Perspectives, roles, and knowledge transfer amongst stakeholders of research data sharing

Live Håndlykken Kvale

Dissertation for the degree of Philosophiae Doctor (PhD)
Department of Archivistic, Library and Information Science
Faculty of Social Sciences
OsloMet - Oslo Metropolitan University

Spring 2022

OsloMet Avhandling 2022 nr 15

ISSN 2535-471X (trykt)
ISSN 2535-5414 (online)

ISBN 978-82-8364-396-1 (trykt)
ISBN 978-82-8364-413-5 (online)

OsloMet – storbyuniversitetet
Universitetsbiblioteket
Skriftserien
St. Olavs plass 4,
0130 Oslo,
Telefon (47) 64 84 90 00

Postadresse:
Postboks 4, St. Olavs plass
0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

Databases and data infrastructures do not simply support research, they fundamentally change the practice and organization of research – the questions asked, how they are asked, how they are answered, how the answers are deployed, who is conducting the research and how they operate as researchers. (Kitchin, 2014, p. 24)

Acknowledgements

The core of this thesis is the empirical material, the data and the people behind the data, the expert participants sharing their experience, ideas, and insight. I want to thank each one of you for sharing your time, and for all the valuable conversations and feedback trough the project.

Special thanks to my supervisor, Professor Nils Pharo and co-supervisor Associate professor Peter Darch for your expertise, experience, patience, and support, which helped me complete this work. Through the years I have remined myself several times that "at least Nils believes I will finish". Thank you for always making sure I stayed (more or less) on track, for saving something for later, and for always being available. Peter has played the role of target and test audience, "if Peter finds this interesting, that must mean I am on to something". You have also taught me much about the art of writing, expanding and condensing text, and how much I have yet to learn about the English language.

A PhD possition in library and information studies with scholarly communication and open science as focus, was a perfect match with my interest in research data management. I am most grateful towards OsloMet and Archival- Library and information science for granting me the possibility to undertake this PhD-project. At OsloMet I have had the pleasure of being a PhD-candidate together with fellow PhD students including Madeleine, Tine, Karin, Kamilla, Marte, Natalie, Jenny, and others whom I want to thank for conversations, inspiration, and motivation throughout the years. I would like to express gratitude the University of Illinois at Urbana- Champaign, School of information science for hosting me as a visiting scholar in the spring of 2018. The people who made this stay possible include Linda Smith, Vetle Torvik, Karen Ruhleder, and Peter whom I was fortunate to recruit as a co-supervisor. Being a visiting doctoral student at the University of Illinois was possible due to funding from OsloMet and Norway-America Association (NORAM). I would also like to acknowledge the invisible yet crucial feedback I have received from reviewers. Your suggestions and feedback have helped shape this thesis and motivated me to see it through to its conclusion. Inger Beate Nylund, Katriina Bystrøm, and Jutta Haider all commented on the three seminars along the way, each of your contributions have been instrumental in realising the finished work.

Special thanks to the University of Oslo, University library for providing the opportunity to combine two positions and approaches of theory and practice and for encouraging me to do so. Thanks, to Live Rasmussen, Bente Andreassen, and Randi Rønningen. At the University of Oslo, I have also had the best PhD-life support team one could hope for; Karoline Moe who shared her knowledge on PhD-life and everything else and Dan Michael Heggø who assisted with scripting and reviewing of xml and data with patience in correcting my many dyslectic typos together with David Pughe, Margaret Fotland, and Elin Stangeland, with whom I can talk about research data management with 24/7. Thanks also to all my other colleagues at the University of Oslo who have provided a supportive and collegial network throughout the PhD process.

Finally, I wish to acknowledge and show appreciation for my family. My husband Angelo for his knowledge on sociological theory in addition to patience and encouragement. Askanio for giving me perspective on the life - work balance, providing the distraction needed to focus when at work, and putting the computer away at home. My parents and in-laws have been commensurably supportive, for which I am deeply grateful.

Oslo, April 2022 **Live Håndlykken Kvale**

Abstract

Data sharing requires collaboration on infrastructure and a knowledge exchange amongst stakeholders while these develop solutions aiming at increasing the quality of research through data curation. Divergent perspectives amongst key stakeholders on how and why data sharing is to take place create frictions in the collaboration and development of infrastructure. In addition, new roles are emerging to facilitate data management and data curation. The professional identity of these struggle with combining and maintaining high-level expertise in multiple domains. The problem statement approaches these frictions by asking: How do the perspectives and expertise of key stakeholders involved in research data sharing affect the collaboration and knowledge transfer amongst these? By addressing this question, the thesis aims at understanding the relation between the perspectives and expertise of key stakeholders involved in research data sharing and how these perspectives affect the collaboration and knowledge transfer amongst the stakeholders.

Designed as a three-phased, modified Delphi study with data collection during a 14-month period, the study captured parts of the development of infrastructure for research data sharing. The participant group consisted of 24 expert shareholders disseminating their experiences and perspectives on the sharing and curation of research data through two interviews and one survey. The results are based on the final analysis of all the data material.

This is a thesis involving compilation of three articles and a narrative (Norwegian ‘kappe’). Each of the three articles addresses specific issues within research data sharing, namely by: 1) Exploring the different data steward roles; 2) Analysing the multiple perspectives on data management plans as a facilitator for data curation and sharing; and 3) How personal privacy can be balanced with high quality research through the research data lifecycle. The narrative lifts the perspective by addressing challenges that connect the three articles as a thread; divergent perspectives, roles, expertise, and the knowledge exchange taking place to facilitate research data sharing.

