in the research projects, had concerns that the research projects and data would be treated like the administrative data, even though there were derogations for research.

Closely related to these compliance challenges, several interviewees expressed concerns about balancing GESIS’ mission as a research infrastructure with data protection and privacy. GESIS05 described this as:

"It's our responsibility to find good balances between what we are required to do and what we want to do, because we understand the responsibility. But also in order to keep, you know, providing quality data for users so that we can push social sciences further and hopefully, join the modern restoration of improving the world with all modesty."

The perspective that GESIS has about this balance is that data protection is an obligation we as an organization have to fulfill. You always need to find compromises between these different goals. I don't think data protection is more important than freedom of research and the striving for best possible research data. I think we always have to be very careful in balancing those different goals (GESIS13).

GESIS05 sees this as an on-going challenge with no set endpoint. GESIS users have also expressed frustration with perceived increased restrictions not from legal concerns but because of the restrictions on accessing, for example, the new digital behavioral and geospatial data that GESIS has started working with in recent years:

"When I go to conferences, and I present myself as a member of the research data center, I'm often blamed for the limited access to regional data that we offer. The researchers don't bother with the legal situation. They don't care. They say to me, ‘you at GESIS, you're so restrictive, we have to come to Cologne to analyze the data’. And I say, ‘Yeah, but it says this is a legal situation’. ‘No, you're just conservative. I think the law is not as restrictive’. I have to defend myself a lot for that, even from persons that are involved in the institutional process of generating the data (GESIS02)."

Data providers also can get angry when issues related to informed consent prevent archiving (GESIS10). The restrictions on processing GESIS usage statistics previously discussed (Section 4.4.5.4) fall into this category as well. GESIS has contractual obligations towards project
funders as much as it does to GESIS users, and even when reports are only aggregated download
statistics, GDPR is cited as why that processing is not permitted.

The last challenge GESIS is still dealing with related to its operations is how to handle
new types of data, specifically the digital behavioral data. A major challenge with the digital
behavioral data is what to do with these data. Logistically, the data sets are often on the terabyte
scale, compared to perhaps a few hundred megabytes for survey responses (GESIS14).

Discussions about whether the content of collected Tweets or Facebook posts or just the
usernames or some other component(s) resolve in different ways depending on the platform and
what the terms of use require at a given time (GESIS14, GESIS15). For example, unlike best
practices for anonymizing or archiving survey data, there is not a lot of guidance available for
the digital behavioral data:

For the social media data, we have so many forking paths, that's another term
some use in statistical analysis, that we can decide, ‘do we exclude variables?
Do we transform them? Do we aggregate them? Can we share these data at all?
Can they exist on the individual level?’ Those are so many decisions that we
have to take there. … For survey data, there's a lot of guidance saying ‘you
should, if you have extreme cases with, for example, income, you should top
code them’, for social media data, we don't have that (GESIS15).

Linking these data to survey data also poses a risk of reidentification, for example, if the research
data subject is also prominent on a social networking platform, then the risks increase and data
use must be more controlled (GESIS04).

The final challenge GESIS faced during this time-period is not actually due to the GDPR
itself. As referenced throughout this dissertation, in December 2019, a hacker accessed and
downloaded the Data Catalog (DBK) and SowiDataNet|datorium registration database containing
the personal data of approximately 75,000 users. GESIS was then blackmailed for €1500 in
bitcoins, which it refused; GESIS followed the reporting requirements and shut down access to
SowiDataNet|datorium (GESIS Leibniz Institut für Sozialwissenschaften, 2019c). According to GESIS12, the hack was possible because the database was an older one and did not have standards and security checks in place; the hack exposed a weakness in the IT infrastructure for the platforms that had not been identified in the previous several years (GESIS01, GESIS12).

After the hack, there were a number of changes GESIS made to its data protection measures. GESIS users had to re-register to access the various services (GESIS04); this had the added benefit that old, no longer active user accounts were removed. While there were already plans to change and centralize the accounts database before the hack, in an effort to prevent something like this ever happening again,

… we are trying to consolidate our services more and more to some common platforms that we can handle much better than every portal has its own implementation, its own software stack, its own special requirements. And well, yes, this is also a lesson learned, I would say, from this hack (GESIS01).

As well, since the hack, “we have a policy that whenever you have a new system online, that there is an external company that has to check your system” (GESIS03, also GESIS12). As discussed previously, the GDPR is cited as why usage data can no longer be processed and reported. However, this policy was not actually changed until after the 2019 hack (GESIS12).

4.4.7 Impacts on GESIS operations

Beginning in 2017, GESIS updated and created anew policies and practices in response to the GDPR. This section will examine what, if any, impact these changes had on the use of GESIS as a data archive based on observable trends in the usage data and reported by the GESIS interviewees. As discussed elsewhere in this case report, after the 2019 hack of the GESIS account registration database, data set download statistics were no longer reported after 2018. From 2010 to 2020, there was a steady increase in data sets available through the Data Archive,
and a slower but steady increase in data sets in SowiDataNet|datorium, the self-deposit archive that opened in 2014 (Fig. 4.21). The vast majority of the data sets in the Data Archive are the Category A data (available only for academic research and teaching purposes) (Fig. 4.22). No Category B (academic research and teaching purposes, results publication require GESIS permission) have been published since 2010.

Figure 4.21 Annual total data sets in the GESIS Data Archive and SowiDataNet|datorium (2010-2021) (GESIS Leibniz Institut für Sozialwissenschaften, n.d.f).

Figure 4.22 Annual total data sets in the GESIS Data Archive by Access Category (2010-2021) (GESIS - Data Archive for the Social Sciences, 2019; Zenk-Möltgen, 2022).
From its inception, SowiDataNet|datorium has had relatively rapid growth in all three access categories: Free access (without registration), Free access (with registration), and Restricted access (Fig. 4.23). There is another category data may be deposited into, Embargo, but the embargo period is a maximum of two years, and the number of embargoed data sets (or which category they were re-classified into at the end of the period) before 2021 is not available. These end-of-year totals are based on the publication year on the SowiDataNet|datorium portal; however, data providers can put the publication year of related journal articles or other publications relations to the data set instead of the year the data are deposited in the archive. For this reason, there are twelve data sets (8 Free access (without registration) and 4 Restricted access) recorded as published prior to when SowiDataNet|datorium opened in 2014.

![End-of-Year Total Data Sets by Access Category](image)

*Figure 4.23 Annual total data sets in SowiDataNet|datorium by access category (2011-2018) (Prinz, 2022b).*

Data set publications in both archives did fluctuate over this period. After 2016, when 245 total data sets were published, the total publications per year decreased by around 50 data sets per year through 2018, when the GDPR entered into force. GESIS02’s perspective on this drop in publications from 2016 to 2018 is that “primary investigators at universities were unsure
about the question whether they could still archive data or not, or transmit data to the Data Archive. So, they refrained from doing so,” though this is not confirmed. SowiDataNet|datorium similarly experienced a decrease in the data set publications from 2016 to 2018, though 2019 and 2020 saw another increase (Fig. 4.24).

According to the interviewees, as quoted above, the GDPR likely had an effect on data provider’s willingness or perceived ability to share their data with data archives such as GESIS. In addition to potential data provider concerns, several interviewees identified the increased strictness about informed consent as one of the biggest changes of the GDPR at GESIS (e.g., GESIS04, GESIS10, GESIS12) that may have affected data provider willingness to deposit their data.

Another measure of GESIS’ operations besides the data sets published is the use of the data. At GESIS, data use is measured through downloads from both archives. Annual Data Archive downloads dropped relatively slightly from 2016 to 2017, before rising again in 2018. The 2016-2017 drop is tied to a decrease of almost 4,000 downloads from Category 0 data (Fig. 4.25), even though downloads from the largest access category, Category A, increased over that
Data set downloads from SowiDataNet|datorium have increased steadily since 2014 (Fig. 4.26).

The other measure of archive use at GESIS is use of the Secure Data Center. In operation since 2014, Secure Data Center use is tracked by number of data users, number of output checks...

Figure 4.25 Annual data downloads for GESIS Data Archive by access category (2010-2018) (GESIS - Data Archive for the Social Sciences, 2019).

Figure 4.26 Annual data downloads for SowiDataNet|datorium (2014-2018) (GESIS - Data Archive for the Social Sciences, 2019).
requested, and total user days per year. There was a slow but overall rise in number of users from 2014 to 2019, with the peak in 2019 at 29 (Fig. 4.27). The numbers did fluctuate a bit in 2016 and 2018, but not significantly. The output checks requested per year also trends upward with the increase in users. For some users, they did not produce any results from their data analysis that they wanted to take with them at the end of their time, and therefore did not request output checking. For others, they produced several different results during their time at the Secure Data Center, or had output rejected and they then went back to their analysis and revised the output; therefore, multiple output checks were requested. Along with the increase in users, the total number of days users were in the Secure Data Center also trended upward over time, with over 100 days in both 2018 and 2019 (Fig. 4.28). In Figures 4.27 and 4.28, the sharp drops in 2020 are due to the COVID 19-related lockdowns that began in March 2020.

Figure 4.27 GESIS Secure Data Center usage by annual number of users and requested output checks (Wiltshire, 2022).

196 The data from 2020 are from January 1, 2020-March 5, 2020, and do not include any users who were already scheduled for the rest of the year before the COVID-19 pandemic lockdowns.
From 2016 to 2018, GESIS experienced an overall increase in data use (as measured by downloads of data sets from the Data Archive and SowiDataNet/datorium and usage of the Secure Data Center). While downloads from the Data Archive did fluctuate over those three years, the total downloads remained over 61,000 downloads each year. GESIS02 described the increased downloads over time as “booming,” and GESIS04 attributes the increase in data use to “users are relying, or they are trusting repositories like GESIS, and others in Germany and across the board, that they do a good job.” While there are no usage statistics after 2018, interviewees did indicate that GESIS data use is still increasing, and requests to use the Secure Data Center have also increased. Between the COVID-19 shutdowns and the renovations to the Secure Data Center, there is a very long waitlist for access (GESIS09, GESIS17).

Overall, GDPR did likely have an effect on data provider willingness to share their data with GESIS, as evidenced by the downward trends of data set publications in both archives in the GDPR transition period (Fig. 4.24), though it did not have a significant effect on data use that can be directly associated with the GDPR or the discussions around it (Fig. 4.25 and 4.26). The main impacts, however, are instead the efforts to document personal data processes, changes to
requirements for data deposit (specifically, the requirement to share the informed consent forms), and the overall increase in awareness across the organization about the importance of data protection. According to GESIS14, “data protection and the legal requirements are a lot more focused in the planning phase of new services around data. It's a lot more present,” a view also shared by GESIS04, GESIS05, GESIS12, GESIS13, and GESIS15.

4.5 Cross-case Comparison

After the GDPR was adopted, all four data archives—ČSDA, FSD, DANS, and GESIS—underwent a period of transition and adaptation to meet the newly imposed expectations and requirements of the GDPR itself and their national (and also, in the case of GESIS, state) data protection laws. While the extent of the changes differs between the four archives, at a high level, there were many similarities in their experiences, whether in terms of changes made, challenges faced, and outcomes. In this section, I will compare the experiences of the four archives in the following areas: timeline and impediments to the response; decisions about archive roles; changes implemented; challenges; and outcomes of the GDPR transition. I will end with a presentation of the major findings of this dissertation.

4.5.1 Organizational timeline and impediments to GDPR adaptation

Within a year of the GDPR’s adoption by the European Parliament in 2016, all four data archives had begun preparing to respond to it. Figure 4.29 presents the timeline of significant events affecting each archive, from the April 14, 2016, adoption to the most recent changes that have taken effect at ČSDA in January 2022. By early 2017, all four archives began reviewing their processes and data holdings to identify potential areas of concern. Two archives had formalized lists of tasks to follow. In 2017, GESIS received a roadmap of tasks from its data
Figure 4.29 Timeline of GDPR-related events at ČSDA (purple), FSD (maroon), DANS (orange), and GESIS (blue) (2016–2022) (timeline developed using the Time.Graphics service (Selected Technologies LLC, n.d.).)
protection officer (DPO) that provided a checklist of tasks it needed to accomplish in order to be compliant. In August 2018, the KNAW’s DPO completed data protection impact assessments on DANS’ three data services—the EASY data archive, NARCIS, and DataverseNL—and the issues identified in these three assessments set DANS’ priorities. Most major formal and informal changes directly related to the GDPR were completed by 2020; however, changes are still on-going at all four archives, with two (ČSDA and DANS) currently transitioning to new archive platforms, and ČSDA only received a GDPR-compliant privacy policy and an affiliated DPO in January 2022.

The ability of the archives (specifically, DANS, GESIS, and ČSDA) to achieve GDPR compliance in a timely fashion was heavily influenced by two external factors: questions about national and state-level legislation, and the role of a parent organization. These two factors affected the archives on an operational level, by increasing the confusion and uncertainty about whether, and how, possible derogations for research use of personal data (allowed by Art. 89(2) GDPR) would affect research and researchers at a national level. DANS, for example, submitted comments about how the then-draft of the Dutch GDPR implementation law (UAVG) did not address whether personal data use for scientific purposes was in the public interest (Doorn, 2017). This uncertainty about what GDPR would mean for Dutch research caused delays that flowed down through the hierarchy of the Dutch research infrastructure. GESIS also faced potentially severe ramifications for its Data Archive because, between the GDPR, the federal Bundesdatenschutzgesetz, and the two state-level data protection laws it must abide by, just what “anonymization” meant in and for Germany might have had to change significantly. The German understanding of “factual anonymization” has been in use since the 1989 BStatG. The GDPR itself does not clearly define what “anonymized data” means, and neither do the BDSG or
the Baden-Württemberg data protection law. Only the North Rhine-Westphalian data protection law defines anonymized data, which it does following the BStatG’s definition. However, given the inconsistencies between the different German states and the lack of clarification at the federal level, as May 2018 approached and the new laws were publicized, it became clear that the question of just what “anonymization” meant after the GDPR was going to be a problem. It was not until July 2018, after the GDPR entered into force, that this question was resolved, and GESIS could confirm that their long-standing interpretation of “anonymized” was still legal. If the resolution was different, a significant portion of the “open access” data at GESIS would have to be restricted in some fashion.

Confusion around just what the GDPR and the Czech GDPR implementation law (ZZOÚ) would mean also created challenges for ČSDA, especially the delays in the ZZOÚ’s adoption. The ZZOÚ only came into effect in late April 2019, nearly a year after the GDPR. The intervening period was one of confusion, as organizations could not fully respond to the GDPR without knowing first what exceptions or derogations the national law might allow for. For ČSDA, this also meant that the Institute of Sociology (SOÚ) was slow to respond to the GDPR (in part because the SOÚ also had to wait for possible direction from the Academy of Sciences of the Czech Republic (AV ČR). Further delays in ČSDA and the SOÚ’s compliance can be attributed to the on-going debate over whether it is the AV ČR or the individual, legally independent institutes which would be responsible for data protection.

Thus, these legislative questions and delays contributed to other delays effecting the data archives that had parent institutions with significant operational input, specifically DANS and ČSDA. At DANS, the confusion about and slow process to finalize the UAVG delayed the KNAW’s own plans and efforts to respond to the GDPR. The KNAW did not finalize its
priorities for itself or for its institutes until after the UAVG was implemented, and even then, the KNAW first focused on its own administrative and operational data. Its focus shifted to the research data only much later in the transition period. DANS was lucky that the new KNAW data protection officer (DPO) was able to complete the data protection impact assessments (DPIAs) for EASY, NARCIS, and DataverseNL relatively soon after May 2018. However, since DANS had to restart the process when the DPOs changed, this meant that significant redundant work and a wholesale pause in DANS’ GDPR response efforts while it waited for the KNAW to catch up. DANS had actually begun their review of systems, data, and processes soon after the GDPR was adopted, and in terms of setting their own expected priorities, DANS was far ahead of the KNAW. Nothing could be finalized, though, until the KNAW established its priorities. Thus, even though DANS is relatively independent in its operations, and indeed, in deciding what it needed to do for the GDPR in addition to the outcomes of the DPIAs, it was also hampered in its efforts because of its parent organization.

ČSDA faced a similar situation with its parent organizations, the SOÚ and the AV ČR. First, because ČSDA is a department within an independent research institute (the SOÚ), decisions about what roles and responsibilities related to personal data were often out of ČSDA’s specific control. For example, ČSDA contributed to the SOÚ’s privacy policy development, but it was the SOÚ’s responsibility to publish it and enforce it. Similarly, the AV ČR and its institutes are still debating whether the institutes need their own DPOs, or if the AV ČR’s DPO was sufficient for the entire Academy. Therefore, for what ČSDA itself was responsible for, the interviewees felt that they were as compliant as they could be, and everything else, they needed to rely on the SOÚ to ensure that the SOÚ was compliant.
While the FSD did not require or receive any GDPR-related priorities or directives from its parent organization(s)—first the University of Tampere, then the merged Tampere University—the parent organization did affect FSD’s experiences during the GDPR transition period. As discussed in Section 4.2, FSD was seen as the expert of its own research data-related needs, so the University felt that FSD was in the best position to determine what was required. However, during this same time period, the University organizational merger into Tampere University did require contributions from FSD, such as further revisions to documentation, websites, agreements, and so forth, to reflect the new legal entity. While this work did not require further changes to data protection and privacy-related policies, it added to FSD staff workloads on top of the day-to-day operations and GDPR-related work.

While legislative considerations and parent organizations presented challenges unique to each archive in their own contexts, they did so in ways that had practical implications on what decisions could be made and when. Therefore, these are two factors that influenced the GDPR experiences of the archives.

4.5.2 Processor vs. controller decision

An important decision that had to be made at each data archive was whether, and for which processes and data, they would be the data processor (who processes data on behalf of the controller) or controller (who determines the purposes and methods of processing and is responsible for demonstrating compliance with the requirements). For the administrative and operational data, all four archives are the controllers (Table 4.1). Only ČSDA, DANS, and GESIS collaborate on or conduct research which produce data, and in general, all three are the controllers for these data as well (ČSDA’s parent organization, the Institute of Sociology (SOÚ) is the controller, specifically, and ČSDA acts in that capacity). GESIS will occasionally either
be a joint controller or a processor for these data; which choice is made depends on the project and any contracts between GESIS, the funder(s), and/or other partners. DANS is the controller during many of the projects; when the data are transferred to the EASY archive, that role changes to processor.

*Table 4.1 Controller (C) or Processor (P) role adopted by each data archive for which type of personal data.*

<table>
<thead>
<tr>
<th></th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archived research data</td>
<td>-</td>
<td>P</td>
<td>P</td>
<td>C</td>
</tr>
<tr>
<td>Administrative/operational data</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Data from archive-involved projects</td>
<td>C</td>
<td>-</td>
<td>C</td>
<td>C/P</td>
</tr>
</tbody>
</table>

The situation is less clear when it comes to the data in the data archive. As discussed in Section 4.1.5.2, because ČSDA considers all of its research data to be anonymous, except for certain data sets from SOÚ researchers which are not available for non-SOÚ reuse, it has not adopted either a processor or controller role for these data. Like ČSDA, FSD has historically only archived anonymous data, or data sets with personal data archived under a “public interest” exemption. Since the GDPR came into effect, FSD has started working with pseudonymized data, which do fall under the purview of the GDPR. For this reason, and because FSD staff may encounter insufficiently anonymized data that mean that FSD is now processing personal data, FSD determined that the processor role was the most appropriate.

The situations at DANS and GESIS are both more complex. The research-affiliated data at DANS exist in three different services. DANS is the processor for the research data in the EASY data archive. The decision to adopt this role was a continuation of the pre-GDPR practice where the responsibility to determine who, and under what conditions, could access and reuse the
research data. This was not a simple decision, however, and is one area where DANS and its parent organization, the KNAW, disagreed. DANS is still debating whether to retain that role or to become the controller instead; the dual role for the projects DANS takes part in is one consideration for this possible change. DANS is also the processor for the data in the DataverseNL platform, because the data are archived, a form of processing, through the DataverseNL infrastructure, while the DataverseNL partners remain the controllers for their data.

GESIS is the only archive of the four to claim the role of controller for the research data in the Data Archive and the self-archiving SowiDataNet|datorium platform. GESIS’ decision to be the controller for the archived research data is intriguing because, as with DANS, GESIS generally leaves the decision about who can access a given data set and for what purpose to the data provider (GESIS Data Archive staff will only intervene to recommend a lower access category if the only reason for restriction is the data provider’s desire for control over the data and is not based on the level of risk in the data set). That DANS does not want to be responsible for granting access to Restricted Access data is the reason that DANS interviewees provided for its choice of the processor role:

DANS doesn't want to be involved in that. And that's also one of the reasons why to choose the data controller or data processor role data, because we don't want to be involved in that. Nevertheless, I mean, we have a certain responsibility, but we don't want to be the person deciding on that. That's something for the depositor should take care of (DANS14).

At DANS, it is thus the data provider’s obligation to demonstrate compliance and ultimate responsibility for the archived data under the Accountability principle (Art. 5(2) GDPR) for the archived data, whereas this is GESIS’ responsibility. GESIS framed its decision to be the controller instead as GESIS “strictly [refraining] from handling personal data, because this would make us a processor. And all of a sudden, we would be in a different legal position”
However, the processor is still obligated to meet the same requirements as the controller regarding data security, and the controller is not restricted from processing personal data. This raises questions then about whether the distinction between the two roles as delineated in the GDPR is as significant in practice, at least for research data archives with personal data, as the GDPR makes it appear.

The decision about what, if any, role the institution adopted for which data was an important decision, and not one that is set in stone. All changes that the data archives made for their services and data were based then on which role applied to the data in question for a given change. This is most evident in the processing agreements that DANS and FSD put into operation. Because GESIS is the controller for its archived data, it determines what processing is done (though this processing is agreed to by the data provider in the given deposit agreement) and does the processing agreed upon, though, as discussed in the GESIS case report (Section 4.4), GESIS does prefer to not do additional anonymization work. Therefore, the only agreement needed is the deposit agreement. However, at DANS and FSD, the data provider, not the archive, is the controller. Therefore, the processors are only allowed to do any processing (including curation, archiving, and dissemination) which the controller authorizes. Thus, the processing agreements.

The processor versus controller decision regarding the other services and data also dictated the changes elsewhere. For example, all four institutions (or, parent organization, in ČSDA’s case) have published a GDPR-compliant privacy policy. For the two institutions which previously had a privacy policy (FSD and DANS), the changes were mainly to update the language and present the information about data subject rights under the GDPR. The content of the new GESIS and ČSDA/SOÚ policies do the same. Because the institutions are all controllers
for the administrative and operational data covered by the privacy policies, they must inform
data subjects who the controller and data protection officer are, as well as about each type of data
collected, for which purposes, and what rights the data subjects have for each type of data.

The decisions here are not final. DANS is the clear example of this, as several DANS
interviewees discussed. DANS may change their role from processor to controller for the
archived data in the future, motivated by limitations of the processor role, the final status of the
transition to the Dataverse-based Data Stations, and what revisions may be made to the UAVG.
Changing this role would lead to changes in the legal agreements and other data deposit and data
use policies, and would increase DANS’ responsibilities to their archived data. ČSDA is also
transitioning to a new Dataverse-based archiving platform and considering future expansions of
its services which may enable wider archiving of and access to non-anonymized data. While the
interviewees did not discuss possible changes to the GDPR role, should one be required, policies
and practices would need to be further revised to reflect that choice. Therefore, the processor
versus controller decision is a significant factor in determining where and how changes to or new
processes and policies to address the GDPR are made.

4.5.3 Changes at the data archives

While I considered the changes made at each of the four archives, I began to sort them
into the different aspects of informational privacy that comprise my theoretical framework. I did
this both by looking at both what the changes actually were (by comparing, for example,
differences in pre- and post-GDPR policy documentation, or identifying entirely new policies or
processes) as well as how my interlocutors at all four archives described the changes and the
motivations behind them. In doing so, I found that the theoretical framework worked well, as the
vast majority of changes fit into one or more of the dimensions, while the remaining changes
reflected a need to manage and enforce compliance with the privacy regulations. Across the four archives, there were a lot of similarities in terms of just what changes were made and where, with the differences tending to lay in the details, determined either by the choice of the processor or controller role for a given data type, or derogations in the national GDPR implementation laws.

4.5.3.1 Regulating information flows

Across the four archives, the greatest consistency in terms of what changes were made reflected managing the flow of data into and out the archives (Table 4.2). Several changes in this category involved bringing existing policies or practices into compliance with the GDPR, for example, updating the text to reflect the new legal terms or legislation or adding information about the processor or controller role. These changes occurred primarily in the agreements required for data deposit and assorted terms of use policies at FSD, DANS, and GESIS (ČSDA made no changes to either of these during this period). FSD, DANS, and GESIS all also began documenting their personal data processing in process register (Art. 30 GDPR); DANS’ reports these processes in their parent organization, the KNAW’s, process register.

All four archives have become stricter about conducting deposit reviews in terms of checking for personal data and requiring specific documentation with the deposited data set. For the most part, this is not a formalized policy change so much as a reflection of the increased awareness that archive staff about the importance of privacy and their obligation to protect the data subjects. GESIS is the only archive which formally changed their deposit review process, though this was not specifically a GDPR-related change. GESIS changed the deposit agreements for both the Data Archive and the self-archiving platform SowiDataNet|datorium in 2021, and there are specific deposit review checks that are only completed for specific agreements based on
cost and effort required. The data protection and privacy components of the GESIS deposit review are the same at all levels of deposit review.

*Table 4.2 Changes or new implementations at the four data archives categorized under "Regulating Information Flows."*

<table>
<thead>
<tr>
<th></th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposit reviews</td>
<td>∆</td>
<td>∆</td>
<td>∆</td>
<td>∆</td>
</tr>
<tr>
<td>Deposit form</td>
<td></td>
<td></td>
<td>∆</td>
<td></td>
</tr>
<tr>
<td>Requirements for deposit documentation</td>
<td>●</td>
<td>●</td>
<td>∆</td>
<td>●</td>
</tr>
<tr>
<td>Deposit agreements</td>
<td></td>
<td>∆</td>
<td>∆</td>
<td>∆</td>
</tr>
<tr>
<td>Terms of use</td>
<td>∆</td>
<td>∆</td>
<td>∆</td>
<td></td>
</tr>
<tr>
<td>Process register</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Processing agreements</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Δ - Changed process, documentation, policy
● - New process, documentation, policy

ČSDA, FSD, and GESIS also now require that data providers provide evidence that data subjects have consented to the archiving of their data, and not just to take part in the data collection. Therefore, the data providers must include a blank copy of the informed consent form with the deposited data; all four archives’ deposit agreements, including DANS, include a provision that the data provider confirms that the data were collected in accordance with all legal and ethical requirements, and therefore that the deposited consent form is the same form the data subjects received. FSD and GESIS require this copy for all data deposits, while ČSDA only requires it in instances where the legal basis for data collection or archiving is informed consent.

Both FSD and DANS implemented a requirement for a processing agreement between the data provider and the archive before the data are deposited. However, they took two different
approaches. FSD requires this agreement for all data sets, based on its experience with “70% of the deposited research data [not being] anonymous although researchers says so” (FSD01). Therefore, it is easier for FSD to simply have this agreement which allows them to process any personal data which may be in a deposited data set without needing to stop the processing to obtain the controller’s permission later in the deposit workflow. At DANS, however, this agreement is only required if the data provider answered “yes” on the pre-deposit “Deposit Form” (separate from the deposit agreement) when asked whether there were personal data in the data set, or if personal data are found even after the data provider answered “no.” This question itself was only added to any agreement at DANS, and the presence of personal data in a data set automatically triggers the requirement for the processing agreement. DANS also requires a processing agreement for all data deposited into DataverseNL, an entirely new requirement reflecting DANS’ formal status as processor for DataverseNL.

4.5.3.2 Minimizing harms

Solove’s (2006, 2008) taxonomy of harmful behaviors identified four groups of activities that may result in privacy harms to data subjects: collection, processing, dissemination, and invasion. My theoretical framework’s dimension, privacy as protection from harm, supported identifying data archive actions that addressed the harms which may result from these activities. Changes at the four archives reflecting the goal of minimizing potential privacy harms to data subjects were primarily related to ensuring that data subjects were aware of their rights under the GDPR regarding the collection, processing, and dissemination of their data, and archive practices that moderate access to data. There was less similarity in what the archives did in this regard (Table 4.3) compared to the changes related to the information flows. Most changes were made at only one or two archives, and there was greater variation in what exactly was done.
Table 4.3 Changes or new implementations at the four data archives categorized under “Minimizing Harms.”

<table>
<thead>
<tr>
<th></th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy policy</td>
<td>●</td>
<td>Δ</td>
<td>Δ</td>
<td>●</td>
</tr>
<tr>
<td>Informed consent for archiving</td>
<td>●</td>
<td>●</td>
<td>Δ</td>
<td>●</td>
</tr>
<tr>
<td>Deposit rejection</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consent for archive-involved projects</td>
<td>Δ</td>
<td></td>
<td></td>
<td>Δ</td>
</tr>
<tr>
<td>Remove or modify already-archived data</td>
<td>Δ</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access categories</td>
<td></td>
<td>Δ</td>
<td>Δ</td>
<td></td>
</tr>
<tr>
<td>Administrative/operational data processing</td>
<td></td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Automatic restricted access</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data protection impact assessment</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Δ - Changed process, documentation, policy
● - New process, documentation, policy

Of all the changes categorized here, the only common target was the privacy policy. FSD and DANS already had existing privacy policies which were revised to bring the language into compliance with the GDPR, while GESIS and ČSDA’s policies are entirely new (because ČSDA is a department within the Institute of Sociology (SOÚ), the privacy policy is actually the SOÚ’s policy). Whether the policies were new or simply revised, they all provide details about the archive user’s rights regarding the administrative and operational data collected about them (for example, personal data provided when the user registers for an account, data provided when
requesting access to restricted data, or collected from event or workshop attendees). The policies all describe the legal bases for the data collection as well as how these data would be used.

As discussed in the previous section, ČSDA, FSD and GESIS each require the data provider to include a copy of the informed consent form as part of the data deposit; both FSD and GESIS include this requirement in guidance for data providers (Finnish Social Science Data Archive, n.d.f; GESIS Leibniz Institut für Sozialwissenschaften, n.d.j). However, the mere presence of the form as a moderator for bringing data into the archive is not its only purpose; the archives do not just verify that a consent form was used. This requirement is included under the “minimizing harms” dimension because archive staff specifically look for whether the informed consent says that the data will only be used for the research project they were collected for, or that the data will not be shared with any third parties, or any other language that informs the data subjects that their data will not be reused by others or for other purposes. The intention of this is to ensure that, as with the privacy policies discussed above, the data subjects are aware of what their rights are regarding their personal data and what will be done with their data. Failure to provide this proof to ČSDA or GESIS may result in the data set being rejected for archiving.

DANS is an interesting exception here, because they did change this requirement in response to the GDPR. However, the change DANS made was to remove the requirement that the data provider provide evidence that the data subjects had consented to data archiving and reuse. This requirement had originally been stated in the pre-GDPR “Privacy Regulation,” but is not in any post-GDPR policy or agreement. The revised deposit agreement only stipulates that the data provider confirms that they collected the data in accordance with the GDPR, but since DANS is “not the controller, we do not need that kind of documentation” (DANS11). At GESIS and ČSDA, changes were made to the consent forms used in the projects both archives take part
in, primarily to utilize the GDPR language about data subject rights and be more transparent to potential survey respondents.

ČSda and DANS also addressed the data already in their archives. ČSda reviewed its data holdings and found data sets with indirect identifiers which were considered to be “anonymous” under the previous legal regime. ČSda staff were less confident about these data after the GDPR, and so they either modified the data to make them “more” compliant, or if they determined that was not possible, the data were removed from the Nesstar catalogue. DANS revised its data access categories from five to two (Open Access and Restricted Access). In so doing, DANS staff have been re-categorizing data archived under the legacy categorizing into the new ones, which has included both making previously restricted data sets Open Access, and previously Open Access data sets Restricted Access. DANS also revised its licensing scheme, with 22 possible open licenses for the Open Access data, and the DANS license, which applies to all Restricted Access data. GESIS is also changing its data access categories, but the change has not been approved or made public yet. One final change at DANS is also triggered, as discussed in the previous section, by whether there are personal data in a deposited data set. That is that all data sets with personal data are now automatically only available under the Restricted Access category. Prior to the GDPR, if the data subjects had consented to their data being made publicly available, the data set could be an Open Access category. However, now, even with that consent, DANS will restrict access to these data.

Whether the archive completed a data protection impact assessment (DPIA) is an interesting difference in how the data archives addressed the GDPR. Of the four archives, only DANS thus far has actually completed a DPIA; with the KNAW data protection officer, it completed one for each of its services (EASY, NARCIS, and DataverseNL) in the fall of 2018.
The results of these DPIAs were used to set the priorities for what DANS would do to become compliant with the GDPR. FSD knows that it needs to do a DPIA, but it is not a priority for the archive; much of the information that FSD provides each year for its internal Data Balance Sheets would be required for a DPIA, so if they are pressed to do one, FSD staff do not feel it would be too burdensome.

Since ČSDA considers its research data to be anonymous, ČSDA staff did not complete a DPIA for the Nesstar catalog or Medard archive. I could not determine whether the SOÚ as the controller completed a DPIA for the administrative and operational data, the SOÚ-produced research data, or the data produced from the various collaborative projects it is involved in (with or without ČSDA); however, the ČSDA interviewees did not mention one for any of these data. GESIS determined that it did not need to conduct a DPIA for any of its services, relying on one of the four different data protection laws which apply to GESIS, specifically, that of the state of North Rhine-Westphalia (DSG NRW). The DSG NRW does not require a DPIA if one had already been completed before the GDPR (there is no such exemption in the other applicable state law from Baden-Württemberg), and GESIS and GESIS’ DPO determined that, since the DPO had been with GESIS for fifteen years and no services had changed in that time, no DPIA would be required. The structure of the German data protection regime (with the GDPR, the national data protection law (BDSG), and state-level data protection laws all applying to different areas) means that, in situations like GESIS, which is based in two different states (and under two different state data protection laws), organizations will have to balance which of the laws to abide by when there are conflicts, since as with the DPIA.
4.5.3.3 Response to technology

The changes reflecting concerns about the technological risks to privacy were primarily found in how the data archives approached data protection and training. Overall, the fewest changes were reported in this category (Table 4.4). FSD, DANS, and GESIS revised their data protection and data management training to reflect the GDPR requirements and introduced GDPR-specific training for their employees. ČSDA developed GDPR-specific training, and, due to its growing reputation as the unit with GDPR expertise within the SOÚ, began providing this training to SOÚ and other Academy of Sciences of the Czech Republic (AV ČR) staff, as well as the wider Czech social science research community. ČSDA, DANS, and GESIS (or rather, for ČSDA and DANS, their respective parent organizations, the SOÚ and KNAW) also implemented restrictions on the use of non-European Union-based 3rd-party tools and applications when processing or working with personal data. These restrictions included recommended local tools and applications instead, such as a parent organization’s own dedicated cloud servers, localized data collection tools, etc., which limit the use of US-based servers and products (e.g., Microsoft Office 365, Google, Dropbox, etc.). These restrictions are not necessarily easily implemented, especially for longstanding multinational European Union or CESSDA projects and collaborations which have used Dropbox or Google Cloud services, but they are a work in progress.
Table 4.4 Changes or new implementations at the four data archives categorized under "Response to Technological Risks."

<table>
<thead>
<tr>
<th></th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account registration database</td>
<td>Δ</td>
<td></td>
<td></td>
<td>Δ</td>
</tr>
<tr>
<td>Recommended list of 3rd-party tools and applications</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
</tr>
<tr>
<td>Data protection training</td>
<td>●</td>
<td>Δ</td>
<td>Δ</td>
<td>Δ</td>
</tr>
</tbody>
</table>

Δ - Changed process, documentation, policy
● - New process, documentation, policy

Both ČSDA and GESIS made changes targeting how the administrative and operational data were handled. At GESIS, the first of these changes primarily focused on documenting and formalizing how these data were handled and harmonizing the data protection practices across its many different services. ČSDA, through the SOÚ, also has new rules in place regarding what specific computers can be used for these data, where the data can be stored, password control, etc. The second change at both archives was to their archive account registration databases. For both GESIS and ČSDA, this was not directly a change due to GDPR requirements, but due to the exposure of a flaw in the data protection system. As discussed in Section 4.4.6, GESIS’ account registration database was hacked in 2019, and in response, GESIS shutdown the affected database and all GESIS users were required to re-register to use GESIS’ services. GESIS also began centralizing its account and services databases, something which had already been planned, and now has a policy that all new GESIS services require an external data security audit. ČSDA was motivated by the GESIS hack to remove its Nesstar catalogue user database.
from the catalog. While ČSDA staff did not think that there was a risk to that database, they chose to err on the side of caution and add that extra layer of protection to its accounts database.

4.5.3.4 Other changes

There were several additional changes made across the four archives which did not directly fit into the privacy framework; these instead focused on managing the data protection processes at the archives. These were all entirely new additions to the archive operations (Table 4.5). A dedicated data protection officer (DPO) was appointed at FSD, an entirely new role, although the person in that position had informally acted as a data protection and privacy advisor before the role was formalized. FSD also hired a temporary legal advisor for this period who had additional expertise working with the archived research data; this position ended in 2020. As a department of the SOÚ, the DPO for ČSDA was appointed by the SOÚ and fills that role for the entire institute. The SOÚ’s parent organization, the AV ČR, also appointed a DPO in 2018, which filled this role for the entire SOÚ as well until 2022. DANS and GESIS already had DPOs before the GDPR (DANS, like ČSDA, uses its parent organization’s DPO). However, both DANS and GESIS determined that they needed additional internal data protection expertise. DANS has two internal staff who also focus on data protection for its various services, though neither of them have legal backgrounds. The staff member in this role for the EASY research data already had this responsibility before the GDPR. However, the KNAW mandated that its institutes establish a privacy coordinator who would focus on the administrative and operational data concerns; therefore, DANS created this position. The privacy coordinator and the advisor for the EASY data collaborate together to address concerns across the organization.
Table 4.5 New roles established to manage data protection and privacy issues.

<table>
<thead>
<tr>
<th>Role</th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data protection officer established</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy/data protection coordinator established</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Internal data protection working group established</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Legal/data protection advisor position established (temporary)</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GESIS likewise realized it needed this additional expertise. This was driven primarily by the fact that the GESIS DPO is an external contracted legal firm which costs money each time it is consulted. Therefore, GESIS has established an internal data protection working group to build greater internal data protection expertise across the organization (the group comprises representatives from every GESIS department) so that when questions arise, they can first be brought to the working group, and are only sent to the DPO if it is something that the working group determines it cannot handle. In addition to this working group, GESIS does have an existing internal legal advisor who is focused on data protection and privacy issues related to the administrative and operational data. But GESIS determined during the transition period that it needed someone who specialized in data protection issues related to the research data, both those data in the two archives and the data from the projects GESIS was involved in. GESIS has established this data protection coordinator role for the research data but has not yet filled it.
4.5.3.5 Implications of the changes

Considering the changes made (or planned) at each archive, it is clear that the archives addressed the GDPR in similar ways, and that the major differences were based on what role the archive chose for a given type of personal data. I considered that decisions across the archives converged when at least two archives made similar changes to the same aspect (e.g., document, policy, or process), recognizing that decisions would rarely be perfectly identical given the different contexts of the archives. The areas of greatest convergence were: archival contracts (agreements for data deposit, usage agreements), policies and processes (data deposit rejection, deposit review, data provider responsibility), and administrative data protection. For the remainder of this section, I will focus on where and why there were still differences in the actions.

As discussed in the previous sections, only FSD, DANS, and GESIS made changes to their agreements for data deposit. The changes to the deposit agreements for all three archives were primarily wording changes that align with the language in the GDPR, including providing more explicit information to the data provider, though there were some more substantial changes to each which were archive specific. The FSD “Deposition Agreement” requires more information from the data provider about the data and documentation which will be archived for reuse. It also replicates parts of the “Processing Agreement” text which had already been agreed to earlier in the deposit process. DANS’ “Deposit Agreement” gives the requirement for the “Processing Agreement,” and also updated the policy regarding personal data in the metadata. GESIS’ new Archiving BASIS/PLUS/PREMIUM service packages were themselves substantial changes from the pre-GDPR “Archive Agreement” (GESIS Leibniz Institut für Sozialwissenschaften, 2014) in that they establish cost and service package-based ingest checks.
and archiving services, but the only significant change is the stated requirement for the inclusion of the informed consent form with the deposit.

The processor or controller decisions only directly influenced the “Deposit Agreements” for FSD and DANS. As processors for the archived research data, both FSD and DANS created entirely new “Processing Agreements” to establish what personal data processing they would do on behalf of the controller. While the deposited data which these apply to differ at each archive—at DANS, the “Processing Agreement” only applies to data sets with personal data, while the FSD “Processing Agreement” applies to all data sets—the content of the two agreements are similar. The DANS agreement is longer and does give more information than the FSD agreement (for example, a page long list of definitions, and other information which FSD instead provides a link to in the PDF version of its “Processing Agreement”). The DANS “Processing Agreement” also is part of a specific workflow for personal data, beginning with the “Deposit Form” and ending with the automatic “Restricted Access” classification of the data.

The “Processing Agreement” and the references to it in the “Deposit Agreement” is one of the main areas of the data acquisition and ingest process which reflect the processor or controller decision.

The other main area is the requirement for proof of informed consent for archiving. ČSDA, FSD, and GESIS all require documented proof that research participants have consented not just to taking part in the given study, but also to their data being archived. ČSDA, FSD, and GESIS archive staff validate this during the deposit review process, where they review a blank copy of the informed consent form (also referred to as “privacy notice” by FSD) to ensure that there is no language restricting data use to the original study only. This is a new requirement at all three archives. If the consent form is not provided, or if the consent form indicates that the
research participants did not consent to data archiving, then the data deposit must be rejected; at
GESIS, this rejection policy applies even if the data in question do not include any personal data.

The main differences in the process, where the data provider is notified of this requirement and when the form is required, is not related to the different processor or controller roles taken by the three archives. DANS, however, interpreted the need to verify the legality of archiving the data differently than FSD, the other processor. FSD determined that it needed to confirm that it was legal for the controller to transfer the data to FSD, by having the controller demonstrate that compliance. DANS instead changed its pre-GDPR policy of requiring that demonstration (as stated in the 2009 “Privacy Regulation” (DANS, 2009)) to one where the data provider confirms through the “Deposit Agreement” that the data were not collected in violation of any legislative restrictions. But DANS, as processor, has decided that it does not need any additional proof, because the liability and responsibility for this, if there are any issues, lies with the controller alone. This difference between interpretations of the processor role and responsibility is the only clearly divergent response to the GDPR, where two or more archives with the same GDPR responsibilities made opposing decisions about the same issue.

The next area of significant convergence relates to archive processes and policies. All four archives, no matter their processor or controller status, implemented a stricter deposit review process. With the exception of the service package-specific deposit review steps at GESIS (wherein all data sets, no matter the service package, undergo the same strict deposit review), the changes were not formalized at any of the archives. The interviewees, even those at GESIS, described the changes as a mindset change. The interviewees considered their existing deposit review processes to be sufficient and effective because they already had to abide by pre-GDPR data protection laws and institutional guidelines, and the changes from the GDPR were a
reinforcement of the importance of data protection and privacy. This reflects a broader, commonly-reported outcome of the GDPR across the four archives: an increased awareness of the importance of data subject privacy as something that should always be at the forefront of work with personal data. It was not new, just reinforced and more prominent. Therefore, during the deposit review, data curators are now stricter with their review for the presence of personal data and, as discussed above for ČSDA, FSD, and GESIS, with verifying that the research subjects consented to data archiving.

ČSDA, DANS, and GESIS all reviewed their archived research data in some fashion to ensure that the manner in which they were archived was GDPR compliant, and made changes to the data or to the accessibility where they were deemed necessary. For example, ČSDA had concerns about whether anonymized data which were archived before the GDPR and the new ZZOÚ met the new expectations for anonymization. If ČSDA determined that a more GDPR-compliant anonymization standard could not be applied to the data, then those data were removed from the Nesstar catalogue and were no longer available for reuse. DANS revised its data access categories to simply “Open Access” and “Restricted Access” and its usage licenses, and in currently working through the archived data to re-categorize data in the legacy categories into the new ones. This project has resulted in the reclassification and re-licensing of data, both from an “Open Access” into a “Restricted Access” setting, and the reverse. GESIS is also re-assessing its data access categories; however, the revised access categories have not yet been approved and are not publicly available. How extensive this change will be, and what affects it will have on already-archived research data remain to be seen.

The next converging policy change relates to who the archives deem responsible for data anonymization. Despite the differing roles, all four archives have reiterated and reinforced the
data provider’s responsibility for anonymization. This is interesting, because it is a response where, given the different role GESIS has taken (as controller) for the archived research data, it would not be surprising were GESIS to have taken on the responsibility of the anonymization work. Interviewees from all four archives described their positions in terms of not wanting to be responsible for this particular type of personal data processing, even with processing agreements in place:

… we communicate with the [non-Institute of Sociology] depositor, and we give the depositor advice, how to anonymize, but in most cases, we don't want to do anonymization because it means that you are working with personal data and we don't want to work with personal data, especially when they are from the outside of the Institute (ČSDA01).

DANS chose to be processor. And deliberately, so almost all of the responsibilities in the GDPR sense or at the controller, which in our case, was the data depositor (DANS02)

Also, now the data must be anonymised by the depositor. Before we often made the anonymisation in FSD (FSD05).

For external data, generally, we check the data. And if there is an issue with anonymization in the data, we ask the primary investigator to revise the data set and resubmit. We are not so keen on changing the data ourselves, because we don't want to get the responsibility for mistakes, of course, that could be done at this process (GESIS02).

Of note, both ČSDA and GESIS have two different approaches to anonymization, based on the source of the data. ČSDA, in its role as an SOÚ service department, will assist with anonymization for SOÚ-produced research data, but will not do so for external data providers. This work falls under the SOÚ’s role as controller for these data. GESIS archive staff will also work with the GESIS-involved projects to support anonymization before data are transferred to the Data Archive, depending on the role that GESIS has taken regarding those projects. FSD did change their practices around anonymization (FSD01, FSD05); before the GDPR, FSD would actively offer anonymization services to support data providers because of the noted frequency
of the presence of personal data in supposedly anonymized data sets. Now, FSD may do some anonymization work if required as part of the “Processing Agreement,” but the expectation is that the data provider should have already done it.

Thus, excepting certain specified situations, for each of the four archives, the data provider is responsible for anonymizing the research data, even though three different decisions about responsibility for the data in the archives were made. The reasoning behind not wanting that responsibility is the same: the archives do not want to be responsible for errors or mistakes in the anonymization, even as they review that work during the deposit review process. Even though, for example, GESIS as controller for the archived data is responsible for demonstrating the GDPR compliance of those data. And, even with that accountability specifically part of the controller’s obligations under the GDPR, the processor is still required to comply with the GDPR even if the accountability principle (Art. 5(2) GDPR) only applies to the controller. This does raise a broader question of whether, at least for research data archives working with personal data, the role of processor or controller really matters? Or, whether there is instead simply a need for a clearer definition of what the processor and controller roles entail in data archives? What role the archives took for which data did dictate many of their later decisions, but in the end, even with those different roles, those later actions were often very similar.

The other policy changes that demonstrate convergence for the archived research data are with the terms of use policies for using the data themselves. For example, in addition to changing the possible access categories from five to two, DANS revised its usage licenses. For the Open Access data, which previously could only license data under the CC0 (public use waiver), there are now 22 possible open licenses (i.e., CC0, other Creative Commons licenses, and various open software licenses). For the data sets which are still categorized under the
legacy Open Access for registered users category as well as the Restricted Access category, the “DANS Licence” and any additional usage restrictions the data provider mandates apply. These were substantial changes to the user’s usage regulations, especially because the access categories and licenses applied to already archived data were and still are being changed to align with the new categories and licenses. GESIS also made substantial changes to the Data Archive “Usage Regulations” (no changes were made to the “SowiDataNet|datorium Terms of Use”). Not only do the new “Usage Regulations” include revised GDPR-compliant language, they also imposed new data storage limitations and stated that GESIS (the controller) would intervene in access category selection if GESIS felt that it was not appropriate for the data. The “Usage Regulations” also included new sections about the data user’s rights and obligations towards the data they use.

Policy changes for administrative and operational personal data also show significant convergence. Whether they instituted a new privacy policy (ČSDA/ SOÚ and GESIS) or revised an existing policy (FSD and DANS), the resulting GDPR-compliant policies are very similar. This is a direct reflection of all four institutions taking the controller role regarding personal data used for operational purposes. As controllers, they were required to identify their data protection officer, the purposes and legal bases for the data collected, and what rights data subjects have for each data collection purpose. For those institutions with parent organizations, the privacy policies also needed to meet parent organization requirements, and ČSDA/SOÚ, FSD, and DANS all drew heavily on the text of their parent organizations, with some minor tailoring to address institution-specific data types and purposes. The most significant difference in the new privacy policies is their implementation date: FSD, DANS, and GESIS all implemented their new policies in May and June 2018, very close to when the GDPR entered into force, while the
SOÚ’s policy applicable to ČSDA only came into effect in early 2022. This delay is due to debates at the parent organization levels about what aspects of Institute data protection are the responsibility of the individual institutes, and which are the Czech Academy of Sciences’. Therefore, the most significant factor here is role of the parent organization.

The changes to policies and processes were not the only outcome of the GDPR response efforts. Over the course of the transition period, all four data archives recognized a need for dedicated staff or external support to manage data protection across the organization. At FSD, this involved formally appointing a data protection officer (DPO); this role was filled by the staff member who had been informally fulfilling many of the duties of the DPO before the GDPR was adopted. FSD also temporarily brought in an additional legal advisor, who happened to have previously worked with the archived research data, to support the DPO, advise on planned changes, and draft and review policy and contract changes. Both the DPO and the legal advisor had experience with research data as well as the legal and data protection expectations, and were able to address concerns that arose during the transition.

GESIS and DANS both already had existing, external DPOs when the GDPR was adopted. GESIS has contracted with an external DPO firm since 2008, while DANS’s parent organization, the KNAW, has a DPO which fills this role for all twelve KNAW institutes and service providers. The former DPO, who had also been a Data Archive employee, is still a member of the DSS staff, and could help bridge the experience gap as related to the data protection considerations for the research data, but this is not a formal role. There is a legal advisor who specializes in the administrative and operational data. However, GESIS is a large organization, and across the other departments and services there was a noticeable lack of internal expertise in issues related to personal data protection for the research projects, and the
former DPO could not take on an additional role. Therefore, GESIS established an internal data protection working group comprising representatives of all departments, and a privacy coordinator position that would focus on the research data. At DANS, in addition to the KNAW DPO, there is also a legal advisor who is also a data processor for the EASY research data. However, the KNAW directed all its institutes to establish a privacy coordinator position that would focus on the administrative and operational data; the privacy coordinator and the legal advisor collaborate with each other and with the KNAW DPO and Chief Privacy Officer to coordinate data protection efforts across DANS.

For the duration of the GDPR transition, ČSDA had no formal DPO that it felt it could turn to, as the SOÚ did not have one until early 2022, and while the AV ČR appointed one in 2018, they did not work with any of the AV ČR institutes on their GDPR efforts. Instead, the ČSDA Expert, the ČSDA staff member who had the most experience addressing data protection and privacy issues for research data at the archive, became the expert advisor for the entire SOÚ, in addition to their normal work within the department. The uncertainty and disagreement about whose responsibility it was to provide a DPO, the AV ČR or the individual institutes, contributed to the almost four years late contracting of an external DPO firm to assist the entire SOÚ in managing data protection and GDPR compliance.

The “path” to GDPR compliance for the research data in particular at FSD and GESIS (with the exception of the 2019 hack) were deemed to be relatively straightforward, if extensive and burdensome, by the interviewees because of the long-term experience with a DPO and internal expertise with data protection for the data archives. Even at DANS, which still struggles to achieve compliance today, the internal expertise with data protection and research data supported planning future changes even as DANS waited on legislative changes and the
KNAW’s priorities, and the on-going collaboration between the DANS legal advisor and privacy coordinator has proven effective and efficient in addressing concerns that arise across the organization. Even at ČSDA, while the ČSDA Expert was being pulled in multiple directions to assist the wider SOÚ, they were still able to guide the archive itself in identifying priorities and understand what would be ČSDA’s responsibility to address, and what they needed to wait for the SOÚ to figure out. And, data protection efforts do not end simply because the organization becomes “compliant”; all four data archives continue to improve their systems and evaluate the effectiveness of their processes and policies. The presence of someone, whether a DPO or an advisor or a privacy coordinator, or a team of people, to coordinate and manage these efforts have greatly aided the data archives, and will continue to do so.

4.5.4 Challenges

While three of the data archives do consider themselves to be compliant, the process of becoming compliant with the GDPR is not “complete” at any of them, for a number of reasons. As ČSDA04 put it,

It is all the same in Europe and every member state, they need to need to follow some rules, they need to have safety measures, technical and organizational measures, they have to realize what data they collect, they process, for how long, they need to change it from now and then, they have to realize that the GDPR data protection is a process, not just a one-point-in-time thing.

There are changes still remaining to be made at each archive; most refer to finalizing or harmonizing policies and procedures across the organization. However, two archives, ČSDA and DANS, are currently transitioning their entire archiving system to Dataverse-based platforms, which will likely result in further data protection-related changes. DANS is also still considering whether to retain its role as “processor” for the archived data; a decision likely will not be made
until the new archiving system is in place, and until a decision is made to revise the Dutch GDPR implementation law. The challenges listed below in Table 4.6 mirror many of the challenges that Almeida Teixeira et al. (2019) identified in their discussion of critical success factors during the GDPR transition process; a lack of financial and human resources, increased workloads, and lack of privacy knowledge and expertise were highlighted as particularly troubling challenges.

Table 4.6 Challenges ČSDA, FSD, DANS, and GESIS faced while responding to the GDPR.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>ČSDA</th>
<th>FSD</th>
<th>DANS</th>
<th>GESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delays in GDPR implementation laws</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Conflicting data protection laws</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Concerns about archiving and research under implementation law</td>
<td>●</td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Lack of dedicated legal &amp; data expertise</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>What are “anonymized data”</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Wait for parent organization decisions</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Increased workload</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Lack of resources</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>External factors</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>How to approach older data</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How to approach “new” data</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Increased assistance to external researchers</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
</tbody>
</table>
The ČSDA, FSD, and GESIS interviewees otherwise consider their organizations to be “mostly” compliant. DANS, on the other hand, was not, and is still not. The reasons for this tie directly to the most influential challenge it faced during the GDPR transition period: the delays in the adoption and implementation of the national data protection laws. While this challenge was also faced by FSD and ČSDA, it was particularly problematic for DANS. While DANS began preparing for the GDPR soon after it was adopted, DANS had to wait for its parent organization, the KNAW, to determine what it thought the different institutes needed to do before it could truly begin making changes. The KNAW itself was waiting on the final version of the national data protection law (UAVG) in order to provide that guidance. Similar delays in the Czech Republic also affected how quickly ČSDA’s parent organizations, the SOÚ and the AV ČR, did the same. At both DANS and ČSDA, this means that changes related to the GDPR are still in progress in 2022, four years after the GDPR entered into force. Other delays and challenges came from concerns about whether data archiving and research would be affected by the national laws (FSD and DANS), and what “anonymization” really meant under the GDPR (GESIS). The timeliness (or lack thereof) of all archives’ ability to be considered “compliant” were severely hampered by the need to wait and see how the various legislation would be interpreted and enforced.

Of the four archives, only FSD did not report issues stemming from a lack of a dedicated advisor with both legal and data protection experience, as the person appointed to be DPO had this experience, as did its temporary legal advisor. The DPOs available for ČSDA, DANS, and GESIS before and/or immediately after the GDPR entered into force were all situated outside of the immediate organizations (either as the parent organization’s DPO (ČSDA, DANS) or as an
external firm (GESIS)), and the archives were not the only clients needing assistance and information.

A lack of resources and an increased workload were frequently reported together as complications during this period, as archive staff had to incorporate GDPR-related tasks into their existing workloads without any significant increase in either personnel or budgetary resources to assist. At FSD and GESIS, external factors—the merger of Tampere University and the 2019 account database hack, respectively—added to those workloads.

4.5.5 Outcomes and impact

ČSDA, FSD, DANS, and GESIS interviewees reported significant confusion and frustration while trying to adapt to the GDPR, particularly in terms of how the law(s) seemingly targeting “big data businesses” would affect and possibly restrict the archives’ abilities to continue functioning. In the end, however, none of the interviewees felt that the GDPR significantly affected data deposits or data reuse. While both ČSDA and GESIS experienced drops in number of annual data sets published between 2016 and 2018 (Fig. 4.3 and Fig. 4.24, respectively), and respondents both hypothesized that this might be due to researcher hesitation about the GDPR, neither archive followed up with their data providers about it. FSD’s fluctuations in this period (Fig. 4.6) as well was tied more to the re-prioritization of FSD’s workloads during the GDPR transition that slowed data processing times, and not to specific restrictions or concerns regarding the GDPR’s possible disruption of research.

GESIS and DANS both saw relatively small drops in data set downloads per year from 2015-2017, and FSD from 2018-2019. For all three archives, these were tied to specific data access categories (GESIS and FSD) or specific disciplines (DANS). None of the interviewees considered the GDPR to be a factor for the data use fluctuations. Since 2017, DANS and GESIS
have seen an increase in usage, and anecdotally, FSD, DANS, and GESIS interviewees reported an increase in usage after the COVID-19 pandemic began. ČSDA does not report its data usage, but ČSDA staff likewise did not think that the GDPR had any effect on data reuse. ČSDA has seen a steady increase in number of registered archive users since 2013, which is attributed to ČSDA branching out into the broader social science research community.

Instead of outcomes measurable by these operational metrics, the interviewees at all four archives reported a variety of outcomes and impacts of the GDPR. The biggest impacts, and the most frequently discussed, were primarily on the employee mindset. Interviewees reported an increased awareness of the importance of privacy and data protection and taking greater care in their work with data and with the information systems. They also reported being more cognizant of the need to protect and safeguard the data subjects’ rights. While this had never not been a responsibility, the discussions around the GDPR brought the importance of protecting these rights very much to the fore in all aspects of archive operations and the daily work.

4.5.6 Summary and key findings

The period of time from 2016-2020 was a busy period for all four data archives in this study. Despite their different places within their research communities, archive sizes, involvement in external and internal research projects, and services, ČSDA, FSD, DANS, and GESIS shared many experiences in terms of changes which were made and challenges that they faced. Revisions to legal agreements and terms of use were similar, bringing pre-GDPR text to reflect GDPR-specific language. The archives enhanced (formally and informally) their deposit review processes, prioritized transparency with data subjects (both research data subjects and archive users), and improved data protection training. The archives also faced similar challenges related to delays in finalized expectations about what would be required (both from delays in
legislation and from parent organizations), limited internal combined legal and data protection expertise, and an overall increase in workload.

There were differences though in how the four archives responded to the GDPR. Some changes were relatively minor, reflecting archive specific wording changes or orienting to the GDPR terminology but without actual policy or practice changes. Others, such as DANS’ change to the “Deposit Form” which can trigger the “Processing Agreement” requirement and an automatic Restricted Access categorization of the data, were more significant. Another significant difference is that, of the three archives that claimed a GDPR role for the research data (FSD, DANS, and GESIS), only one has completed any data protection impact assessments (DPIA). It is possible that, had GESIS completed a DPIA, the flaw in the user database system might have been identified before the system was hacked in 2019.

Thus, there are three significant findings of this dissertation. The first is that, despite national differences in the importance or relevance of privacy and effects of the GDPR on general public awareness of data protection and privacy, these differences did not significantly affect how ČSDA, FSD, DANS, and GESIS as archives addressed their obligations under the GDPR. Instead, the most influential factors were: what, if any, processor or controller role was adopted for which data (Section 4.5.2); national and sub-national legislation; and, a parent or supervisory organization (Section 4.5.1). These latter two factors are largely outside of the individual data archive’s control, yet both caused delays and confusion for the archives in their responses to the GDPR.

The second finding is that the changes that were made and the biggest impacts felt at the four data archives focused mainly on two of the three dimensions of the theoretical framework: regulating information flows and minimizing harms. More changes were made reflecting both of
those dimensions; and more entirely new policies and practices were implemented in both of those dimensions. The prioritization of regulating how data moved into and out of the archive and increasing transparency about data subjects’ rights was evident in how the interviewees from all four archives described the changes in these areas.

All four archives considered that the data protection and network security protocols already in place before 2016 were sufficient, and therefore, changes reflecting the technological risks to privacy which could be directly connected to the GDPR were limited to: improvements to data protection and GDPR-specific training for archive staff; enhanced rules for working with administrative and operational data; and, recommended lists of approved, Europe-produced personal data processing tools and applications. Significant changes to further secure personal data were made to GESIS and ČSDA’s archive account registration databases were made after the GDPR came into effect. However, both were a response to a specific external threat, and not to bring either database into compliance, as they were both considered compliant before December 2019. Thus, data archives should not take for granted that existing data protection measures are sufficient to protect privacy due to prior presumed compliance or a lack of previous data breaches. The other priorities for the archives reflected a recognized need to have dedicated staff or external advisors with legal expertise who were also familiar the needs of research data, and who could manage and advise data archive staff locally. The appointment of data protection officers, privacy coordinators, and the promotion of internal data protection and privacy expertise to help support the on-going work to manage and coordinate work in all three areas.

The final significant finding of this dissertation is that the GDPR did not measurably affect use of the archives as data archives. While there were fluctuations in annual data deposits and/or publications and data downloads from 2015-2019, these fluctuations were relatively
small. No archive considers with certainty that the GDPR played any role in those fluctuations, although interviewees from both GESIS and ČSDA hypothesized that the GDPR may have led to a period of researcher or data provider hesitance. Instead, the more significant impacts were mindset changes on the part of archive staff regarding data protection and data subject rights, and certain policy changes, like the processing agreements at DANS and FSD.
Chapter 5 Discussion and Conclusion

5.1 Introduction

In Chapter 4, I described the experiences of four European Union-based social science data archives—ČSDA, FSD, DANS, and GESIS—from 2016-2021 as they reacted and adapted to the General Data Protection Regulation. I presented each of the four case studies and compared their experiences to identify key considerations and challenges. I did so by evaluating their experiences through the theoretical model of informational privacy, in order to align what aspects of informational privacy the changes addressed.

This theoretical model was developed to address the insufficiency of Clarke’s (2016) definition of informational privacy: “the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves.” The challenge with this definition, which is one of the most commonly used definitions of informational privacy, is that, in the context of data archives working with personal data, such as social science data archives, that individual has become far removed from their personal data. First, their data were given to the research team, then to the data archive, then to the data user. Therefore, a construct of informational privacy focusing only on the individual’s immediate control of their own data is insufficient to consider the different considerations affecting how informational privacy is maintained in the context of research data archives.

Instead, the informational privacy framework used in this dissertation understands informational privacy as grounded in three different constructs of privacy: privacy from regulating information flows (Barocas & Nissenbaum, 2014; Nissenbaum, 2010), privacy as
minimizing harms to data subjects (Solove, 2006, 2008), and privacy in response to technological risks (Gavison, 1992; Warren & Brandeis, 1890). I incorporated these three constructs into my theoretical framework because they reflected three aspects of data archive operations that I considered most likely to be affected by new or revised data protection and privacy regulations: collecting, processing, and disseminating personal data; establishing access and usage policies; and, activities related to data protection. This dissertation allowed me to test the utility of this model in framing how informational privacy is conceived of and acted upon at these archives.

This comparative case study was motivated by two overarching research questions:

- How do the ways in which General Data Protection Regulation has been enacted increased our understanding about approaches to privacy in data archives?
- How has the full adoption of the General Data Protection Regulation affected data sharing and subsequent data reuse at four European Union data archives?

There were several subsidiary research questions intended to support answering these overarching questions:

1. What factors determined the four social science data archives’ responses to the General Data Protection Regulation?
2. What do the different archive implementations of the General Data Protection Regulation indicate about what aspects of privacy are prioritized?
3. What are the factors motivating the archives’ data protection policy and practice adaptations to the General Data Protection Regulation?
4. What are the mechanisms used at the data archives to address their stated goals of General Data Protection Regulation implementation?
5. In what ways have the different data archives’ policies and practices converged and diverged as a result of the General Data Protection Regulation?
6. How have the data archives conceptualized “contextually appropriate” information flows in response to the General Data Protection Regulation requirements?
I found that, while there were a lot of broad similarities between the four data archives—new or revised policies for similar operations, common internal and external challenges—there were three main factors that most influenced how each archive responded to the GDPR: which processor or controller role, if any, the archive adopted regarding their personal data; national or sub-national legislative considerations; and, the role of a parent or supervisory organization.

While there were a number of shared adaptations at the four data archives, implying a level of standardization of practices and policies as intended by the GDPR, there were also several areas where the data archives took different routes. These areas, including the processor or controller decision, application of certain agreements, requirements for data deposit documentation, and interpretations of data disclosure risk, etc., however, do not reflect differing concerns over the importance of personal data protection and privacy at the different archives, but instead reflect the level of responsibility that each archive is willing take on when it comes to enforcing and complying with the data protection laws.

When I evaluated and compared the experiences at ČSDA, FSD, DANS, and GESIS, I found that the changes made at each archive, whether formal, documented changes or more abstract changes in mindset, closely aligned with two of the three dimensions of my revised model of informational privacy: regulating the flow of information into and out of the data archive (Barocas & Nissenbaum, 2014; Nissenbaum, 2010) and minimizing harm to data subjects (Solove, 2006, 2008). The other dimension, changes aimed at responding to technological risks to privacy, was present, though not to the same degree as the other two dimensions, in part because all archives considered their data security measures before the GDPR was adopted to mostly be sufficient, so minimal changes were deemed necessary. The remaining changes at the archives were aimed at managing the archives’ compliance in this area.
Finally, while the GDPR resulted in procedural and mindset changes at all four archives, the GDPR itself could not be directly correlated with measurable operational outcomes, nor did any interviewee feel that the GDPR had any effect. Both GESIS and ČSDA staff hypothesized that data providers might have been more hesitant to deposit their data with an archive immediately after the GDPR was adopted in 2016, neither archive investigated this possibility, and data deposits did begin to increase after 2018 at both archives. Similarly, while data set download counts fluctuated from 2015 to 2019, again, these fluctuations could not be conclusively connected to the GDPR. Though it began in 2020, at the end of the period I intended to focus on in this dissertation, the COVID-19 pandemic was seen as having a greater impact on data reuse than the GDPR or any of the changes the data archives made in response to it. Instead, the most significant outcomes of the GDPR for all archives lay in the mindset changes about the importance of data privacy and data subject rights and contract changes related to data processing by the archive.

5.2 Discussion of major findings

In this section, I will examine the findings discussed above.

5.2.1 Three factors played the biggest roles in the GDPR response

Discussions of privacy and data protection, especially comparative studies, highlight national and cultural differences in how “privacy” is conceived. Whitman (2004) characterized the “European” approach to privacy as one focused on respect and personal dignity (further discussion of this point in Germany, for example, can be found in Borneman, 1997; Espindola, 2011), in contrast to the United States approach focusing instead on liberty and freedom from intrusion. The boundaries of public and private, whether in the physical sphere (e.g., Garvey,
2005; Mols & Janssen, 2017; Van Der Horst & Messing, 2006; Vera, 1989) or the digital (e.g., Gawrecká, 2013; Kozyreva et al., 2021; Krasnova et al., 2012; Krasnova & Veltri, 2011; Vasalou et al., 2010), are not universal and frequently differ even between neighboring countries. Additionally, 20th century personal data abuses, including events that directly affected the countries my field sites are based in (specifically, the Holocaust and post-World War II communist rule) were motivators for the adoption of data protection legislation beginning in the 1970s and the classification of specific data points as especially at-risk for abuse. Given these experiences, and building on research into privacy motivations and perceptions in the Czech Republic, Finland, the Netherlands, and Germany, I considered that these cultural differences in interpretation and prioritization of privacy and the historical experiences would at least in part answer the first of my subsidiary research questions: What factors determined the four social science data archives’ responses to the General Data Protection Regulation?

However, the findings of my dissertation do not support this. These cultural and historical considerations did influence national data protection efforts, even before the Data Protective Directive implementations. For example, as discussed in Section 4.1, the Czech Act on Personal Data Protection (2002) did not allow for personal data reuse for scientific purposes, a restriction based on members of parliament experiences under the communist regime (ČSDA02). In Finland, the emphasis on employee personal data privacy has its foundations in the early 20th Century “black lists of trade unionists and trouble-makers” which were used both to refuse employment and to fire union members already employed (Suviränta, 1995, p. 46). In the Netherlands, after World War II, religious identifiers were removed from official registers and decennial census questionnaires, as the identifiers in the registers supported Nazi efforts to identify and transport Dutch Jews to concentration camps (Ketelaar, 2020; Seltzer & Anderson,
2001). The Nazi atrocities and the East German experiences with the Stasi have been highlighted as the main motivators for the German focus on both personal data protection and privacy, especially in terms of protection from government surveillance and data abuses (DeSimone, 2010; GESIS04). It is unsurprising that the first data protection law in the world was a West German state law targeting the state government’s creation of a centralized database of residents’ personal data, which then-Hessian Minister-President Albert Osswald said would prevent the establishment of an Orwellian surveillance state in Hesse (“EDV im Odenwald,” 1971).

As influential as these experiences were in establishing data protection and privacy legislation, by the time the GDPR was adopted, the precedents and obligations of data archives (and other institutions working with personal data) derived from several decades of data protection laws were already well established. The GDPR itself, building on the historical legacies that motivated the 46 years of preceding data protection efforts across Europe, is the EU’s attempt to codify a standardized norm of data protection of privacy in line with Manners’ (2002) concept of “normative power” and what Coman (2018) calls a “community of values” (Section 2.3.2). Thus, these experiences did not factor into the decisions made at each archive while responding to the GDPR, though they are reflected in the constraints on and specific protections for types of personal data which have historically been abused. Instead, the factors most influential in how ČSDA, FSD, DANS, and GESIS responded to the GDPR were instead: what, if any, processor or controller role was adopted for which data (Sections 4.5.2 and 4.5.3.5); national and sub-national legislation (Section 4.5.1); and, a parent or supervisory organization.

---

197 The full quote reported in Der Spiegel, May 9, 1971: “Die Orwellsche Vision des allwissenden Staates, der die intimsten Winkel menschlicher Lebenssphäre aus forscht, wird in unserem Land nicht Wirklichkeit werden” (The Orwellian vision of the omniscient nation probing the most intimate corners of human sphere of life will not become reality in our state). Translation by the author.
(Section 4.5.1). These are important findings because of the lessons that data archives just beginning their own responses to new data protection laws can learn.

5.2.1.1 To be processor or controller?

The first decision that each archive made was what responsibility it would take on regarding any personal data it works with, whether these data are research data from on-going, archive-involved projects, research data in the data archive, and any administrative and operational data. Research data archives which archive and disseminate personal data from external data providers have to decide which role to adopt for these data; organizations may have more clearly defined roles for other data, such as “controller” for administrative and operational data necessary for operations. There is no legal guidance or requirement either that research data archives choose a specific role for these data. If the archive takes on the role of a controller in line with the GDPR’s definition, then they have a certain level of prescribed responsibility and liability for the data, including demonstrating compliance. The controller also decides what processing will be conducted on the personal data. If the archive instead takes on the role of a processor in line with the GDPR’s definition, then, while the processor must still comply with the GDPR, it only processes data in accordance with what the controller determines, and it is not required to demonstrate their compliance under the Accountability Principle (Art. 5(2) GDPR) other than to the controller.\(^\text{198}\) The choice between processor and controller will dictate many of the following changes, as seen in DANS and FSD’s “Processing Agreements,” or the content of

\(^{198}\text{According to Article 28(3)(h), the processor must “[make] available to the controller all information necessary to demonstrate compliance with the obligations laid down in [Article 28].” Therefore, the processor does need to demonstrate their GDPR compliance to the controller they have a processing contract with, but the controller demonstrates their compliance to the wider community, including by not only implementing data protection measures but also demonstrating their effectiveness (Recital 74 GDPR).}
all four archives’ privacy policies, which conform to the obligations of the controllers, and is thus the first major factor in the GDPR responses at each archive.

While the decisions made here about the research data differed across the four data archives, and role specific changes were made such as the implementation of a processing agreement, in other ways, the choice of processor or controller did not cause divergence in the archive policies and practices, such as the components of the deposit review and who would be responsible for anonymization. There are direct liability and responsibility differences based on role, certainly, but for the research data in a data archive, especially in light of the on-going discussions at DANS about changing their role, the question remains whether the role truly matters? Does it matter that a processor is not beholden to the Accountability Principle (Art. 5(1) GDPR), in that it does not need to demonstrate its compliance with the GDPR principles to anyone other than the controller (who is responsible for demonstrating that the processor is complying), when Article 28 GDPR reiterates that the processor is still obligated to protect personal data according to Article 32 GDPR, and to comply with the GDPR and national laws themselves?

In Chapter 4, I described the ways that the choice of specific roles, particularly with the archived research data, dictated following decisions and changes to archive policies and processes. However, I have also made clear that even when specific roles were adopted, the implementations of those responsibilities were not consistent. DANS and FSD both implemented “Processing Agreements” as required by Article 28(3) GDPR; however, FSD requires them with every deposit and DANS only when the data set contains personal data. FSD and GESIS adopted different roles, and ČSDA no role, and yet all through have implemented a new requirement that data providers include proof of consent for data archiving with data
deposits (though again, as with the processing agreements and DANS, this is not a universal requirement applied to all data deposits at ČSDA). All four data archives require that any necessary data anonymization be completed by the data provider before the data are deposited. However, the reasonings given for this decision, despite the differing roles, were essentially the same: the data archives did not want to be responsible for the results of that particular type of processing. Yet at the same time, all four data archives indicated that they would do anonymization work that amounted to small, quick changes or corrections (or in the case of ČSDA’s review of potentially problematic data sets, in-house attempts to further anonymize the data set) to data sets. More extensive anonymization requirements would see the data set returned to the data provider, but allowances could be made for small things. For example,

… what we do is that with the ingest check, if we come across things that need refinements, whatever, if it's minor things, we will do it. … So, if we see that it’s major changes that have to be made, we pass the data back to the data providers (GESIS04).

FSD’s “Processing Agreement” could also include a “anonymization” processing task, but even then, FSD’s preference and expectation is that the data provider has done this fully before the data ever reach the data archive.

The practical implementations of these roles therefore indicate a fluidity to the intended responsibilities and expectations associated with the roles when it comes to archived or archivable research data that goes beyond a consideration of the potential legal and ethical liability for personal data. Perhaps the question is not whether the choice matters for a research data archive, but whether the processor and controller roles as currently conceived of in the GDPR can, or should, even be applicable to the archived research data aspect of data processing work. If so, further guidance about the expectations and obligations for each role related to the research data would certainly be helpful. The changes and approaches to the other personal data
these archives worked with—the data resulting from archive-involved research and the administrative and operational data—were relatively straightforward. The archived research data, however, are in a unique context, where the personal data are processed and shared with the expectation of further use by yet other data users. This is not a context that is clearly defined in the GDPR or in the national legislation, and is an area of significant confusion as there is no guidance for data archives on how to determine which approach is best for the archive, or what the best practices would be for the archives once they adopt a specific role. I do not have an answer to this question at this time, though as I discuss in Section 5.5, this is a valuable path for future research.

5.2.1.2 National and sub-national legislation

The data protection legislation itself is also a significant factor in how data archives approach any operational changes. If, as with the GDPR and European Union-based data archives, there are multiple levels of applicable legislation, delays in national or sub-national legislation implementation may only allow organizations to adapt their operations only up to a point. The GDPR, for example, allows for national legislation to derogate from the GDPR on certain points: Article 89(2) GDPR allows for national derogations from the data subjects rights (Art. 15, 16, 18, and 21 GDPR) to their personal data when those data are used for research or statistical purposes, derogations which all four Member States in this research (Section 2.3.5.3) utilized (the Dutch UAVG only limited the Article 15, 16, and 18 GDPR rights). These derogations are important for the initial research data collection as well as personal data reuse from the data archives.

Confusion about just how the GDPR and its implementations would actually affect the research communities were highlighted in the interviews as the main source of challenges during
the GDPR transition period. The confusion about the text was hampered by the delays in adopting and implementing the laws; in no Member State was this a fast or efficient process. The delays in finalizing and adopting the new data protection legislation, and in clarifying the points of confusion discussed above, meant delays in identifying further necessary measures and finalizing planned changes to policies and legal contracts. What changes were needed, and how quickly they could be made, relied on timely adoption and interpretation of national legislation, not just an understanding of what the GDPR itself said. Therefore, the national and sub-national legislation are another significant factor influencing the data archives’ responses to the GDPR.

5.2.1.3 Parent or supervisory organizations

The last significant factor in how the data archives responded to the GDPR is the presence of a parent or supervisory organization. Whether a data archive has a parent organization (either as a dependent component of that parent organization (ČSDA, FSD) or as a separate entity within a larger hierarchy (ČSDA/SOÚ, DANS)) and how much influence and oversight that parent organization has over long-term strategic planning can support an archive’s efforts towards compliance or complicate those efforts. A parent organization’s recognition of the data archive’s expertise about their own needs (as occurred with the University of Tampere and FSD) may lead to minimal interference and external confusion about what the data archive actually needs. Depending on the level of dependence between the data archive and its parent organization, such as ČSDA’s status as a department within the SOÚ, the data archive may also rely the parent organization to adopt or adapt policies that bring the data archive closer to GDPR compliance. The legislative delays and confusion discussed above can further complicate matters at data archives which need to wait for their parent organizations to first interpret that
That these three factors—the choice of a processor or controller role, the national and sub-national legislation, and the parent organization—were significant to the GDPR transition process, and the cultural and historical perspectives on privacy and data protection were not, was a surprise. I began this dissertation wondering whether, and in what way, those cultural and historical perspectives would affect the archives’ approaches in some way, anticipating that they would. My findings indicate instead that they did have an impact on the early (pre-Directive or pre-GDPR) data protection legislation (Aldhouse, 2018; C. J. Bennett, 1992; Hondius, 1980; Kirby, 1980; Riccardi, 1983); this has been discussed in the literature on data protection efforts in Europe and was also referenced by my interviewees. These perspectives also influenced how my interviewees thought about privacy in their personal lives; however, in terms of their work with the data archives and responding to the GDPR, the impact was already in the past, setting up the pre-GDPR national approaches to data protection and privacy that were not significantly altered by the GDPR. Instead, my findings indicate that, for archives which had pre-existing data protection or privacy legislation that they needed to abide by, these three factors were most important, because they directly affect what the adaptation process looked like, and how quickly the archives could become compliant.

5.2.2 Changes focused on data flows and harm minimization

While, as discussed elsewhere, the GDPR is focused on personal data protection, and privacy is a hopeful outcome of data protection efforts, the majority of my research participants framed their archive’s responses to the GDPR in terms of protecting privacy through various means, as evidenced throughout Chapter 4. My theoretical framework was designed around legislation and answer those questions before developing priorities for their subordinate institutes.
three dimensions of privacy that support informational privacy—privacy by regulating information flow, privacy from minimizing harms, and privacy as a response to technology—which I conceived of as reflecting the processes and procedures at the archives aimed at (re)enforcing data protection and privacy. Which, if any, of the framework dimensions the data archives’ decisions and changes reflected was an important question I wanted to address through my interviews and analysis of the policy documentation. As discussed in Section 4.5.6, the archives’ responses (both changes to existing policies and procedures and entirely new ones) focused on two of the dimensions: regulating the flow of data through the archive and minimizing harm to the data subjects. More changes were deemed necessary for archive operations related to these two dimensions (e.g., agreements for data deposit or usage, privacy policies, deposit review, data access categories, etc.), and those changes were considered significantly different from the pre-GDPR operations.

Does this answer the second of my subsidiary research questions: What do the different archive implementations of the General Data Protection Regulation indicate about what aspects of privacy are prioritized? Despite the bulk of changes at the data archives focusing on the two dimensions—regulating the flow of information and minimizing harm—my findings indicate that this does not mean that the technical aspects of data protection (specific data protection measures, restrictions on use of certain tools and applications, and employee training) were not considered important, just that numerically fewer changes were made. Technical and awareness-raising privacy protection measures were and are on-going priorities at the data archives, whether considering the research data (both in the archive(s) or collected in archive-involved projects) or the administrative and operational data used by the archives. All four data archives began their GDPR responses by evaluating various aspects of their operations to determine where changes
might be needed within their own contexts. ČSDA and FSD established their own priorities without input from external parties, while GESIS’s external data protection officer also developed a checklist to guide GESIS through the necessary changes. DANS also did their own review, but the KNAW data protection officer also conducted GDPR’s data protection impact assessments on the three DANS services (EASY, NARCIS, and DataverseNL). The archives’ priorities established that changes were needed in specific areas aligning with all three dimensions, and that the majority focused on the data flows and minimizing harms.

But this does not mean that the technological considerations for privacy were not a priority for the data archives. Prior to the 2019 hack of GESIS’ Data Archive and SowiDataNet|datorium account registration database (where 75,000 users’ personal data were stolen), the IT infrastructure supporting data protection at all four data archives was deemed sufficient and compliant with the GDPR requirements for “technical and organizational measures to ensure a level of security appropriate to the risk” to data subjects (Art. 32(1)). Therefore, the priorities for this dimension instead targeted non-European Union-based third-party applications and tools for the administrative and operational data, and data protection and data management training for archive staff as well as external parties. This is in line with how the other three organizations approached this dimension. And when the security flaw was identified, GESIS addressed it and moved forward with planned changes to the account registration databases for the various GESIS services (Sections 4.4.5.7 and 4.4.7). ČSDA observed what happened at GESIS, and while it had also determined that its Nesstar account database was secure and GDPR compliant, ČSDA staff decided to make similar changes to make the system even more secure.

The training component highlighted in Section 4.5.3.3 in particular approached the potential technological risks to privacy from an awareness and identification perspective, and not
a reactive perspective. As many interviewees described, by having privacy and data protection at the forefront of their work, they feel more capable of identifying aspects of their work that may be risky; this is especially important for those staff members whose work is less directly related to personal data (i.e., those who do not work directly with the archived research data, or the user or personnel data). Thus, protecting data subject privacy in the face of technological risks remains a priority, even if the focus of the archives’ changes were not on the technical measures themselves, just as do those controlling whether and how data come into or leave the archive or implementing changes that reduce the harms to the data subjects. All three aspects of informational privacy are prioritized at the data archives, and where they focused attention while responding to the GDPR do not imply that certain dimensions are prioritized differently.

The goal of all four data archives was to both be compliant with the GDPR and be able to demonstrate their compliance with the GDPR, even for the systems for which they were not the controller. Even at DANS, which for example no longer requires proof that research subjects consented to data archiving because of their role as processor, if questions arise about whether the EASY data are GDPR compliant, DANS11 still emphasized that DANS will address those concerns with data providers or data users. GDPR compliance is not just a singular instance, but an on-going process; the data archives cannot just say that they are GDPR compliant as of a certain date and not consider it again. In their review of GDPR implementations at Portuguese health clinics, Lopes et al. reiterated that “data protection is not a destination, it’s a journey” (Lopes et al., 2020, p. 504), a perspective shared by my interviewees. Being, demonstrating, and maintaining compliance with the GDPR and other data protection legislation thus answers the third of my subsidiary research questions: What are the factors motivating the archives’ data protection policy and practice adaptations to the General Data Protection Regulation?
The changes implemented at ČSDA, FSD, DANS, and GESIS closely followed one of the mechanisms of organizational change described by van de Ven and Poole (1995; Poole & van de Ven, 2004): teleology. Teleological theories of organizational change view organizational change as a “repetitive sequence of goal formulation, implementation, evaluation, and modification” (van de Ven & Poole, 1995, p. 516) which is purposeful and requiring consensus among organization members. Unlike other organizational change theories, the teleological approach is not prescriptive in how change is implemented, instead focusing on goal attainment (Sune & Gibb, 2015). For the data archives in this study, this goal was GDPR compliance, as previously discussed.

The first thing the data archives did once that goal was set was to review their operations and establish initial priorities to meet the goal, either alone or with external support. Then, over a period of time, changes to policies, processes, and documentation were made to meet those priorities. As evidenced by on-going revisions and re-evaluations of those changes (e.g., updates and revisions to research consent forms (FSD), privacy policies (DANS, GESIS), processor versus controller roles for certain data (DANS)), none of the archives, even those which consider themselves to be GDPR compliant, view their work to be complete. When there are perceived or actual deviations between the intended goals and the current outcome, such as after the 2019 hack at GESIS, decisions and changes are re-evaluated, and modifications made. The consensus necessary to support these efforts was and is continuously developed through organizational training and oversight of the data protection work by data protection officers, legal advisors, privacy coordinators, and other internal and external advisors. Therefore, the mechanism driving the data archive changes in response to changes in legal requirements is teleological: purposeful changes to aspects of data archive and personal data management operations to meet a specific
goal, with consensus across the organization, and evaluation and revision when and where those changes fall short. This addresses the fourth of my subsidiary research questions: What are the mechanisms used at the data archives to address their stated goals of General Data Protection Regulation implementation?

As discussed in Section 2.3.5, the goal of the GDPR was and is to harmonize the data protection regime across the EU, and as discussed later in Section 2.3.5.3, at the level of the data protection laws, this has occurred. There are still important differences in how certain components of the GDPR have been applied; the German utilization of “factual anonymization” is one; Finland’s “freedom of expression” exemption to the requirement for informed consent when archiving identifiable personal data about “publicly important” (FSD04) people is another. While the timeframe of changes (Section 4.5.1) and what specific changes were made (Section 4.5.3) across the four data archives did differ, for the reasons discussed in Section 5.2.1, the four data archives’ changes do indicate that that goal of harmonization is not impossible.

There were relatively few changes that all four data archives made in similar fashion (these few are the increased strictness in the data deposit review process, the privacy policies, responsibility for anonymization, data protection training, and data protection management). However, changes that were made rarely occurred in isolation. For example, DANS’ decision to make all data sets with personal data “Restricted Access” no matter the legal basis for processing or willingness of the data subjects (in contrast to FSD’s “public interest” basis) is tied to DANS’ interpretation of its role as “processor” and is a direct result of other changes made because of that role.

If harmonization and convergence is defined as all organizations making the exact same changes (role dependent) in response to the same legal interpretations, then the answer to my
fifth subsidiary research question—In what ways have the different data archives’ policies and practices converged and diverged as a result of the General Data Protection Regulation?—would have to be that the four archives diverged in almost every aspect. The national and sub-national implementations of the GDPR are still different, though with less variation than before 2018. The changes even by those data archives which chose the same role for a given type of personal data did not result in identical policies and practices, reinforcing the question raised previously of whether the processor or controller choice matters.

However, if convergence instead is considered a process which takes into account that the GDPR itself allows for national differences in interpretation and does not demand a definitive end state of perfect unity, then the four data archives are converging in their operations and processes. This approach to convergence aligns with Kerr’s definition of convergence as “the tendency of societies to grow more alike, to develop similarities in structures, processes and performances” (Kerr, 1983, p. 3). Dunlop et al (1975) further highlighted that convergence is a process towards change, not an end state in itself. The GDPR seeks convergence but not perfect uniformity, and therefore, taking into account the national legal and the archives’ operational contexts, the four data archives’ approaches to data protection through revising policies and processes are converging, and are more similar than they were before. Except for ČSDA’s reaction to the 2019 GESIS hack, none of the archives in this study indicated that they formally consulted or mimicked the decisions made at other EU-based social science data archives (which, given the various delays and complications at each archive, may not have been possible before 2018, at the earliest), even though they are all members of CESSDA). However, the converging that we do see aligns with institutional theory, where this convergence comes from a shared institutional logic about how policies and practices addressing specific goals or
expectations are best implemented, based on common norms and rules (DiMaggio & Powell, 1991; W. R. Scott, 2005, 2008).

Where this convergence is most apparent is in how the data archives regulate the flow of personal data into and out of the organization. Nissenbaum (2010) framed privacy not in terms of individual control over personal data, but as a contextually appropriate flow of information. Even before the GDPR, the data archives implemented procedures that established these appropriate flows through access restrictions (e.g., requirements for account registration to download data, data provider permission) and the deposit review process (e.g., verification of sufficient anonymization, affirmation that the data provider can legally transfer the data). These procedures were revised during the GDPR transition period as were the data they applied to (e.g., DANS realigning already-archived data with the new data access categories and restrictions, ČSDA changed the account requirements to access data through the Nesstar catalogue, FSD’s new incorporation of pseudonymized data, privacy policies at all four data archives identifying the appropriate legal bases for data processing). However, these revisions did not significantly alter those appropriate contexts from the pre-GDPR flows; instead, they simply formalized and clarified the why of this data processing that had already been on-going.

ČSDA, FSD, and GESIS however did implement a policy change that did establish an additional context under which data archives interact with personal data in their research data: establishing a requirement that data subjects have not only consented to take part in a research study, but also to the resulting data being archived. This is in accordance with the GDPR’s Article 6(1)(a) legal basis for processing: that “the data subject has given consent to the processing of his or her personal data for one or more specific purposes.” It is important to note that there are other legal bases for processing (including archiving) that also apply at the data
archives, including those not provided for in the GPDR (for example, the Finnish “freedom of expression” basis); however, the vast majority of personal data archived at the four data archives fall under the “consent” basis. As Custers et al. state, consent “is the only [legal basis] that is not based on necessity,” and thus is the only basis that allows for the data subject’s ability to freely choose to have their data be processed (Custers et al., 2022, p. 463), an ethical and moral priority reflecting the historical motivations for data protection and privacy rights in Europe.

The German BDSG explicitly requires that consent must be given “in a particular case” (§ 46(17))—i.e., for each type of processing—as does the Czech ZZOÚ, though the Finnish Data Protection Act does not explicitly require it. All three data archives, however, have started requiring proof that the data subjects are aware of and consent to their data being archived. While ČSDA, as discussed in Section 4.1, does not as a rule archive personal data outside of the SOÚ-provided data not made available outside of the institute, it does still require proof from the research team that those research subjects consented to archiving (when informed consent was the basis for collection and archiving). FSD applies this new requirement to all data sets archived after the GDPR came into effect, with the additional requirement that the “privacy notice” the data subject agree to must identify FSD specifically. GESIS also requires this for all data deposited into the Data Archive (but not the self-archiving SowiDataNet|datorium) under the new 2021 archiving service package contracts.

Data providers comply with this requirement by including a copy of the consent form or privacy notice (unsigned) given to all research participants with the other data set documentation. Failure to provide this proof, or if the terms of the form are considered insufficient, the data

---

199 § 43(4), ZZOÚ: “Prior to granting his or her consent, the data subject shall be informed of the purpose of processing and for what personal data is the consent being granted, to which controller and for what period of time. The controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data throughout the entire term of the processing.” This can be interpreted as requiring the research team or data collector to not only require consent for the initial data collection, but also to future data archiving or dissemination.
archive may reject the data deposit, or only allow an anonymized form to be deposited instead. The interviewees described this new requirement both as a means to regulate whether and how research data can be archived and as a means of demonstrating and prioritizing the data subjects’ rights to their data. As discussed in Section 4.3, DANS no longer requires that the data providers include this proof, based on its interpretation of its responsibilities as a processor. However, as controller, the data provider still must confirm that they collected the data in a legal manner. Because DANS does not collect the research consent forms, it cannot verify if research subjects did or did not consent to data archiving; if there are concerns about this, or data subjects wish to withdraw their consent, they must address this with the data provider, which then works with DANS to further address and rectify any concerns. Therefore, consent for personal data processing, to include data archiving, has been established as the primary appropriate context for data archiving. When other legal bases possible under Article 6(1) GDPR apply to personal data archived at or used in the course of the data archive’s operations, they are clearly identified through the GDPR-compliant privacy policies or other documentation. The Article 6(1) GDPR bases, and other legal bases identified in the national data protection laws, thus provide the contexts that address my sixth research question: How have the data archives conceptualized “contextually appropriate” information flows in response to the General Data Protection Regulation requirements?

5.2.3 Biggest impacts are immeasurable

A motivating question for this dissertation was whether, and how, the GDPR and the revised national data protection laws would affect European Union-based data archives working with personal data. In particular, I was interested in whether there were any measurable changes to data deposit, data archiving, and data use, and hypothesized that there would be changes
which could be connected to the GDPR after the 2016 adoption. For this reason, I examined usage statistics from all four data archives (Section 3.3.2.2). However, as discussed in Section 4.5.5, even though there were fluctuations in both data deposit (or data publication, depending on how the data archive reported their statistics) and data usage from 2015-2019, these changes cannot be conclusively connected to the changed legal requirements. Both ČSDA and GESIS did hypothesize that the decrease in their data set publication numbers during that period could be related to data provider or researcher hesitation in the immediate aftermath of the GDPR’s adoption. However, as with this dissertation, neither data archive thus far has investigated this with their data providers. At FSD, the rate of data set publications also dropped from 2017-2019; while the interviewees did not specifically identify the GDPR as a possible cause, the Annual Reports (Finnish Social Science Data Archive, 2018c, 2019b, 2020b) did report fluctuations in the percentage of data sets processed in nine months or less which corresponds to the period that the interviewees discussed a re-prioritization of normal operations and duties.

Overall, even with these fluctuations, since the GDPR entered into force, usage of the data archives as archives (i.e., as locations in which to deposit research data, store them, and make them available for reuse) has increased. Since 2020, as well, increased use of the archived data has been anecdotally connected to shifting work and research practices during COVID-19-related restrictions at FSD, DANS, and GESIS.

The biggest impacts, instead, were not the measurable ones. As discussed in Section 2.3.5, at a high level, the most important contributions of the GDPR are the standardization it has required for personal data protection across 27 different legal regimes (Albrecht, 2016; Aldhouse, 2018; Mondschein & Monda, 2019) and the increased transparency about what will happen to an individual’s personal data and why (Junge, 2018). Beyond this, however,
according to my interviewees, the biggest impact that the GDPR had at the data archives was the changes in the data archive employee mindsets regarding data protection and privacy. Despite, and perhaps because of, the tradition of data protection efforts at all four of the data archives, the data archive staff all pointed to a growing awareness of their importance. These were not new concepts by any means; the changes instead were in how they moved to the forefront of the work. GDPR-related considerations are a part of all planning processes, and all data processing work is reinforced through data protection training and workshops. All four archives also have actively encouraged this increased awareness outside of the archives; they also provide training, outreach, and user support to researchers in their national research communities. In their collaborations with CESSDA as well, they are contributing to the reinforcement of data privacy as a professional value and norm among the broader European community of data archivists and researchers (Pyrko et al., 2017).

This is an important finding because of the pre-2018 fears and confusion over what the GDPR would mean for research and data archiving across the European Union, much less other personal data processing. As reported in the case reports in the previous chapter, panic was one of the main reactions to the GDPR, often compounded by delays in adopting the national data protection legislation and legal speculation about what limitations would be imposed. The interviewees reported the same panic and confusion when it came to the data archive operations. However, at least for these four social science data archives, while their actual operations changed, as discussed in depth above, the feared break down of data sharing and data use did not occur. This is important for other research data archives, and may allay fears that new data protection legislation will negatively affect data sharing practices. While there may be short-term fluctuations in data deposits and data reuse, my findings indicate that they will not be long-
term decreases, and will restart growth as data providers and data users become more familiar with any new expectations or requirements. Data archives may contribute to this recovery, as ČSDA, FSD, DANS, and GESIS have, through their user support, data protection training, and outreach efforts. These efforts not only develop and reinforce data protection and privacy in the minds of the archive users, but also in the data archive staff. CESSDA also seeks to be a sustainable social science research infrastructure; through its tools, research resources, training, collaboration on EU-wide projects, and best practices for researchers and data archives, it is likewise fostering this mindset on a broader scale. ČSDA, FSD, DANS, and GESIS’ involvement as CESSDA nodes contributes to these norms and standardization across the Consortium and to the convergence discussed in the previous section.

5.3 What does the GDPR mean for informational privacy?

The theoretical framework underpinning the analysis of this dissertation was a reconceptualization of “informational privacy.” I determined that this new framework was necessary at the beginning of my research because other interpretations of informational privacy centered the data subject in ways that did not seem sufficient in the context of a data archive. Clarke (2016) frames informational privacy as an individual’s immediate control over their information. Floridi (2013) similarly defined information privacy as an individual’s restricting external knowledge about their personal information. Floridi’s privacy violation is thus a violation of “the informational self” (Floridi, 2013, p. 260). Studies of privacy and personal data sharing often focus on the data subject’s perspective, and the findings highlight concerns about their lack of control (e.g., Damschroder et al., 2007; Haug, 2017; Hondius, 1980).

However, these studies about and definitions of informational privacy focus on the person, the individual data subject. Informational privacy in data archives, where the amount of
control that the data subject will have is significantly distanced and therefore reduced, thus should consider this context. My theoretical framework for this dissertation thus reflected the aspects of data archive operations which intersect with data protection and privacy. These are operations related to: the flow of data into and out of the organization (“privacy from regulating information flows”), minimizing privacy risks to data subjects (“privacy by minimizing harm”), and data protection measures (“privacy as a response to technology”). Once I developed this framework, the next task during the data analysis was to understand whether and how the data archives’ GDPR response actions aligned with it.

As discussed in Section 5.2.2 based on the pre-GDPR data protection and operational measures which were changed in order to become GDPR compliant, the data archives’ priorities focused on regulating the data flows and minimizing privacy risks and harms. This does not mean that the technological risks were not considered; all four data archives did review their data protection systems and procedures as part of their GDPR responses. However, because the data protection measures for the IT infrastructure already had to comply with previous data protection legislation and other best practices, the changes in this regard did not focus on the technical infrastructure at the archives until after GESIS was hacked in late 2019. Instead, the majority of substantial changes can be found addressing the flows of data into and out of the organizations and intending to minimize harms to the data subjects.

A consistent challenge all four data archives faced was confusion about whether, and then how, the GDPR and national legislation would affect the organization. It is not enough, nor is it easy, to simply read a complex legal text and interpret its requirements for a common understanding. Those requirements then had to be converted into actual, actionable practice and policy changes; doing so required effective implementation strategies that fit the requirements to
the needs, goals, contexts, and resources of the specific data archive (O’Toole Jr. & Meier, 2015). This was further complicated where there was greater reliance on the supervisory organizations or where there was limited internal or external expertise in the legal and data management aspects of data protection. Where they could, the archives did turn to knowledgeable colleagues for guidance—this is most noticeable at FSD, ČSDA, and GESIS—though with varying levels of formality. For ČSDA and GESIS, unlike FSD, this resulted in adding to the workloads of the staff as they took on the additional tasks of being (or becoming, in the case of GESIS’ internal data protection working group) the main point of contact for organizational questions and concerns over the GDPR (Kellogg, 2014).

While all four archives made staffing changes, both permanent and temporary, targeting management of the immediate changes and the longer-term compliance maintenance and coordination by using “experts” in either the data or the data protection (though ideally both), these occurred based on the specific contexts of the data archives (Kanter, 1983; O’Toole Jr. & Meier, 2015). For example, both FSD and ČSDA had temporary GDPR advisory support from already existing staff with experience in data protection for research data, though for different reasons and at different levels of willingness. These two archives are also relatively small, and data protection efforts primarily focus on the administrative and operational data and the archived research data, unlike DANS and GESIS which are much more involved in collaborative, multinational research projects. Therefore, while they utilized these temporary advisors, they did not have a need to establish dedicated support for each of the different types of personal data as DANS and GESIS did, in addition to the establishment of a data protection officer. This aspect, however, has no place in the original framework of informational privacy that I proposed at the start of this dissertation. Therefore, my findings indicate that this
framework does provide a way to understand specific organizational changes intended to maximize data and data subject privacy protection through revisions to or replacement of existing policies and procedures related to personal data, whether in the data archive or used elsewhere in the organization. But, it is missing something.

I return now to the first of my overarching research questions that guided this dissertation: How do the ways in which General Data Protection Regulation has been enacted increase our understanding about approaches to privacy in data archives? I conceived of this framework as a way to understand whether and how “informational privacy” had been reconceptualized at European Union-based social science data archives. However, as I collected and analyzed my data, I realized that the model instead reflects observable “informational privacy behaviors,” which I define here as

the actions and decisions organizations take to protect informational privacy in the form of policies, documentation, and processes, whether the control of the personal information remains with the data subject or is controlled by an organization on the data subject’s behalf.

“Informational privacy” of data subjects, whether research participants, employees, or users of the online platforms, was always a concern for data archive staff, and my findings indicate that the GDPR did not lead to any re-conceptualizations of privacy in those areas. As discussed in Section 5.2.3, the importance of informational privacy, and the archives already existing responsibility to protect data subjects, was reinforced, and all for organizations continue their efforts to raise awareness about it and increase their understanding of compliance.

My findings demonstrate that the changes ČSDA, FSD, DANS, and GESIS made to their operations from 2016-2021 in order to be and be able to demonstrate that they were compliant with the GDPR and revised national and sub-national data protection laws ultimately fall into four categories. These are the three dimensions from the theoretical framework—regulating the
flow of personal data through the data archive, minimizing data subject harm, and responding to technological risks to privacy—as well as changes in how data protection and privacy concerns are managed. However, these changes are influenced significantly by the three mediating factors discussed earlier in this chapter (Section 5.2.1). Therefore, the theoretical framework that I began this dissertation using is more accurately depicted as follows:

![Theoretical framework of informational privacy behaviors in data archives](image)

*Figure 5.1 Theoretical framework of informational privacy behaviors in data archives*

Of the three mediating factors, two are external to the data archive: the national and sub-national legislation derogations from the GDPR and the influence of a parent or supervisory organization. As discussed in Sections 5.2.1.1 and 5.2.1.2, legislative delays influence the how long it takes an organization to respond to the new legislation, whether there is an intermediary organization they must wait for, such as at DANS and ČSDA. Legislative requirements and supervisory mandates also may influence which processor or controller decision is made for which personal data at the organization, as indicated by the small arrow between the external and
internal mediating factors in Fig. 5.1. These mediating factors together influence what then happens at the data archive.

In this new framework, developed based on the findings in this dissertation, each of the three dimensions of the original framework remain. However, instead of representing specific approaches to privacy themselves, the three dimensions represent categories of organizational actions or behaviors that promote those aspects of informational privacy within a data archive, represented by the three arrows connecting each behavior category to the resulting informational privacy. The bi-directional arrow between the three categories of actions represent opportunities for overlapping motivations for the same behaviors; this was observed at my field sites in the two different motivations (to regulate data flows and to minimize harm) behind the new requirements that data providers demonstrate consent for archiving, for example. The red box surrounding right side of the model represents the fourth set of organizational actions. This fourth aspect itself does not result in informational privacy, but rather is the management component that mediates and directs the specific behaviors represented by the other three dimensions.

As the types of data and the methods of collecting and processing continue to advance, and as data archive holdings of data continue to grow, it has become more and more necessary for there to be someone (at least one person) at the organization with the expertise and knowledge of both the legal expectations around data protection and privacy and the needs of the data themselves, and to coordinate with external data protection experts. The complexities and influences of the internal and external factors on the data archives support O’Toole and Meier’s (2015) argument that external and internal contexts affect management of public organizations while reinforcing that the management of these implementation strategies is vital. As DANS and GESIS determined as well, because of the size and complexity of their organizations and the
diverse needs of the different categories of personal data they worked with—research data in the
data archives, research data and administrative data from internal and external research projects,
and the administrative and operational data used for the data archive’s daily operations—they
needed more than just one data protection officer to take on this role. They needed dedicated
internal expertise as well, in the form of legal advisors, privacy coordinators focusing on specific
types of personal data, and other organizational representatives.

The importance of this new framework is that, by centering informational privacy as a
possible result of specific organizational behaviors intended to comply with data protection
requirements, organizations can identify which processes and policies should be reviewed, and
potentially changed, in order to protect and preserve it. Once those reviews have been conducted
(such as through a data protection impact assessment or similar in-depth review), specific
changes or new requirements can be implemented. The inclusion of the management component
is vital as well, as it serves to remind organizations to not simply look at the processes and
policies; they need to consider how the immediate organizational changes will be coordinated to
ensure effective and ready compliance, and also how that compliance can be maintained over
time. Including the mediating factors in the model identifies what additional internal and
external may influence those behaviors. As ČSDA04 reminds us, “GDPR data protection is a
process, not just a one-point-in-time thing,” and it is a process that demands effective
management.

5.4 Contributions

This dissertation makes three main contributions to the body of knowledge around social
science data management and data archiving. The first is a new theoretical framework for
understanding how informational privacy is created and maintained at data archives. I originally
proposed a theoretical framework to better understand informational privacy itself in this context, because of a perceived inadequacy of previous conceptualizations that centered privacy on the data subject itself. The first proposed framework (Figure 2.1) conceived of it as the result of three connecting but distinct conceptualizations of privacy interwoven together. In this framework (Fig. 5.1), informational privacy exists when harm to the data subject, from advance capabilities to collect and process data are used in ways the data subject did not intend, is prevented. This framework recognizes that, for personal data in a data archive this prevention may be outside of the data subject’s personal control, in contrast to many views of privacy when center the data subject as the sole controller (e.g., Allen, 2011; Clarke, 2016; Junge, 2018; Tavani, 2008).

However, my findings demonstrate that what this framework models instead are the data archive’s actions taken to preserve and maintain a data subject’s informational privacy. When the interviewees discussed their organizations processes and changes, they framed them as mechanisms for maintaining or improving the ability to protect privacy, and not how privacy itself was conceptualized. Pre-GDPR data archive policies and processes regarding personal data, whether in the research data or used for the archive’s operational needs, addressed at least one of the three dimensions, preserving privacy by regulating how the data flowed through the organization, actively minimizing the privacy risks through controlling access, and implementing data protection measures in the IT systems and training archive staff in data protection and security. All but a few changes or new developments at ČSDA, FSD, DANS, and GESIS were made to improve the efficacy of the policies and procedures along those same dimensions.

Where the proposed framework also anticipated significant interplay between all three dimensions, my findings indicated that where there was overlap in motivations or interpretations
of specific changes, this only occurred between two: behaviors aimed at regulating information flows and minimizing harm.

The remaining changes represent an expansion on the proposed framework: the management of informational privacy behaviors. As discussed in the case reports and in Section 5.3, data archives, especially larger research infrastructures like DANS and GESIS which contend with not only large research data collections with personal data, but external projects and collaborations and their own administrative and operational data needs, found themselves needing to identify and document all of their personal data processing as part of their response to the GDPR. Research data and administrative and operational data differ in a number of ways, including their legal bases for collecting and processing, who may use or reuse them and how, and how long the data archive intends to retain them. No single policy exists at any data archive to direct processing of these different data sources in the same way, because the data and their purposes are so different. A particular challenge all four data archives encountered (to varying levels of disruption) related to internal organizational expertise in data management (research data or administrative and operational data) and in the legal requirements for data protection. Therefore, the final version of my theoretical framework of informational privacy behaviors (Fig. 5.1), identifies each of the three categories of privacy protecting behaviors, inside the fourth aspect, privacy protection management. This framework enables institutions to similarly evaluate their own data and privacy protection policies and procedures based on the intended outcomes and understand whether and how those efforts are coordinated and managed.

While social science data archives and repositories have formally existed for over sixty years, recent developments and operational changes in response to growing technological risks to privacy have not been studied in the context of these data archives. The 2016 adoption of the
General Data Protection Regulation brought uncertainty and confusion to the European Union’s social science research community; from the perspectives of the archives, this confusion was exacerbated by a lack of clear guidance and consistent interpretations of just what the GDPR would mean for research, and the personal data frequently used in that research, across the social science disciplines. Since that time, there has been very little discussion of the impacts of the GDPR on research data management and archives in the social sciences either from a research perspective or from a practical guidance perspective.

This dissertation addresses both of these limitations in succession. The second contribution is the empirical study of four social science data archives in the aftermath of the GDPR. Unlike the literature referencing that GDPR was considered during data archive development (e.g., Lee et al., 2022; van den Heuvel et al., 2020) or contemplating possible effects on research in the social sciences (e.g., Breuer et al., 2021; Mourby et al., 2018; Yuill, 2018), my research reports on the process of GDPR adaptation, the motivations and considerations behind what changes were made, and what challenges they faced. I have identified the three primary factors influencing the process and outcomes of these changes—the choice of processor or controller role for personal data, the national and sub-national legislation, and the presence and influence of a parent or supervisory organization (Section 5.2.1). My findings also indicate that the predicted effects of the GDPR on data archiving and reuse after 2016, if they occurred, were only short-term effects that cannot be connected exclusively to the GDPR (Section 4.5.5).

From this second contribution to research on data archive practices and considerations when working with personal data, comes the third contribution of this study: practical guidance for data archives and institutions addressing data protection requirements and multiple different
types of personal data. To develop a list of potential types of policies or other documentation where changes might be visible, I had to rely on models and reports of organizational changes from other industries. The types of changes discussed in Chapter 4, and the lists of reviewed policy documents in Appendix E (Tables E.1-E.4), provide a practical guide to other data archives which can be tailored to their own needs.

As discussed at the start of each case report, the four data archives in this dissertation have important places within their national research infrastructures. All four are members and national nodes of the European-wide CESSDA, and each archive is the main data archive in each country for social science research data. Despite their joint and long-standing membership in CESSDA, and involvement in numerous CESSDA and EU-focused infrastructure and research projects, as discussed in this dissertation, the archives’ GDPR experiences, policies, practices, and limitations are not identical. Instead, they heavily reflect their different legal and supervisory contexts. As exemplars of four different experiences in four different contexts, their experiences and insights (and the practical guide discussed above) provide valuable lessons to other research institutions and data archives, whether in their own countries or across Europe and the world. I also identified the choice of processor or controller as an important factor to how the four data archives responded to the GDPR. The influence this choice has over where, and what, changes are required is also an informative insight for future institutional responses.

Indeed, the key lessons learned from my four case studies are as follows:

- processor vs controller decision will influence what changes will be made;
- changes to existing policies and documentation will focus on those affecting the flow of data into and out of the archive and on those intended to minimize data subject harm;
• if there are multiple levels of legal obligation (state vs. national), depending on the system, there may be conflicts in applications;

• organizations with parent institutions may face delays in finalizing plans if required to wait for the parent’s response; and,

• dedicated internal expertise with both legal aspects of data protection and research data management is vital.

5.5 Future work

One of the motivations of this dissertation was to address the limited empirical research on the outcomes of the GDPR on research infrastructures, and the specific lack of research about social science data archives in particular. My four case studies and the preceding discussion have contributed to efforts to remedy that gap. However, more work is necessary to fully understand the effects of the GDPR on the research community. My dissertation focuses on four data archives supporting one specific (though broad) discipline area: social science. However, the various social science fields are not the only ones to use (or want to use) personal data for research purposes. The next steps include expanding the scope of the comparative case study to include data archives in these other fields, including health, genomics, and other life sciences research. For example, the COVID-19 pandemic and the increase in multinational medical and public health research has also resulted in both the establishment of new data sharing portals and new collections of COVID-19 related research data in existing data archives (including in the social sciences). While, as discussed, there were a lot of commentaries and reports of GDPR’s impact on medical research, the outcomes for medical research data infrastructures, as with the social sciences, is also limited.

One important finding of my dissertation is that, at these four data archives, the GDPR did not produce any measurable significant, long-term changes in data set publications or use, and neither did data archive staff consider the GDPR to do so. Staff at GESIS and ČSDA, which
did see short term decreases in data deposit during the transition period, hypothesized that it was possible that, for a time after the GDPR entered into force, data providers were more hesitant or cautious about sharing their data. However, neither archive discussed this with their data producers in any formal or informal process, and this dissertation likewise only focused on the data archives themselves.

Expanding the archive’s stakeholder groups (initially conceived of for this dissertation to be advisory boards or other sources of external oversight) to include data producers and providers is another path for future research. A future study that investigates the data provider’s perspectives on and experiences with the GDPR and their research, and with the data archives changing policies, may provide additional insight into the operational impacts of the GDPR, as well as insights into the practical effects the GDPR has had on research within the EU.

This study focused on data archives within the European Union. However, as discussed in Chapter 2, the number of global data protection laws has been growing since 2016, and many of the new laws are modeled on the GDPR itself. Data archives and research infrastructures responding to these new laws have faced the similar challenge that motivated this dissertation—a lack of guidance, best practices, and research on the implications of the GDPR in their contexts—and as with my four field sites, had to find their own way towards compliance. A study that focuses on non-EU-based data archives responding to new data protection legislation (whether modeled on the GDPR), could provide further insight into operational impacts of privacy legislation, what changes are deemed necessary and why, and also expand our understanding of how informational privacy is enacted in regions where the Eurocentric histories of personal data abuse are more removed.
Finally, a lingering question from this dissertation remains: for data archives and the research data with personal data that they work with, does the choice of processor or controller truly matter? As discussed in Chapter 4, the question of whether the best choice was made is still being debated at at least one EU-based data archive: DANS. The primary deficiency in the decision to remain in the “processor” role, according to several DANS interviewees, was in the limited ability to continue in their mission for long-term data preservation and dissemination in support of science, when the controller may no longer be reachable or when the controller could decide to remove their data from the public at will. However, at the same time, because the controller is able to process data while also determining how and why the processing is done, and because the processor must also comply with data protection requirements, from those perspectives, being the processor is not that different from being the controller. The difference lies in the level of responsibility over the data and accountability for compliance.

This is a significant difference in terms of liability. But does it truly make a difference for the research data, or the operation of the data archive? In this dissertation, one data archive claimed the role of “controller” for these data, two claimed the role of “processor,” and one claimed neither role for the data it made publicly available. An investigation of other data archives or research infrastructures, selected based on their choice of processor, controller, and/or neither, may provide insight into whether these roles matter for data archives, and if so, how. The GDPR has been criticized since it was first adopted for not sufficiently recognizing the valuable utility of personal data and special categories of personal data for research and knowledge creation; these criticisms can also be made about the national implementations of the GDPR. As a regulation primarily targeting the “big” data industries, a common frustration during the adaptation period was how this law that was not aimed at research would affect
research and research infrastructures, and whether the obligations and requirements in the GDPR would or could actually be applicable or effective. This question about the processor and controller role for the archived research data that DANS continues to wrestle with reflects this and raises that broader question of if the roles as currently conceived of ought to be applied to the data archiving work of research infrastructures at all. A deeper understanding of what these roles mean at, and what each role expects from, data archives and research infrastructures would not only better support these institutions themselves. It would also provide valuable feedback to the European Commission and the Member States as they prepare revisions and updates to the GDPR and associated national laws in the coming years, and support development of guidance about for data archives as they adopt or revise their roles regarding their research data.
Appendices

Appendix A Data Archive Staff Interview Protocol

Respondent background
- What is your academic background?
- How long have you been working at ARCHIVE?
- What positions have you held at ARCHIVE?
- Prior to your position here, have you had any experience as a data manager or primary/principle investigator?

Questions about current work
- I would like to start by hearing more about your current position.
  - What is your current position?
  - (If applicable) Were you working in this position prior to May 2018, when the GDPR went into effect? (If not) What was your position at that time?
  - (If applicable) How has your role changed since GDPR went into effect?
- In your current position, do you interact with data producers/data depositors/principle investigators? (If yes) What types of interactions do you have?
- In your current position, do you interact with potential data users (specifically, those who are interested in using data at ARCHIVE)? (If yes) What types of interactions do you have?

Institutional policies and procedural changes
- To what extent has your institution transitioned to meet the GDPR and/or national legal implementation requirements?
  - (If not complete) What remains to be done?
  - How did you prioritize what was done first?
• What challenges did ARCHIVE (or your department) face when adapting to the GDPR?

• How much discretion did you/your institution have in this transition?

• How involved were you in the transition process?

• Please discuss the transition process.
  ▪ When did it begin? Who was involved? What were the points of contention?

• Was input sought and/or provided by external institutions or persons?

• What data protection measures are put in place for data used from ARCHIVE?
  ▪ How were these measures developed/determined?
  ▪ Were new measures for data storage and protection put in place to ensure data protection?
  ▪ How were ARCHIVE staff trained to the new requirements?
  ▪ Did ARCHIVE staff access to data or certain positions change in response to the new legal requirements?

• Have you received feedback from data producers or data users about how GDPR has affected their work?
  ▪ (if yes) What type of feedback did you receive?

• What changes to data acquisition/ingest policies and practices were implemented in response to the GDPR?
  ▪ How were these changes developed?
  ▪ Were data producers consulted about planned/discussed changes to policies or practices?
  ▪ How have data producers responded to these changes?
    ▪ How were data producers made aware of these changes?
    ▪ What feedback have data producers provided regarding their research and GDPR’s requirements?
    ▪ What questions have data producers had about GDPR and ARCHIVE?
    ▪ Have data producers expressed concerns about sharing their data?
    ▪ (If yes) What are those concerns?

• What changes to data use practices and policies were implemented in response to the GDPR?
  ▪ How were these changes developed?
Were data users consulted about planned/discussed changes to policies or practices?

How have data users responded to these changes?

- How were data users made aware of these changes?
- Has access to or requests for data at your archive changed since GDPR?
- What questions have data users had about GDPR and their ability to use data from your archive?
- Have data users expressed concerns about using data?
- (If yes) What are those concerns?

Are your data used by researchers outside of the European Union?

- Does your ARCHIVE archive data produced in another European Union Member State?
  - (If yes) What is the process for bringing those data into ARCHIVE?
  - (If yes) Prior to GDPR’s adoption, what challenges did ARCHIVE face in sharing these data between Member States? (probe if necessary: What challenges in acquiring the data? What challenges in sharing the data with new users?)
  - (If yes) Is this process easier/less complex than before GDPR’s adoption?

- What type of oversight does ARCHIVE have…?
  - To what extent does a Data Protection Officer or Data Protection Authority currently influence ARCHIVE policies and practices?
  - What role did they play in the transition process?
  - Have there been conflicts in priorities between ARCHIVE and the Data Protection Officer or Data Protection Authority?

Questions about data and the archive

- What type or types of data does your archive ingest and make available for later use?
- Are there any data with use restrictions?
- What is the workflow/process for evaluating research data once they are deposited/shared with ARCHIVE?
- What criteria are used to determine how the data are made available for later use?
  - How were those criteria established at ARCHIVE?
  - How are those criteria evaluated?
- Are data containing identifiable information (to include information identified as “special” under the GDPR) archived at ARCHIVE?
• (If no) Why are these data not archived?

Are data which have been anonymised or pseudonymised archived at ARCHIVE?
• (If yes) Who is responsible for anonymising or pseudonymising these data?
• What components of the data are anonymised or pseudonymized?

Questions about the GDPR

• What does privacy mean to you?
• What was the reaction in COUNTRY to the GDPR?
• What has been the most significant change at ARCHIVE in response to the GDPR?
  • How has data acquisitions changed in response to the GDPR?
  • How has data reuse changed in response to the GDPR?
• What barriers are there for data producers to share their data with ARCHIVE?
• What barriers exist for data users who want to use data from ARCHIVE?
• How are your data used by users outside of the European Union?
  • How do the GDPR’s “adequacy decision” requirements about non-European Union countries affect non-European Union data use?
• In your opinion, what is the most important impact/component of the GDPR?
  • Why do you say this?
  • How well is this opinion reflected in ARCHIVE’s response to GDPR?
• In your opinion, what is the greatest challenge of the GDPR?
  • Why do you say this?
  • How well is this opinion reflected in ARCHIVE’s response to GDPR?
• Do you think that the GDPR is an effective tool to protect data subject privacy?
  • Why do you say this?
  • How widely shared is your opinion reflected in ARCHIVE’s response to GDPR?
• Do you think that GDPR goes too far in its requirements for data protection?
  • Why or why not?
• Do you think that GDPR promotes data sharing?
• What measures or changes would you recommend to improve its effectiveness in protecting data subject privacy while also encouraging data use?
Wrap-up questions

- Is there anything about ARCHIVE’s response to the GDPR that you think I should know about but that I did not ask you?
- Is there anyone else that I should talk to?
- Do you have any questions for me about this project?
Appendix B Legal Interview Protocol

Respondent background

- What is your academic background?
- How long have you been working in the legal field?
- Prior to your work with the ARCHIVE, did you have any experience working with research data?
  - [Follow-up] with special categories of data?
- Prior to your work with the ARCHIVE, what experience did you have with data protection and privacy law?

Questions about work with ARCHIVE

- I would like to start by hearing more about your work with the ARCHIVE.
  - When did you begin working with ARCHIVE? Are you still working with ARCHIVE?
  - Why did you begin working with ARCHIVE?
  - What was your role with the ARCHIVE? [probe: consultant? Employee?]
- Who were you working with at the ARCHIVE while you were there? (position, not their name)
  - In what capacity?

Institutional involvement

- How involved were you in ARCHIVE management’s discussions about addressing the GDPR?
- How involved were you in ARCHIVE’s transition process?
- To what extent did you interact with an institutional Data Protection Officer?
  - In what capacity?
• To what extent has ARCHIVE transitioned to meet the GDPR and/or national legal implementation requirements? [IF no longer affiliated, change to “to what extent had your…”]
  o (If not complete) What remains to be done? [IF no longer affiliated, change to “what remained to be done?”]

• Did you seek advice or guidance from anyone or any organization outside of ARCHIVE during your time there?
  o (if yes) Who did you consult with?
  o (if yes) What type of guidance were you looking for?

• To what extent did you interact with the Data Protection Authority [Data Protection Ombudsman]?
  o Have there been conflicts in priorities between your archive and the Data Protection Officer or Data Protection Authority? [IF no longer affiliated, change to “were there conflicts…?”]

GDPR transition experience

• In COUNTRY, what was the most significant change that resulted from GDPR?

• How different is COUNTRY’S GDPR implementation from the previous data protection law?

• How different is COUNTRY’s GDPR implementation from the GDPR itself?
  o Why are there differences? [Probe: due to COUNTRY approach to international law incorporation into national law? Culture? Legal system structure?]

• What legal concerns must data archives in COUNTRY consider when archiving personal data?
  o How prevalent is personal data archiving in COUNTRY?
  o Are these legal concerns shared by all data producers?
  o Is personal data processing by public institutions regulated in the same way as processing by private institutions?

• What legal concerns must data archives in COUNTRY consider when sharing data with non-EU/EEA data users?
  o What legal barriers are there for personal data sharing in COUNTRY?

Opinion questions about the GDPR

• What does privacy mean to you?
• What was the reaction in COUNTRY to the GDPR?
• What has been the most significant change at ARCHIVE in response to the GDPR?
• In your opinion, what is the most important impact/component of the GDPR?
  o Why do you say this?
  o How well is this opinion reflected in your archive’s response to GDPR?
• In your opinion, what is the greatest challenge of the GDPR?
  o Why do you say this?
  o How well is this opinion reflected in your archive’s response to GDPR?
• Do you think that the GDPR is an effective tool to protect data subject privacy?
  o Why do you say this?
  o How widely shared is your opinion reflected in your archive’s response to GDPR?
• Do you think that GDPR goes too far in its requirements for data protection?
  o Why or why not?
• Do you think that GDPR promotes data sharing?
• What measures or changes would you recommend to improve its effectiveness in protecting data subject privacy while also encouraging data use?

Wrap-up questions

• Is there anything about ARCHIVE’S response to the GDPR that you think I should know about but that I did not ask you?
• Is there anything about your role in ARCHIVE’S response to the GDPR that you think I should know about but that I did not ask you?
• Is there anyone else that I should talk to?
• Do you have any questions for me about this project?
Appendix C Archive Stakeholder Interview Protocol

Respondent background

• What is your academic background?
• What is your current position?
• How long have you been working with ARCHIVE?
• Prior to your work with the ARCHIVE, did you have any experience working with research data?
  o [Follow-up] with special categories of data?
• Prior to your work with the ARCHIVE, what experience did you have with data protection and privacy law/regulations?

Questions about work with ARCHIVE

• I would like to start by hearing more about your work with ARCHIVE.
  o When did you begin working with ARCHIVE? Are you still working with ARCHIVE?
  o Why did you begin working with ARCHIVE?
  o What was your role with the ARCHIVE? [probe: consultant? Employee?]
• Who were you working with at the ARCHIVE while you were there? (position, not name)
  o In what capacity?

Institutional involvement

• What were the SUPERVISORY ORG’s goals for the GDPR response?
  o How were those goals communicated?
  o What were the priorities in the transition process?
  o How were challenges and disagreements between SUPERVISORY ORG’s and ARCHIVE’s response settled?
• What feedback and/or concerns were you receiving from researchers and other data stakeholders in response to GDPR?
  o What was SUPERVISORY ORG’s response to this feedback?

• What external pressure was felt by SUPERVISORY ORG (e.g., governmental pressure?) in relation to GDPR?
  o How did that affect the ARCHIVE’s responses?

• How involved were you in ARCHIVE management’s discussions about addressing the GDPR?
  o What was your role?
  o What challenges did ARCHIVE face in its transition process?

• How involved were you in ARCHIVE’s transition process?

• To what extent did you interact with an institutional Data Protection Officer?
  o In what capacity?

• What changes were made at ARCHIVE in response to the GDPR?

• To what extent has ARCHIVE transitioned to meet the GDPR and/or national legal implementation requirements? [IF no longer affiliated, change to “to what extent had your…”]
  o (If not complete) What remains to be done? [IF no longer affiliated, change to “what remained to be done?”]

• Did you seek advice or guidance from anyone or any organization outside of ARCHIVE during your time there?
  o (if yes) Who did you consult with?
  o (if yes) What type of guidance were you looking for?

• To what extent did you interact with the national Data Protection Authority [Data Protection Ombudsman]?
  o Have there been conflicts in priorities between your archive and the Data Protection Officer or Data Protection Authority? [IF no longer affiliated, change to “were there conflicts…?”]

Opinion questions about the GDPR

• What does privacy mean to you?
• What was the reaction in COUNTRY to the GDPR?
• What has been the most significant change at ARCHIVE in response to the GDPR?
• In your opinion, what is the most important impact/component? of the GDPR?
• In your opinion, what is the greatest challenge of the GDPR?
  o Why do you say this?
  o How well is this opinion reflected in your archive’s response to GDPR?
• Do you think that the GDPR is an effective tool to protect data subject privacy?
  o Why do you say this?
  o How widely shared is your opinion reflected in your archive’s response to GDPR?
• Do you think that GDPR goes too far in its requirements for data protection?
  o Why or why not?
• Do you think that GDPR promotes data sharing?
• What measures or changes would you recommend to improve its effectiveness in protecting data subject privacy while also encouraging data use?

Wrap-up questions
• Is there anything about ARCHIVE’S response to the GDPR that you think I should know about but that I did not ask you?
• Is there anything about your role in ARCHIVE’S response to the GDPR that you think I should know about but that I did not ask you?
• Is there anyone else that I should talk to?
• Do you have any questions for me about this project?
### Appendix D Qualitative Code Set

Table D.1 Qualitative code set used for interview and archive documentation analysis.

<table>
<thead>
<tr>
<th>Parent Code</th>
<th>Child Code</th>
<th>Sub Child Code</th>
<th>Definition</th>
<th>Notes/Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicable GDPR article</td>
<td></td>
<td></td>
<td>Use for direct references to requirements dictated in GDPR articles (USE even if the article number is NOT specifically mentioned)</td>
<td></td>
</tr>
<tr>
<td>Archive history</td>
<td></td>
<td></td>
<td>Use for discussion of history/development/establishment of archive</td>
<td></td>
</tr>
<tr>
<td>Archive operations</td>
<td></td>
<td></td>
<td>General code for discussions of &quot;how things work&quot; at the archive/organization (also use with respondents who do not directly work in the archive/organization, when describing how they work with their own institutions - e.g., a staff member of a parent institution, talking about higher level interactions, etc.)</td>
<td>May be dual coded with some of the &quot;Respondent information&quot; codes</td>
</tr>
<tr>
<td>Archive services</td>
<td></td>
<td></td>
<td>These codes are to be used with discussions of services that the data archive itself offers for data that are deposited/archived there. These codes are NOT to be used with other services that the institution may offer (such as non-archive provided training, community outreach, etc.) that are not led by archive staff.</td>
<td>Only use this parent code if the child codes &quot;anonymization or pseudonymization,&quot; &quot;Translation,&quot; &quot;Curation,&quot; &quot;Advising,&quot; or &quot;Geo-referencing&quot; are not applicable.</td>
</tr>
<tr>
<td>Provided by archive</td>
<td></td>
<td></td>
<td>Use for discussion of archive services provided by the archive</td>
<td>Dual code with the other child codes if applicable</td>
</tr>
<tr>
<td>Not provided by archive</td>
<td></td>
<td></td>
<td>Use for discussion of archive services not provided by the archive</td>
<td>Dual code with the other child codes if applicable</td>
</tr>
<tr>
<td>Anonymization or pseudonymization</td>
<td></td>
<td></td>
<td>Use for discussion of archive provided anonymization/pseudonymization services</td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td></td>
<td></td>
<td>Use for discussion of archive provided translation services</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Curation</td>
<td></td>
<td></td>
<td>Use for discussion of archive provided data curation services</td>
<td>Curation services can include: ingest checking, data cleaning, reformatting, creating/adapting metadata, checking for identifiers (disclosure risk), documentation development, etc.</td>
</tr>
<tr>
<td>Geo-referencing</td>
<td></td>
<td></td>
<td>Use for discussion of archive provided geo-referencing services</td>
<td></td>
</tr>
<tr>
<td>Advising</td>
<td></td>
<td></td>
<td>Use for discussion of providing advice to data producers or data users about the GDPR, consent, data management, archiving standards, etc.</td>
<td></td>
</tr>
<tr>
<td>Access to sensitive data</td>
<td></td>
<td></td>
<td>Use for discussion about providing access to sensitive/risky data through secure rooms, etc.</td>
<td></td>
</tr>
<tr>
<td>Data linking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archive role</td>
<td></td>
<td></td>
<td></td>
<td>DO NOT USE PARENT CODE</td>
</tr>
<tr>
<td>Processor</td>
<td></td>
<td></td>
<td>Use for discussion of role of the archive as a processor under the GDPR</td>
<td>Art. 4(8) - ‘processor’ means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;</td>
</tr>
<tr>
<td>Controller</td>
<td></td>
<td></td>
<td>Use for discussion of role of the archive as a data controller under the GDPR</td>
<td>Art. 4(7) - ‘controller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;</td>
</tr>
<tr>
<td>Country specific contexts</td>
<td></td>
<td></td>
<td></td>
<td>Use this code only if one of the child codes are not applicable</td>
</tr>
<tr>
<td>National legislation vs GDPR</td>
<td></td>
<td></td>
<td>Use for discussion or comparison of GDPR requirements and the national/state-level implementations of GDPR in the local context</td>
<td>Use even if Respondent says there are no differences</td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Reaction to GDPR within country</td>
<td>Cultural view of privacy</td>
<td></td>
<td>Use for discussion of the national (not necessarily respondent-specific) reactions to the GDPR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legal status of privacy</td>
<td></td>
<td>Use for when the respondent discusses privacy from the perspective of the culture/country the archive is in and/or the respondent is from/currently in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social factors for privacy</td>
<td></td>
<td>Use for when the respondent discusses privacy from a societal level, including discussions/comments on surveillance, influence of politics, etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Historical influences on privacy</td>
<td></td>
<td>Use for when the respondent discusses historical experiences of privacy or privacy violations in the country the archive and/or the respondent are in/from</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Status of data archiving</td>
<td></td>
<td>Use for discussion of the prevalence and/or status of data archiving, data sharing, data use, or data processing in the country the archive/organization is in</td>
<td></td>
</tr>
<tr>
<td>Status of research</td>
<td></td>
<td></td>
<td>Use for discussion about research, research practices, within country the archive/organization is in</td>
<td></td>
</tr>
<tr>
<td>Data at the archive</td>
<td>Organizational administrative data</td>
<td></td>
<td>Referring to data that the archive/organization uses for its own operation (including data about staff, customer data, user data, etc.)</td>
<td>Use the child codes to identify types of data that the archive or respondent works with. Use all that apply.</td>
</tr>
<tr>
<td></td>
<td>Survey data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social media data</td>
<td></td>
<td>May also be referred to as &quot;digital trace data&quot; or &quot;digital behavioral data&quot; etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Humanities data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social science data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Life sciences data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualitative data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archaeology data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from archive-led projects</td>
<td></td>
<td></td>
<td>Use for data from projects that the archive/organization leads or is involved in which are archived at the archive</td>
<td></td>
</tr>
<tr>
<td>Audio/visual data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data formats</td>
<td></td>
<td></td>
<td>General code for when the respondent talks specifically about data formats that the archive can OR cannot accommodate</td>
<td></td>
</tr>
<tr>
<td>Data protection and security measures</td>
<td></td>
<td></td>
<td>Only use if there is no applicable child code</td>
<td></td>
</tr>
<tr>
<td>Geographically distinct storage</td>
<td></td>
<td></td>
<td>Use for discussion of storing data in different places as data protection measure</td>
<td></td>
</tr>
<tr>
<td>Encrypted storage</td>
<td></td>
<td></td>
<td>Use for discussion of encrypted data storage as data protection measure</td>
<td></td>
</tr>
<tr>
<td>Encrypted transfer</td>
<td></td>
<td></td>
<td>Use for discussion of encrypted data transfer/transmission as data protection measure</td>
<td></td>
</tr>
<tr>
<td>Anonymization</td>
<td></td>
<td></td>
<td>Use for discussion of anonymization/pseudonymization specifically as data protection measure</td>
<td></td>
</tr>
<tr>
<td>Rejection of data</td>
<td></td>
<td></td>
<td>Use for discussion of rejecting the archiving of data as a data protection measure</td>
<td></td>
</tr>
<tr>
<td>Access restrictions</td>
<td></td>
<td></td>
<td>Use for discussion of restrictions on data access through different access categories or classifications of the data as a data protection measure</td>
<td></td>
</tr>
<tr>
<td>Approval required</td>
<td></td>
<td></td>
<td>Use if the access restriction requires explicit approval from the archive or the data producer in order for a data user to access the data.</td>
<td></td>
</tr>
<tr>
<td>Usage contract required</td>
<td></td>
<td></td>
<td>Use for discussion of a usage contract as a data protection measure (including requirements or</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>questions about data protection that must be established in the contract for access to the data</td>
<td></td>
</tr>
<tr>
<td>Account registration</td>
<td></td>
<td></td>
<td>Use for discussion of requirements to register for an account in order to request/download data</td>
<td></td>
</tr>
<tr>
<td>Archive responsible</td>
<td></td>
<td></td>
<td>Use for discussion of the archive/institution as the one responsible for data protection</td>
<td>May be dual coded with above codes</td>
</tr>
<tr>
<td>Data provider responsible</td>
<td></td>
<td></td>
<td>Use for discussion of the data provider as the one responsible for data protection</td>
<td>May be dual coded with above codes</td>
</tr>
<tr>
<td>Deposit review</td>
<td></td>
<td></td>
<td>Use for discussion of reviewing/checking data upon deposit for identifiers (direct and indirect), other privacy risks, (as well as other components of the deposit review).</td>
<td>May be used WITHOUT child codes if the reference to deposit review only talks about “checking for identifiers” but without specifics</td>
</tr>
<tr>
<td>Direct identifiers</td>
<td></td>
<td></td>
<td>Use for discussion of reviewing/checking deposited data for direct identifiers</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td>Geographical information</td>
<td></td>
<td></td>
<td>Use for discussion of reviewing/checking deposited data for geographical information as a privacy risk</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td>Other indirect identifiers</td>
<td></td>
<td></td>
<td>Use for discussion of reviewing/checking deposited data for other indirect identifiers (excluding geographical information)</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td>Data protection impact assessment</td>
<td></td>
<td></td>
<td>Use for discussion of conducting a data protection impact assessment</td>
<td></td>
</tr>
<tr>
<td>Response to data breach</td>
<td></td>
<td></td>
<td>Use for discussion of what happens when/if there is a data breach, personal data violation, hack, etc.</td>
<td></td>
</tr>
<tr>
<td>Retention limits</td>
<td></td>
<td></td>
<td>Use for discussion of limits on how long the archive can keep the data. Use for BOTH specific references to GDPR guidance about data retention and other limits on retention from local laws, organizational practices, etc. DO NOT USE FOR EMBARGO PERIODS.</td>
<td></td>
</tr>
<tr>
<td>Restrictions on third-party applications</td>
<td></td>
<td></td>
<td>Use for discussion of limits, restrictions, evaluations, etc., regarding the use and/or installation of 3rd party applications (such as the Google Suite, Dropbox, etc. that were not developed by the institution or affiliated institution). This may include discussions of requiring specific contracts governing that usage.</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Feedback or questions about GDPR</td>
<td>From data users</td>
<td></td>
<td>Use for questions about the GDPR</td>
<td>Only use if any of the child codes are not applicable. Do NOT code for absence</td>
</tr>
<tr>
<td></td>
<td>From data producers</td>
<td></td>
<td>Use for questions from data producers to the archive about the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>From archive or institution staff</td>
<td></td>
<td>Use for questions from archive/institution staff about the GDPR (including staff within the archive)</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td>Types of feedback or questions</td>
<td></td>
<td></td>
<td></td>
<td>USE WITH CHILD CODES</td>
</tr>
<tr>
<td></td>
<td>Insecurity about data collection</td>
<td></td>
<td>Use for discussion about uncertainty/insecurity about data collection under the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Insecurity about data archiving</td>
<td></td>
<td>Use for discussion about uncertainty/insecurity about data archiving under the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Insecurity about data use</td>
<td></td>
<td>Use for discussion about uncertainty/insecurity about data use under the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Openness of the data</td>
<td></td>
<td>Use for discussion about uncertainty/insecurity about open data under the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Questions about informed consent</td>
<td></td>
<td>Use for discussion about informed consent under the GDPR</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Questions about merging datasets</td>
<td></td>
<td>Use for discussion of questions about whether they can, or how they can, merge data sets</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Questions about compliance</td>
<td></td>
<td>Use for discussion of questions about whether XX is in compliance with GDPR or other data protection laws</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Questions about data protection</td>
<td></td>
<td>Use for discussion about questions about data protection, data security, etc. that are NOT just about how to comply with the law</td>
<td>USE WITH PARENT CODE</td>
</tr>
<tr>
<td>GDPR transition challenges</td>
<td>Legal ability to ingest data</td>
<td></td>
<td>Use for archive/respondent questions or concerns about the legal ability to ingest/archive data</td>
<td>Use only with archive staff concerns</td>
</tr>
</tbody>
</table>

GDPR transition challenges: Parent code for ARCHIVE challenges (not data provider/user challenges or questions) May be dual coded with the child codes under "Organizational changes"
<table>
<thead>
<tr>
<th>Parent Code</th>
<th>Child Code</th>
<th>Sub Child Code</th>
<th>Definition</th>
<th>Notes/Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal ability to disseminate data</td>
<td></td>
<td></td>
<td>Use for archive/respondent questions or concerns about the legal ability to disseminate data</td>
<td>Use only with archive staff concerns</td>
</tr>
<tr>
<td>Confusion about requirements</td>
<td></td>
<td></td>
<td>Use for archive/respondent questions or concerns about the GDPR requirements that are/are not applicable to the archive</td>
<td></td>
</tr>
<tr>
<td>Recategorization of data</td>
<td></td>
<td></td>
<td>Use for discussion of needing to re-categorize data in the archive in response to the GDPR</td>
<td></td>
</tr>
<tr>
<td>Depositor contract updates</td>
<td></td>
<td></td>
<td>Use for discussion of needing to have data depositors re-sign/agree to new contract terms for data already in the archive</td>
<td></td>
</tr>
<tr>
<td>Lack of expertise in law and research data</td>
<td></td>
<td></td>
<td>Use for discussion that lawyers with experience with research data or data archives are very rare, which complicate the archive/institution approaches to GDPR</td>
<td></td>
</tr>
<tr>
<td>How GDPR applied to research</td>
<td></td>
<td></td>
<td>Use for discussion about questions/concerns/confusion about how GDPR applied to research, data reuse for research purposes, research data archives, etc.</td>
<td></td>
</tr>
<tr>
<td>Unclear definitions</td>
<td></td>
<td></td>
<td>Use for discussion of a challenge related to lack of clarity in definitions, principles, etc. in the GDPR</td>
<td></td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td></td>
<td>Use for discussions of challenges in compliance with the GDPR, enforcing compliance, being compliant “in time,” effort required for compliance, etc.</td>
<td></td>
</tr>
<tr>
<td>Ethical mindset vs fear-based compliance</td>
<td></td>
<td></td>
<td>Use for discussions about questions about ethical research under GDPR, concerns over people complying out of fear of penalties instead of ethical mindset, etc.</td>
<td></td>
</tr>
<tr>
<td>Points of contention</td>
<td></td>
<td></td>
<td>Use when respondent describes or discusses challenges in the GDPR response that specifically relate to tensions/disagreements between different areas of the organization in how to respond (not just different opinions that are held by different staff, but actual instances of policy-making disagreements)</td>
<td></td>
</tr>
<tr>
<td>Insufficient resources</td>
<td></td>
<td></td>
<td>Use when respondent describes or discusses a lack of resources (e.g., funding, manhours, materials,</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>GDPR transition oversight</td>
<td></td>
<td></td>
<td>staff, etc.) as a challenge or concern during the GDPR transition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Role of the DPO</td>
<td></td>
<td>Used for discussion about the role of the archive's data protection officer in influencing the archive's efforts to respond to the GDPR's requirements.</td>
<td>DO NOT USE PARENT CODE</td>
</tr>
<tr>
<td></td>
<td>Role of the DPA</td>
<td></td>
<td>Used for discussion about the role of the data protection authority in influencing the archive's efforts to respond to the GDPR's requirements.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>State or regional DPA</td>
<td></td>
<td>Use with the &quot;Role of the DPA&quot; code to identify that the DPA is a state/regional/district data protection authority</td>
<td>USE ONLY FOR GERMANY</td>
</tr>
<tr>
<td></td>
<td>Federal DPA</td>
<td></td>
<td>Use with the &quot;Role of the DPA&quot; code to identify that the DPA is the national data protection authority</td>
<td>USE ONLY FOR GERMANY</td>
</tr>
<tr>
<td></td>
<td>Role of parent organization</td>
<td></td>
<td>Use for discussion about the role of a parent/supervisory organization to the archive in the archive's efforts to respond to the GDPR's requirements.</td>
<td>Some parent organizations are also funders of the archives, so this code may be dual coded with &quot;Role of funder&quot; based on usage</td>
</tr>
<tr>
<td></td>
<td>Role of funder</td>
<td></td>
<td>Use for discussion about the role of an archive funder in the archive's efforts to respond to the GDPR's requirements.</td>
<td>Some parent organizations are also funders of the archives, so this code may be dual coded with &quot;Role of parent organization&quot; based on usage</td>
</tr>
<tr>
<td></td>
<td>Role of advisory boards</td>
<td></td>
<td>Use for discussion about the role of an advisory board in the archive's efforts to respond to the GDPR's requirements.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Role of external parties</td>
<td></td>
<td>Use for discussion of the role of other external parties in the archive's efforts to respond to the GDPR's requirements.</td>
<td></td>
</tr>
<tr>
<td>Impact of the GDPR</td>
<td></td>
<td></td>
<td>Use the parent code if child codes do not apply</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Added bureaucracy</td>
<td></td>
<td>Use for discussion of bureaucratic impact of the GDPR on the archive.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increased awareness</td>
<td></td>
<td>Use for when the respondent indicates that an impact of the GDPR is an increased awareness of</td>
<td>May dual code with the &quot;Privacy meaning&quot; child codes if needed</td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
<td>Sub Child Code</td>
<td>Definition</td>
<td>Notes/Rules</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>the importance of data privacy and/or data protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adequacy decision</td>
<td></td>
<td>Use for discussions about the impact of the adequacy decision requirements on the archive</td>
<td>Non-EU countries/users ability to use data from EU-based archives</td>
</tr>
<tr>
<td></td>
<td>Increased fear of penalties</td>
<td></td>
<td>Use for discussion of the GDPR increasing fears (within the archive, with data depositors, with data users, or other external parties) of the penalties from data use violations (fees, fines, etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Privacy vs usability</td>
<td></td>
<td>Use for discussion related to balancing data or privacy protection and the archive being able to provide usable/useful data for reuse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requests for data deletion</td>
<td></td>
<td>Use for discussion of receiving requests from data subjects to have their data deleted from the archive (from research data within the archive as well as data the archive has about customers/users/etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increased standardization</td>
<td></td>
<td>Use for discussion of the GDPR resulting in increased data protection standardization across the EU, in practices within the organization, etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increased complexity</td>
<td></td>
<td>Use for discussion of an increasing complexity in how privacy is protection, data protection requirements, work practices, etc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transparency</td>
<td></td>
<td>Use if the respondent discusses transparency as a significant impact</td>
<td></td>
</tr>
</tbody>
</table>

**Respondent information**  

<table>
<thead>
<tr>
<th>Job description</th>
<th>Use for respondent descriptions of their job or jobs at the archive/organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactions with data producers</td>
<td>Use for general descriptions of respondent interactions with data producers that DO NOT fit under &quot;Feedback and questions...&quot; (e.g., discussions about refining anonymizations, etc.)</td>
</tr>
<tr>
<td>Interactions with data users</td>
<td>Use for general descriptions of respondent interactions with data users that DO NOT fit under &quot;Feedback and questions...&quot;</td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>GDPR</td>
<td>organizational changes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Non-EU data use</td>
<td></td>
</tr>
<tr>
<td>Changes remaining to be completed</td>
<td></td>
</tr>
<tr>
<td>Harmonization</td>
<td></td>
</tr>
<tr>
<td>New documentation requirements</td>
<td></td>
</tr>
<tr>
<td>Changes to data user access</td>
<td></td>
</tr>
<tr>
<td>Data protection measure changes</td>
<td></td>
</tr>
<tr>
<td>Changes in staff access to data</td>
<td></td>
</tr>
<tr>
<td>Changes not directly related to GDPR</td>
<td></td>
</tr>
<tr>
<td>Changes in acquisition</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Changes in usage</td>
<td></td>
</tr>
<tr>
<td>Transition status</td>
<td></td>
</tr>
<tr>
<td>Transition priorities</td>
<td></td>
</tr>
<tr>
<td>Process register</td>
<td></td>
</tr>
<tr>
<td>Staff changes or additions</td>
<td></td>
</tr>
<tr>
<td>Non-EU data or data use</td>
<td>Non-EU data or data use</td>
</tr>
<tr>
<td>Privacy meaning</td>
<td>Privacy meaning</td>
</tr>
<tr>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>Responsibility to data subjects</td>
<td></td>
</tr>
<tr>
<td>Difference between personal and professional view</td>
<td></td>
</tr>
<tr>
<td>Change in perspective due to GDPR</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Respondent perspectives on GDPR</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legal restrictions</td>
</tr>
<tr>
<td></td>
<td>Informed consent as barrier</td>
</tr>
<tr>
<td></td>
<td>Desire for greater control</td>
</tr>
<tr>
<td></td>
<td>Cost</td>
</tr>
<tr>
<td></td>
<td>Barriers to data use</td>
</tr>
<tr>
<td></td>
<td>Effort to find data</td>
</tr>
<tr>
<td></td>
<td>Effort to access</td>
</tr>
<tr>
<td></td>
<td>Depositor permission</td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>Most important impact or component overall</td>
<td></td>
</tr>
<tr>
<td>Biggest challenge overall</td>
<td></td>
</tr>
<tr>
<td>Effectiveness as tool to protect privacy</td>
<td></td>
</tr>
<tr>
<td>Data sharing promotion</td>
<td></td>
</tr>
<tr>
<td>Recommendations to improve</td>
<td></td>
</tr>
<tr>
<td>Does GDPR go too far</td>
<td></td>
</tr>
<tr>
<td>Role of the adequacy decision</td>
<td></td>
</tr>
<tr>
<td>Timeline</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Internally provided</td>
<td></td>
</tr>
<tr>
<td>Externally provided</td>
<td></td>
</tr>
<tr>
<td>Provided to external trainees</td>
<td></td>
</tr>
<tr>
<td>On GDPR</td>
<td></td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Organization specific</td>
<td>Anonymization training</td>
</tr>
<tr>
<td></td>
<td>Data protection training</td>
</tr>
<tr>
<td></td>
<td>Research data management</td>
</tr>
<tr>
<td>Who was involved in transition</td>
<td>IT</td>
</tr>
<tr>
<td></td>
<td>Organizational working group</td>
</tr>
<tr>
<td></td>
<td>Director</td>
</tr>
<tr>
<td></td>
<td>Department heads</td>
</tr>
<tr>
<td></td>
<td>Privacy coordinator</td>
</tr>
<tr>
<td></td>
<td>Internal DPO</td>
</tr>
<tr>
<td></td>
<td>External DPO</td>
</tr>
<tr>
<td>Parent Code</td>
<td>Child Code</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>External parties</td>
<td></td>
</tr>
<tr>
<td>Lawyer or legal advisor</td>
<td></td>
</tr>
<tr>
<td>Archive staff</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix E List of Data Archive Documentation

Table E.1 List of Documentation for the Czech Social Science Data Archive

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Category</th>
<th>Year</th>
<th>Pre or Post GDPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>About the Czech Social Science Data Archive (English)</td>
<td>Archive information</td>
<td>2017</td>
<td>Pre</td>
</tr>
<tr>
<td>About the Czech Social Science Data Archive (English)</td>
<td>Archive information</td>
<td>2020</td>
<td>Post</td>
</tr>
<tr>
<td>Archivační řád Českého sociálněvědního datového archivu (ČSDA) Verze 1.2 (Archiving rules for ČSDA version 1.2)</td>
<td>Archive policy</td>
<td>2016</td>
<td>Pre</td>
</tr>
<tr>
<td>Czech Social Science Data Archive preservation policy Version 1.3 (English)</td>
<td>Archive policy</td>
<td>2016</td>
<td>Pre</td>
</tr>
<tr>
<td>Introduction to the management of social survey data</td>
<td>Archive publication</td>
<td>2014</td>
<td>Pre</td>
</tr>
<tr>
<td>GDPR and the Research Data – introduction into the topic</td>
<td>Archive training</td>
<td>2019</td>
<td>Post</td>
</tr>
<tr>
<td>Agreement on Data Deposition at the Czech Social Science Data Archive, Institute of Sociology AS CR (English)</td>
<td>Deposit agreement</td>
<td>2014</td>
<td>Pre</td>
</tr>
<tr>
<td>Document Name</td>
<td>Category</td>
<td>Year</td>
<td>Pre or Post</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Dohoda o předání datového souboru do Českého sociálněvědního datového archivu Sociologického ústavu AV ČR, v.v.i. (Agreement on Data Deposition)</td>
<td>Deposit agreement</td>
<td>2019</td>
<td>Post</td>
</tr>
<tr>
<td>Dohoda o předání datového souboru do Českého sociálněvědního datového archivu Sociologického ústavu AV ČR, v.v.i. (Agreement on Data Deposition)</td>
<td>Deposit agreement</td>
<td>2021</td>
<td>Post</td>
</tr>
<tr>
<td>Data description form for dataset deposition in Czech Social Science Data Archive (English)</td>
<td>Deposit documentation</td>
<td>2014</td>
<td>Pre</td>
</tr>
<tr>
<td>Formulář pro archivaci datových souborů v Českém sociálněvědním datovém archivu Sociologického ústavu AV ČR, v.v.i. (Data description form for dataset deposition in Czech Social Science Data Archive)</td>
<td>Deposit documentation</td>
<td>2014</td>
<td>Pre</td>
</tr>
<tr>
<td>Komunikace depozitora s archivem (Communication of the depositor with the archive) (English translation available)</td>
<td>Deposit guidance</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Příprava dat k archivaci (Preparing data for archiving) (English translation available)</td>
<td>Deposit guidance</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Příprava dohody o depozici (Preparation of deposit agreement) (English translation available)</td>
<td>Deposit guidance</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Příprava dokumentace dat (Preparation of data documentation) (English translation available)</td>
<td>Deposit documentation</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Assessment Information: CoreTrustSeal Requirements 2017-2019—Czech Social Science Data Archive (CSDA)</td>
<td>Governance and oversight</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Methodology of evaluation of research and professional activity of research-oriented institutes of the Czech Academy of Sciences for the period 2015–2019 (AV ČR) (English)</td>
<td>Governance and oversight</td>
<td>2021</td>
<td>Post</td>
</tr>
<tr>
<td>Statutes and Rules of Procedure of the Scientific Advisory Board of the Czech Social Science Data Archive (CSDA-SAB) (English)</td>
<td>Governance and oversight</td>
<td>2016</td>
<td>Pre</td>
</tr>
<tr>
<td>Informace o zpracování Vašich osobních údajů (Information about the processing of your personal data) (AV ČR)</td>
<td>Privacy policy</td>
<td>2020</td>
<td>Post</td>
</tr>
<tr>
<td>Ochraza osobních údajů (Privacy policy) (SOÚ)</td>
<td>Privacy policy</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Privacy policy (SOÚ) (English)</td>
<td>Privacy policy</td>
<td>2022</td>
<td>Post</td>
</tr>
<tr>
<td>Podmínky přístupu k datům v ČSDA (Conditions of access to data in ČSDA)</td>
<td>Terms of use</td>
<td>2020</td>
<td>Post</td>
</tr>
<tr>
<td>Registrace (Registration – Includes terms of use for data access)</td>
<td>Terms of use</td>
<td>2021</td>
<td>Post</td>
</tr>
<tr>
<td>Document / Regulation</td>
<td>Type</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Appendix 3: Information System Description (Aila)</td>
<td>Archive information</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Appendix 4: Information System Description (Tiipii)</td>
<td>Archive information</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>About us</td>
<td>Archive information</td>
<td>n.d.</td>
<td>Post</td>
</tr>
<tr>
<td>FSD Operational Guidelines</td>
<td>Archive policy</td>
<td>2017</td>
<td>Pre</td>
</tr>
<tr>
<td>FSD Operational Guidelines</td>
<td>Archive policy</td>
<td>n.d.</td>
<td>Post</td>
</tr>
<tr>
<td>Appendix 1: Documents and document series generated by FSD</td>
<td>Archive policy</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>EU's Data Protection Regulation Ended Up Fairly Research-friendly</td>
<td>Archive publication</td>
<td>2016</td>
<td>Pre</td>
</tr>
<tr>
<td>New contract models for archiving research data</td>
<td>Archive publication</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Agreement on the deposition of data to the Finnish Social Science Data Archive</td>
<td>Deposit agreement</td>
<td>2013</td>
<td>Pre</td>
</tr>
<tr>
<td>Deposition Agreement</td>
<td>Deposit agreement</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Agreement on personal data processing for assessing the suitability of research data for archiving</td>
<td>Deposit agreement</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Informing research participants about the processing of their personal data</td>
<td>Deposit guidance</td>
<td>2019</td>
<td>Post</td>
</tr>
<tr>
<td>FSD’s Anonymization plan template</td>
<td>Deposit guidance</td>
<td>2019</td>
<td>Post</td>
</tr>
<tr>
<td>Depositing data through Aila</td>
<td>Deposit guidance</td>
<td>2020</td>
<td>Post</td>
</tr>
<tr>
<td>Guidelines for depositing data</td>
<td>Deposit guidance</td>
<td>n.d.</td>
<td>Post</td>
</tr>
<tr>
<td>Anonymisation and Personal Data</td>
<td>Deposit guidance</td>
<td>n.d.</td>
<td>Post</td>
</tr>
<tr>
<td>Assessment Information: CoreTrustSeal Requirements 2020-2022—Finnish Social Science Data Archive</td>
<td>Governance and oversight</td>
<td>2020</td>
<td>Post</td>
</tr>
<tr>
<td>Administration</td>
<td>Governance and oversight</td>
<td>n.d.</td>
<td>Post</td>
</tr>
<tr>
<td>Appendix 2: Privacy Notice</td>
<td>Privacy policy</td>
<td>2018</td>
<td>Post</td>
</tr>
<tr>
<td>Privacy Policy</td>
<td>Privacy policy</td>
<td>2015</td>
<td>Pre</td>
</tr>
</tbody>
</table>
### Table E.3 List of Documentation for DANS

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Category</th>
<th>Year</th>
<th>Pre or Post</th>
<th>GDPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet van 6 juli 2000 bescherming persoonsgegevens, Nr. 92. (6 July 2000), (Personal Data Protection Act)</td>
<td>Legislation</td>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wet van 9 december 2015, houdende bundeling en aanpassing van regels op het terrein van cultureel erfgoed (Erfgoedwet), Stb. 2015, 511. (Heritage Act)</td>
<td>Legislation</td>
<td>2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home—EASY</td>
<td>Archive information</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>DANS 2021-2025 - Strategy</td>
<td>Archive information</td>
<td>2022</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>About DANS</td>
<td>Archive information</td>
<td>2022</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>DataverseNL</td>
<td>Archive information</td>
<td>n.d.</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Legal information</td>
<td>Archive policy</td>
<td>2017</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Legal information</td>
<td>Archive policy</td>
<td>2020</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Disclaimer EASY</td>
<td>Archive policy</td>
<td>2017</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Disclaimer EASY</td>
<td>Archive policy</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Property rights statement EASY</td>
<td>Archive policy</td>
<td>2017</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Preservation Plan</td>
<td>Archive policy</td>
<td>2018</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Provenance Document: Data Processing Protocol</td>
<td>Archive policy</td>
<td>2021</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>DANS en de Levenswetenschappen: Beleidsnota (DANS and the life sciences: Policy note)</td>
<td>Archive policy</td>
<td>2019</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Memo – Beleid persoonsgegevens datasets EASY (Memorandum – Personal data policy EASY)</td>
<td>Archive policy</td>
<td>2020</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Procedure afsluiten verwerkersovereenkomst (handmatig) (Procedure for entering into a processor agreement (manual))</td>
<td>Archive policy</td>
<td>n.d.</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Procedure omgaan met rechten betrokkenen DANS-KNAW 2 dec 2018 (Procedure dealing with rights of data subjects DANS- KNAW 2 Dec 2018) (KNAW)</td>
<td>Archive policy</td>
<td>2018</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Type</td>
<td>Year</td>
<td>Status</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Toelichting standaardovereenkomsten duurzame archivering (Explanation standard agreements sustainable archiving)</td>
<td>Archive policy</td>
<td>2020</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Jaarverslag 2007 Data Archiving and Networked Services</td>
<td>Archive publication</td>
<td>2008</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Reactie van DANS t.b.v. De internetconsultatie met betrekking tot de Uitvoeringswet Algemene Verordening Gegevensbescherming (AVG) (Reaction of DANS with regard to the internet consultation on the Dutch General Data Protection Regulation (AVG))</td>
<td>Archive publication</td>
<td>2017</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Evaluation of DANS EASY Repository Based on the FAIR Principles</td>
<td>Archive publication</td>
<td>2018</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Archiving and managing research data: Data services to the domains of the humanities and social sciences and beyond: DANS in the Netherlands</td>
<td>Archive publication</td>
<td>2020</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>NARCIS update: February 2022</td>
<td>Archive publication</td>
<td>2022</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>GDPR for researchers: Making your data management GDPR proof</td>
<td>Archive training</td>
<td>2019</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>De AVG en evenementen organiseren met een online dienst (The GDPR and organizing events with an online service)</td>
<td>Archive training</td>
<td>2020</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>License agreement</td>
<td>Deposit agreement</td>
<td>2016</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Deposit agreement</td>
<td>Deposit agreement</td>
<td>2019</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Processing agreement sustained archiving</td>
<td>Deposit agreement</td>
<td>2019</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>EASY Deposit Form</td>
<td>Deposit agreement</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>DANS template for a data management plan</td>
<td>Deposit guidance</td>
<td>2015</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Depositing Data Manual - Before depositing</td>
<td>Deposit guidance</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Depositing Data Manual - During depositing</td>
<td>Deposit guidance</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Depositing Data Manual - After depositing</td>
<td>Deposit guidance</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Rate information</td>
<td>Deposit guidance</td>
<td>2021</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>NWO-DANS data contract</td>
<td>External agreement</td>
<td>2015</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Samenwerkingsovereenkomst inzake DataverseNL (Collaboration agreement DataverseNL)</td>
<td>External agreement</td>
<td>2016</td>
<td>Pre</td>
<td></td>
</tr>
<tr>
<td>Samenwerkingsovereenkomst inzake DataverseNL (Collaboration agreement DataverseNL)</td>
<td>External agreement</td>
<td>2022</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Verwerkingsovereenkomst DataverseNL</td>
<td>External agreement</td>
<td>2022</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Nederlandse gedragscode wetenschappelijke integriteit (Netherlands Code of Conduct for Research Integrity) (KNAW)</td>
<td>Governance and oversight</td>
<td>2018</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Document Name</td>
<td>Category</td>
<td>Year</td>
<td>Pre or Post GDPR</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>--------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td><em>Grundgesetz der Bundesrepublik Deutschland.</em> (1949, last amended 29.9.2020). (Basic Law)</td>
<td>Legislation</td>
<td>1949</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Gesetz zur Anpassung des allgemeinen Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Nordrhein-Westfälisches Datenschutz-Anpassungs- und Umsetzungsgesetz EU (Datenschutzgesetz)</td>
<td>Legislation</td>
<td>2018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Legal References

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Date</th>
</tr>
</thead>
</table>

### Archive and Information

<table>
<thead>
<tr>
<th>Topic</th>
<th>Type</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facts 2014 pamphlet</td>
<td>Archive</td>
<td>2014</td>
</tr>
<tr>
<td>Secure Data Center</td>
<td>Archive</td>
<td>2017</td>
</tr>
<tr>
<td>GESIS - Leibniz-Institute for the Social Sciences</td>
<td>Archive</td>
<td>2017</td>
</tr>
<tr>
<td>Computational Social Science</td>
<td>Archive</td>
<td>2018</td>
</tr>
<tr>
<td>Secure Data Center</td>
<td>Archive</td>
<td>2022</td>
</tr>
<tr>
<td>SowiDataNet</td>
<td>atorium</td>
<td>Archive</td>
</tr>
<tr>
<td>New data—New opportunities</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>Survey Data Curation (SDC)</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>Drawing of samples</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>Research Data Centers</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>German Microdata Lab</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>GESIS Leibniz-Institute for the Social Sciences</td>
<td>Archive</td>
<td>n.d.</td>
</tr>
<tr>
<td>Legal Imprint</td>
<td>Archive policy</td>
<td>2010</td>
</tr>
<tr>
<td>Legal Imprint</td>
<td>Archive policy</td>
<td>2015</td>
</tr>
<tr>
<td>Was wir sammeln (What we collect)</td>
<td>Archive policy</td>
<td>2013</td>
</tr>
<tr>
<td>Licence compliance policy</td>
<td>Archive Policy</td>
<td>2021</td>
</tr>
<tr>
<td>Overview of ingest check levels</td>
<td>Archive policy</td>
<td>2022</td>
</tr>
<tr>
<td>Conceptualizing a spatial data infrastructure for the social sciences: An example from Germany</td>
<td>Archive publication</td>
<td>2016</td>
</tr>
<tr>
<td>Using georeferenced data in social science survey research: The method of spatial linking and its application with the German General Social Survey and the GESIS Panel</td>
<td>Archive publication</td>
<td>2019</td>
</tr>
<tr>
<td>Access data from the GESIS data catalogue hacked</td>
<td>Archive publication</td>
<td>2019</td>
</tr>
<tr>
<td>Title</td>
<td>Type</td>
<td>Year</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships</td>
<td>Archive publication</td>
<td>2020</td>
</tr>
<tr>
<td>Integrating survey data and digital trace data: Key issues in developing an emerging field</td>
<td>Archive publication</td>
<td>2020</td>
</tr>
<tr>
<td>An infrastructure for spatial linking of survey data</td>
<td>Archive publication</td>
<td>2020</td>
</tr>
<tr>
<td>Informed consent for linking survey and social media data—Differences between platforms and data types</td>
<td>Archive publication</td>
<td>2021</td>
</tr>
<tr>
<td>The missing link? Linking surveys with geospatial and digital behavioral data</td>
<td>Archive publication</td>
<td>2021</td>
</tr>
<tr>
<td>Measuring the time spent on data curation</td>
<td>Archive publication</td>
<td>2022</td>
</tr>
<tr>
<td>GDPR and research one year on – Experiences across Europe</td>
<td>Archive training</td>
<td>2019</td>
</tr>
<tr>
<td>Archive agreement</td>
<td>Deposit agreement</td>
<td>2014</td>
</tr>
<tr>
<td>Archiving PLUS service contract (Version 12)</td>
<td>Deposit agreement</td>
<td>2021</td>
</tr>
<tr>
<td>Archiving PREMIUM service contract (Version 12)</td>
<td>Deposit agreement</td>
<td>2021</td>
</tr>
<tr>
<td>SowiDataNet price list</td>
<td>Deposit guidance</td>
<td>2019</td>
</tr>
<tr>
<td>SowiDataNet</td>
<td>atorium: Submitting and describing research data (1.1)</td>
<td>Deposit guidance</td>
</tr>
<tr>
<td>Archiving PREMIUM/PLUS</td>
<td>Deposit guidance</td>
<td>n.d.</td>
</tr>
<tr>
<td>How-to guide data sharing</td>
<td>Deposit guidance</td>
<td>n.d.</td>
</tr>
<tr>
<td>Preparation of data transfer (Pre-Ingest)</td>
<td>Deposit guidance</td>
<td>n.d.</td>
</tr>
<tr>
<td>Agreement on procedures</td>
<td>Verfahrensvorberaung (Datenzugang: Gastaufenthalt (FDZ) (Data Access: Guest Stay) (FDZ-IAB)</td>
<td>External agreement</td>
</tr>
<tr>
<td>Geschäftsordnung des Kuratoriums: Beschluss auf der Sitzung des Kuratoriums am 05. Dezember 2014 in Mannheim (Rules of Procedure of the Board of Trustees: Adopted at the meeting of the Board of Trustees in Mannheim on December 05, 2014)</td>
<td>Governance and oversight</td>
<td>2014</td>
</tr>
<tr>
<td>Rules of Procedure: Scientific Advisory Board (Kuratorium)</td>
<td>Governance and oversight</td>
<td>2016</td>
</tr>
<tr>
<td>Assessment Information: CoreTrustSeal Requirements 2017-2019—GESIS Data Archive for the Social Sciences</td>
<td>Governance and oversight</td>
<td>2017</td>
</tr>
<tr>
<td>The Leibniz Association Senate Evaluation Procedure—Basic Principles (Leibniz Association)</td>
<td>Governance and oversight</td>
<td>2018</td>
</tr>
<tr>
<td>Geschäftsordnung des Nutzerberirats von GESIS gemäß Beschluss des GESIS-Kuratorium vom 25.06.2021 (Rules of procedure of the GESIS User Advisory Board)</td>
<td>Governance and oversight</td>
<td>2021</td>
</tr>
<tr>
<td>Title</td>
<td>Document Type</td>
<td>Year</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>------</td>
</tr>
<tr>
<td>Data protection at GESIS</td>
<td>Privacy policy</td>
<td>2018</td>
</tr>
<tr>
<td>Data protection at GESIS</td>
<td>Privacy policy</td>
<td>2021</td>
</tr>
<tr>
<td>Data protection information for users and interested persons of the GESIS Leibniz Institute for the Social Sciences</td>
<td>Privacy policy</td>
<td>2018</td>
</tr>
<tr>
<td>Usage regulations (Data Archive)</td>
<td>Terms of use</td>
<td>2007</td>
</tr>
<tr>
<td>Usage regulations (Data Archive)</td>
<td>Terms of use</td>
<td>2018</td>
</tr>
<tr>
<td>Data Use Agreement (SDC)</td>
<td>Terms of use</td>
<td>2015</td>
</tr>
<tr>
<td>Data Use Agreement (Off-site)</td>
<td>Terms of use</td>
<td>2018</td>
</tr>
<tr>
<td>Data Use Agreement (Off-site)</td>
<td>Terms of use</td>
<td>2018</td>
</tr>
<tr>
<td>Guidelines for the Secure Data Center Safe Room</td>
<td>Terms of use</td>
<td>2016</td>
</tr>
<tr>
<td>Nutzungsbedingungen (SowiDataNet</td>
<td>datorium)</td>
<td>Terms of use</td>
</tr>
<tr>
<td>Terms of Use (SowiDataNet</td>
<td>datorium)</td>
<td>Terms of use</td>
</tr>
<tr>
<td>Leaflet on secure handling of research data provided with a GESIS data use agreement</td>
<td>User guidance</td>
<td>2014</td>
</tr>
<tr>
<td>Charges – GESIS Dept. Data Archive and Data Analysis</td>
<td>User guidance</td>
<td>2016</td>
</tr>
<tr>
<td>Access</td>
<td>User guidance</td>
<td>2017</td>
</tr>
<tr>
<td>Registration 2.0—ZACAT, DBK and datorium</td>
<td>User guidance</td>
<td>2015</td>
</tr>
<tr>
<td>Register</td>
<td>User guidance</td>
<td>n.d.</td>
</tr>
<tr>
<td>Preisliste für die manuelle Datenbereitstellung im Datenarchiv für Sozialwissenschaften (DAS) des GESIS – Leibniz-Institut für Sozialwissenschaften (Price list for manual data provision in the Data Archive for the Social Sciences (DAS) of GESIS - Leibniz Institute for the Social Sciences)</td>
<td>User guidance</td>
<td>2019</td>
</tr>
<tr>
<td>Data linking – Linking survey data with geospatial, social media, and sensor data Version 1.0</td>
<td>User guidance</td>
<td>2021</td>
</tr>
</tbody>
</table>
## Appendix F Eurobarometer 83.1 Data Tables


### Table F.1 National responses for European Union and field site locations to: QB4 (Eurobarometer 83.1): How much control do you feel you have over the information you provide online?

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Responses</th>
<th>European Union n=17484</th>
<th>Czechia n=715</th>
<th>Finland n=724</th>
<th>Germany n=1028</th>
<th>The Netherlands n=915</th>
</tr>
</thead>
<tbody>
<tr>
<td>QB4: How much control do you feel you have over the information you provide online, e.g. the ability to correct, change or delete this information?</td>
<td>Complete control</td>
<td>2974 (17.01%)</td>
<td>119 (16.64%)</td>
<td>106 (14.64%)</td>
<td>51 (4.96%)</td>
<td>80 (8.74%)</td>
</tr>
<tr>
<td></td>
<td>Partial control</td>
<td>9118 (52.15%)</td>
<td>392 (54.83%)</td>
<td>481 (66.44%)</td>
<td>410 (39.88%)</td>
<td>536 (58.58%)</td>
</tr>
<tr>
<td></td>
<td>No control at all</td>
<td>4987 (28.52%)</td>
<td>194 (27.13%)</td>
<td>128 (17.68%)</td>
<td>472 (45.91%)</td>
<td>294 (32.13%)</td>
</tr>
<tr>
<td></td>
<td>It depends on the website or application</td>
<td>405 (2.32%)</td>
<td>10 (1.40%)</td>
<td>9 (1.24%)</td>
<td>95 (9.27%)</td>
<td>5 (0.55%)</td>
</tr>
</tbody>
</table>

### Table F.2 National responses for European Union and field site locations to: QB5 (Eurobarometer 83.1): How concerned are you about not having complete control over the information you provide online?

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Responses</th>
<th>European Union n=14406</th>
<th>Czechia n=588</th>
<th>Finland n=615</th>
<th>Germany n=970</th>
<th>The Netherlands n=835</th>
</tr>
</thead>
<tbody>
<tr>
<td>QB5: How concerned are you about not having complete control over the information you provide online? Would you say you are…?</td>
<td>Very concerned</td>
<td>2442 (16.95%)</td>
<td>158 (26.87%)</td>
<td>53 (8.62%)</td>
<td>198 (20.41%)</td>
<td>76 (9.10%)</td>
</tr>
<tr>
<td></td>
<td>Fairly concerned</td>
<td>6940 (48.17%)</td>
<td>297 (50.51%)</td>
<td>336 (54.63%)</td>
<td>470 (48.45%)</td>
<td>337 (40.36%)</td>
</tr>
<tr>
<td></td>
<td>Not very concerned</td>
<td>4361 (30.27%)</td>
<td>123 (20.92%)</td>
<td>196 (31.87%)</td>
<td>293 (30.21%)</td>
<td>362 (43.35%)</td>
</tr>
<tr>
<td>Question</td>
<td>Possible Responses</td>
<td>European Union n=27980</td>
<td>Czechia n=1066</td>
<td>Finland n=1016</td>
<td>Germany n=1594</td>
<td>The Netherlands n=1008</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>QB17: Should your explicit approval be required before any kind of personal information is collected and processed?</td>
<td>Yes, in all cases</td>
<td>19645 (70.21%)</td>
<td>776 (73.08%)</td>
<td>670 (65.94%)</td>
<td>1060 (66.50%)</td>
<td>587 (58.23%)</td>
</tr>
<tr>
<td></td>
<td>Yes, in the case of personal information required online</td>
<td>4322 (15.45%)</td>
<td>146 (13.70%)</td>
<td>191 (18.80%)</td>
<td>309 (19.38%)</td>
<td>259 (25.69%)</td>
</tr>
<tr>
<td></td>
<td>Yes, in the case of sensitive information whether online or offline (e.g. health, religion, political beliefs, sexual preferences, etc.)</td>
<td>3442 (12.30%)</td>
<td>158 (14.82%)</td>
<td>113 (11.12%)</td>
<td>312 (19.57%)</td>
<td>175 (17.36%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1419 (5.078%)</td>
<td>27 (2.53%)</td>
<td>48 (4.72%)</td>
<td>66 (4.14%)</td>
<td>50 (4.96%)</td>
</tr>
</tbody>
</table>

Table F.3 National responses for European Union and field site locations to: QB17 (Eurobarometer 83.1): Should your explicit approval be required before any kind of personal information is collected and processed?
Appendix G Special Eurobarometer 487 Data Tables


Table G.1 National responses for European Union and field site locations to: QB17 (Special Eurobarometer 487): Have you heard of the General Data Protection Regulation, which came into force in 2018? and to: QB18(1-6) Special Eurobarometer 487: The General Data Protection Regulation guarantees a number of rights. Have you heard of each of the following rights?

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Responses</th>
<th>European Union (incl. UK) n=27524</th>
<th>Czechia n=1068</th>
<th>Finland n=1000</th>
<th>Germany n=1507</th>
<th>The Netherlands n=1017</th>
</tr>
</thead>
<tbody>
<tr>
<td>QB17: Have you heard of the General Data Protection Regulation (GDPR), which came into force in 2018?</td>
<td>Yes and you know what it is</td>
<td>9867 (35.85%)</td>
<td>529 (49.53%)</td>
<td>349 (34.90%)</td>
<td>627 (41.61%)</td>
<td>614 (60.37%)</td>
</tr>
<tr>
<td></td>
<td>Yes but you don’t know what exactly it is</td>
<td>8600 (31.25%)</td>
<td>373 (34.93%)</td>
<td>313 (31.30%)</td>
<td>560 (37.16%)</td>
<td>277 (27.24%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>8784 (31.91%)</td>
<td>164 (15.36%)</td>
<td>334 (33.40%)</td>
<td>303 (20.11%)</td>
<td>126 (12.39%)</td>
</tr>
<tr>
<td></td>
<td>Don’t know</td>
<td>273 (0.99%)</td>
<td>2 (0.19%)</td>
<td>4 (0.40%)</td>
<td>16 (1.06%)</td>
<td>0 (-)</td>
</tr>
<tr>
<td>QB18.1: The General Data Protection Regulation (GDPR) guarantees a number of rights. Have you heard of each of the following rights? The right to access your data</td>
<td>Yes and you have exercised it</td>
<td>4836 (17.57%)</td>
<td>120 (11.24%)</td>
<td>270 (27.00%)</td>
<td>200 (13.27%)</td>
<td>261 (25.66%)</td>
</tr>
<tr>
<td></td>
<td>Yes but you have not exercised it</td>
<td>13074 (47.50%)</td>
<td>614 (57.49%)</td>
<td>485 (48.50%)</td>
<td>892 (59.19%)</td>
<td>615 (60.47%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>8455 (30.72%)</td>
<td>275 (25.75%)</td>
<td>211 (21.10%)</td>
<td>373 (24.75%)</td>
<td>128 (12.59%)</td>
</tr>
<tr>
<td></td>
<td>Don’t know</td>
<td>1158 (4.21%)</td>
<td>59 (5.52%)</td>
<td>33 (3.30%)</td>
<td>42 (2.79%)</td>
<td>13 (1.28%)</td>
</tr>
<tr>
<td>QB18.2: ... Have you heard of each of the following</td>
<td>Yes and you have exercised it</td>
<td>6474 (23.52%)</td>
<td>123 (11.99%)</td>
<td>393 (39.30%)</td>
<td>440 (29.20%)</td>
<td>425 (41.79%)</td>
</tr>
<tr>
<td>Question</td>
<td>Yes but you have not exercised it</td>
<td>Yes and you have exercised it</td>
<td>Yes but you have not exercised it</td>
<td>No</td>
<td>Don’t know</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
<td>------------------------------</td>
<td>----------------------------------</td>
<td>----</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td><strong>QB18.3:</strong> Have you heard of each of the following rights? The right to correct your data if it is wrong</td>
<td>9756 (35.45%)</td>
<td>460 (43.07%)</td>
<td>404 (40.40%)</td>
<td>598 (39.68%)</td>
<td>395 (38.84%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>10044 (36.49%)</td>
<td>413 (38.67%)</td>
<td>178 (17.80%)</td>
<td>419 (27.80%)</td>
<td>187 (18.39%)</td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>1251 (4.55%)</td>
<td>72 (6.74%)</td>
<td>24 (2.40%)</td>
<td>51 (3.38%)</td>
<td>11 (1.08%)</td>
<td></td>
</tr>
<tr>
<td><strong>QB18.4:</strong> Have you heard of each of the following rights? The right to have your data deleted and to be forgotten</td>
<td>12294 (44.67%)</td>
<td>573 (53.65%)</td>
<td>559 (55.90%)</td>
<td>811 (53.81%)</td>
<td>561 (55.16%)</td>
<td></td>
</tr>
<tr>
<td>Yes and you have exercised it</td>
<td>4481 (16.28%)</td>
<td>83 (7.77%)</td>
<td>196 (19.60%)</td>
<td>201 (13.34%)</td>
<td>253 (24.88%)</td>
<td></td>
</tr>
<tr>
<td>Yes but you have not exercised it</td>
<td>12235 (44.45%)</td>
<td>530 (49.63%)</td>
<td>444 (44.40%)</td>
<td>832 (55.21%)</td>
<td>579 (56.93%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3690 (13.41%)</td>
<td>75 (7.02%)</td>
<td>129 (12.90%)</td>
<td>206 (13.67%)</td>
<td>202 (19.87%)</td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>1207 (4.39%)</td>
<td>63 (5.90%)</td>
<td>31 (3.10%)</td>
<td>50 (3.32%)</td>
<td>18 (1.77%)</td>
<td></td>
</tr>
<tr>
<td><strong>QB18.5:</strong> Have you heard of each of the following rights? The right to have a say when decisions are automated</td>
<td>9035 (32.83%)</td>
<td>356 (33.33%)</td>
<td>337 (33.70%)</td>
<td>573 (38.02%)</td>
<td>359 (35.30%)</td>
<td></td>
</tr>
<tr>
<td>Yes and you have exercised it</td>
<td>2168 (7.88%)</td>
<td>47 (4.40%)</td>
<td>76 (7.60%)</td>
<td>75 (4.98%)</td>
<td>75 (7.37%)</td>
<td></td>
</tr>
<tr>
<td>Yes but you have not exercised it</td>
<td>14647 (53.22%)</td>
<td>576 (53.93%)</td>
<td>539 (53.90%)</td>
<td>786 (52.36%)</td>
<td>542 (53.29%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9035 (32.83%)</td>
<td>356 (33.33%)</td>
<td>337 (33.70%)</td>
<td>573 (38.02%)</td>
<td>359 (35.30%)</td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>1268 (4.61%)</td>
<td>62 (5.80%)</td>
<td>31 (3.10%)</td>
<td>46 (3.32%)</td>
<td>14 (1.38%)</td>
<td></td>
</tr>
<tr>
<td><strong>QB18.6:</strong> Have you heard of each of the following rights? The right to move your data from one provider to another</td>
<td>12324 (44.78%)</td>
<td>509 (47.66%)</td>
<td>422 (42.0%)</td>
<td>590 (39.15%)</td>
<td>382 (37.56%)</td>
<td></td>
</tr>
<tr>
<td>Yes and you have exercised it</td>
<td>3661 (13.30%)</td>
<td>65 (6.09%)</td>
<td>157 (15.7%)</td>
<td>223 (14.80%)</td>
<td>184 (18.09%)</td>
<td></td>
</tr>
<tr>
<td>Yes but you have not exercised it</td>
<td>10146 (36.86%)</td>
<td>422 (39.51%)</td>
<td>378 (37.80%)</td>
<td>639 (42.40%)</td>
<td>421 (41.40%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>12324 (44.78%)</td>
<td>509 (47.66%)</td>
<td>422 (42.0%)</td>
<td>590 (39.15%)</td>
<td>382 (37.56%)</td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>1393 (5.06%)</td>
<td>71 (6.65%)</td>
<td>42 (4.20%)</td>
<td>55 (3.65%)</td>
<td>30 (2.95%)</td>
<td></td>
</tr>
</tbody>
</table>
Table G.2 National responses for European Union and field site locations to: QB9 (Special Eurobarometer 487): How much control do you feel you have over the information you provide online?

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Responses</th>
<th>European Union (incl. UK) n=18975</th>
<th>Czechia n=799</th>
<th>Finland n=787</th>
<th>Germany n=1129</th>
<th>The Netherlands n=956</th>
</tr>
</thead>
<tbody>
<tr>
<td>QB9: How much control do you feel you have over the information you provide online, e.g. the ability to correct, change or delete this information?</td>
<td>Complete control</td>
<td>2681 (14.13%)</td>
<td>131 (16.39%)</td>
<td>126 (16.01%)</td>
<td>64 (5.67%)</td>
<td>92 (9.62%)</td>
</tr>
<tr>
<td></td>
<td>Partial control</td>
<td>9563 (50.40%)</td>
<td>412 (51.56%)</td>
<td>516 (65.56%)</td>
<td>507 (44.91%)</td>
<td>594 (62.13%)</td>
</tr>
<tr>
<td></td>
<td>No control at all</td>
<td>5635 (29.70%)</td>
<td>208 (26.03%)</td>
<td>123 (15.63%)</td>
<td>456 (40.39%)</td>
<td>255 (26.67%)</td>
</tr>
<tr>
<td></td>
<td>It depends on the website or application</td>
<td>627 (3.30%)</td>
<td>82 (10.26%)</td>
<td>4 (0.51%)</td>
<td>77 (6.82%)</td>
<td>7 (0.73%)</td>
</tr>
<tr>
<td></td>
<td>Don’t know</td>
<td>379 (1.99%)</td>
<td>7 (0.88%)</td>
<td>17 (2.16%)</td>
<td>25 (2.21%)</td>
<td>7 (0.73%)</td>
</tr>
</tbody>
</table>

Table G.3 National responses for European Union and field site locations to: QB10 (Special Eurobarometer 487): How concerned are you about not having complete control over the information you provide online?

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Responses</th>
<th>European Union (incl. UK) n=15915</th>
<th>Czechia n=639</th>
<th>Finland n=644</th>
<th>Germany n=1040</th>
<th>The Netherlands n=857</th>
</tr>
</thead>
<tbody>
<tr>
<td>QB10: How concerned are you about not having complete control over the information you provide online? Would you say you are…?</td>
<td>Very concerned</td>
<td>2616 (16.44%)</td>
<td>96 (15.02%)</td>
<td>89 (13.82%)</td>
<td>175 (16.83%)</td>
<td>89 (10.39%)</td>
</tr>
<tr>
<td></td>
<td>Fairly concerned</td>
<td>7340 (46.12%)</td>
<td>259 (40.53%)</td>
<td>308 (47.83%)</td>
<td>442 (42.50%)</td>
<td>321 (37.46%)</td>
</tr>
<tr>
<td></td>
<td>Not very concerned</td>
<td>4919 (30.91%)</td>
<td>198 (30.99%)</td>
<td>221 (34.32%)</td>
<td>354 (34.04%)</td>
<td>380 (44.34%)</td>
</tr>
<tr>
<td></td>
<td>Not at all concerned</td>
<td>920 (5.78%)</td>
<td>77 (12.05%)</td>
<td>26 (4.04%)</td>
<td>59 (5.67%)</td>
<td>66 (7.70%)</td>
</tr>
<tr>
<td></td>
<td>Don’t know</td>
<td>120 (0.75%)</td>
<td>10 (1.56%)</td>
<td>0 (-)</td>
<td>10 (0.96%)</td>
<td>1 (0.12%)</td>
</tr>
</tbody>
</table>
Bibliography


Bernard, T. S., Hsu, T., Perlroth, N., & Lieber, R. (2017, September 7). Equifax says cyberattack may have affected 143 million in the U.S. *New York Times (Online)*.


Consortium of European Social Science Data Archives. (n.d.a). *History—CESSDA*. CESSDA. https://www.cessda.eu/About/History


DANS. (2015a). *NWO-DANS Data Contracts*. DANS.
DANS. (2015b). *General conditions of use* (5.3 UK).


DANS. (2019b, March 26). *Registration—EASY*. EASY.


DANS. (2019d). *Processing agreement sustained archiving* (UK 26-09-2019 1.2). DANS.


https://easy.dans.knaw.nl/ui/home


DANS. (2021c, September 15). *Deposit your data*. EASY. https://easy.dans.knaw.nl/


DANS. (2021g, December 10). *Centre of expertise & repository for research data*. DANS.
https://dans.knaw.nl/en/


DANS. (2022b). *Samenwerkingsovereenkomst inzake DataverseNL* (Standaard - 15 februari 2022). DANS.


DataverseNL. (n.d.). *DataverseNL*. DataverseNL. https://dataverse.nl/


Elgesem, D. (1999). The structure of rights in directive 95/46/EC on the protection of individuals with regard to the processing of personal data and the free movement of such data. *Ethics and Information Technology, 1*, 283–293.

Espindola, J. (2011). *The disclosure of respect: The public exposure of Stasi informers after the German reunification* [Dissertation, University of Michigan]. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/84520/espimata_1.pdf?sequence=1&isAllowed=y


European Commission. (2012). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data* (General Data...


GESIS Leibniz Institut für Sozialwissenschaften. (2015a). Registration 2.0—ZACAT, DBK, and datorium. DBK.


GESIS Leibniz Institut für Sozialwissenschaften. (2018e). Data protection information for users and interested persons of the GESIS Leibniz Institute for the Social Sciences. GESIS Leibniz Institut für Sozialwissenschaften.

https://data.gesis.org/sharing/docs/SowiDataNet_Prices_english.pdf

https://www.gesis.org/fileadmin/upload/Datenservices/Preisliste/Preisliste_DSS.PDF

495


Hajduk, G. K., Jamieson, N. E., Baker, B. L., Olesen, O. F., & Lang, T. (2019). It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too! *BMJ Global Health, 4*(4), e001550. https://doi.org/10.1136/bmjgh-2019-001550


Hestermeyer, H. P. (2014). The implementation of European Union Law in Germany. In E. de Wet, H. P. Hestermeyer, & R. Wolfrum (Eds.), The implementation of international law in Germany and South Africa (pp. 444–467). Pretoria University Law Press.


Inter-university Consortium for Political and Social Research. (2019). About ICPSR. ICPSR. https://www.icpsr.umich.edu/icpsrweb/content/about/index.html


Koninklijke Nederlandse Akademie van Wetenschappen, Nederlandse Federatie van Universitair Medische Centra, Nederlandse Organisatie voor Wetenschappelijk Onderzoek, TO2-Federatie, Vereniging Hogescholen, & Vereniging van Universiteiten. (2018). *Nederlandse gedragscode wetenschappelijke integriteit (Netherlands Code of Conduct*


Kraaikamp, E. (2020b). *De AVG en evenementen organiseren met een online dienst*. DANS.


Ministerstvo školství, mládeže a tělovýchovy České republiky. (2010). *Roadmap for Large Research, Development and Innovation Infrastructures in the Czech Republic*. Ministerstvo školství, mládeže a tělovýchovy České republiky. https://www.vyzkumnneinfrastruktury.cz/wp-content/uploads/2019/06/Cestovn%C3%AD-mapavelk%C3%BDch-v%C3%BDzkumn%C3%BDch-infrastruktur%C4%8CR_2010_EN.pdf


509
http://www.monitor.upeace.org/archive.cfm?id_article=707

https://doi.org/10.1109/ACCESS.2016.2535120

https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf

https://nerc.ukri.org/research/sites/data/


510


Poole, M. S., & van de Ven, A. H. (2004). Theories of organizational change and innovation processes. In M. S. Poole & A. H. van de Ven (Eds.), Handbook of Organizational Change and Innovation. Oxford University Press.


Prinz, C. (2022a). *AW: How many datasets in our Archive? (Also by A,B,C)?*


https://doi.org/10.1142/S0218488502001648

https://doi.org/10.2478/mmcks-2020-0005

Tartwijk, M. van. (2014, February 11). Dutch Minister of Interior Fights for His Political Life; Faces Possible No-Confidence Vote He Misled Public. *Wall Street Journal (Online).*


http://uk.practicallaw.thomsonreuters.com/6-382-4256?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1


University of Tokyo. (n.d.). *What is SSJDA?* Center for Social Research and Data Archives. https://csrda.iss.u-tokyo.ac.jp/en/ssjda/about/