The thesis contributes to the understanding of research data curation as a key to high quality research. Policy and infrastructural development are interconnected with the different stakeholder groups and the motivations and expertise they hold. However, the effort made to create strong infrastructural organizations risks changing the target. The findings show how additional goals and agendas amongst stakeholders risk obscuring the focus on research quality as the goal of data curation in interpretation and application of policy.

For providers of research data services to succeed, it is necessary to combine practice, learning and recognition. This is best achieved through maintaining active memberships in multiple communities; primarily, it is necessary to keep the combination of researcher and research support up to date and to apply best practice from both communities. There is a need to rethink research data support services with a focus on the identity of data stewards as domain specialists and as data management experts. This requires community building and incentives for recognition of multiple memberships. Furthermore, different research support services within universities need to work together and re-think research data services based on a common goal of creating better research.

In collaborations and infrastructure development, agreements on standards, entities and definitions help to facilitate knowledge exchange. These must be developed dynamically through experience and application. A transfer of knowledge occurs between the stakeholders as standards are applied and updated; this requires lines of communication where the stakeholders with multiple identities and stewardship communities function as translators of various perspectives and creators of common understandings.

Sammendrag

Samarbeid og kunnskapsutveksling mellom involverte aktører er nødvendig når infrastrukturløsninger for deling av forskningsdata utvikles. Hensikten ved å tilrettelegge for kuratering av forskningsdata er å styrke kvaliteten på forsking. De sentrale aktørenes ulike perspektiver på hvordan og hvorfor deling av forskningsdata skal finne sted skaper uenigheter i infrastruktursamarbeidet. Parallelt med dette oppstår en ny rolle med ansvar for å legge til rette for god datahåndtering og kuratering av forskningsdata. I den profesjonelle identiteten til denne rollen ligger det et behov for ekspertise innenfor flere fagfelt. Disse problemstillingene leder til følgende spørsmål: Hvordan påvirker perspektiver og ekspertise hos sentrale aktører involvert i deling av forskningsdata samarbeidet og kunnskapsutvekslingen mellom disse? Ved å stille dette spørsmålet ønsker jeg å forstå forholdet mellom tilnærminger og ekspertise hos de sentrale aktørene involvert i deling av forskningsdata. Videre ser jeg på hvordan perspektivene påvirker samarbeid og kunnskapsutvekling mellom aktørene.

Studien er utformet som en Delphi-studie i tre faser med datainnsamling over 14 måneder. Slik fanger studien deler av utviklingen av infrastruktur for deling av forskningsdata. Gruppen med deltakere bestod av 24 aktører med høy ekspertise på feltet. Deltakerne delte sine erfaringer og syn på deling og kuratering av forskningsdata i to intervjurunder samt en spørreundersøkelse. Resultatene som presenteres er basert på den samlede analysen av alle innsamlede data.

Avhandlingen består av tre artikler og en kappe. Hver av de tre artiklene går inn i spesifikke tema innen forskningsdatahåndtering: 1) Utforskning av de ulike datarøkterrollene; 2) Analyse av ulike syn på hvordan datahåndteringsplanen kan legge til rette for kuratering og deling av forskningsdata; 3) Hvordan hensynene til forskningskvalitet og personvern balanseres gjennom livssyklusen til forskningsdata.

Avhandlingen bidrar til forståelse av kuratering av forskningsdata som avgjørende for å styrke kvaliteten på forskning. Politikk og infrastrukturutvikling er sammenkoblet med de ulike aktørene og deres agenda og sakkyndighet. Samtidig er det en risiko for at resursene som settes inn for å skape en ny sterk infrastrukturorganisasjon fjerner fokus fra målet. Resultatene peker på at når politiske føringer skal tolkes og følges kommer andre mål og agendaer blant aktørene i veien for søkelyset på forskingskvalitet som mål for datahåndtering.

For at forskningsdatastøttetjenester skal lykkes, trengs en kombinasjon av praksis, læring og annerkjennelse. Dette oppnås ved å tilrettelegge for at datarøktere kan beholde aktive medlemskap i ulike fagmiljøer, og må til for at personene som tilbyr forskningsdatastøte skal være oppdatert og anvende beste praksis fra både forsker- og forskerstøttemiljøene. Med utgangspunkt i datarøkere som både domeneeksperter og datahåndteringseksperter er det behov for å tenke nytt rundt støtte til datahåndtering. Det er både et behov for å utvikle felleskap og for få på plass insentiver som bidrar til annerkjennelse internt i de ulike fagmiljøene hvor datarøktere har sine tilhørigheter. Tettere samarbeid mellom de ulike tilbyderne av støttetjenester for forskningsdata er etterspurt og kan møtes ved at man tenker nytt om disse tjenestene der et felles mål om bedre forskning får trumfe egne agendaer.

Enighet om standarder, enheter og definisjoner er med på å legge til rette for kunnskapsutveksling i samarbeid om infrastrukturutvikling. Disse må oppdateres og videreutvikles gjennom anvendelse og erfaring. I det standarder tas i bruk og holdes oppdaterte skjer en kunnskapsutveksling mellom de ulike aktørene. Denne kunnskapsutvekslingen fordrer kommunikasjonslinjer hvor aktører med tilhørighet i ulike fagfelt er i stand til å utveksle og oversette de ulike tilnærmingene og skape felles forståelser.

Content

List of included articles

Article 1

Kvale (2021). Using Personas to Visualize the Need for Data Stewardship. *College and research libraries* 82 (3). DOI: **https://doi.org/10.5860/crl.82.3.332**

Article 2

Kvale & Pharo (2021). Understanding the Data Management Plan as a Boundary Object from a Multi-Stakeholder Perspective. *International Journal of Digital Curation* 16 (1). DOI: **https://doi.org/10.2218/ijdc.v16i1.746**

Article 3

Kvale, L.H., & Darch, P. (2022). Privacy protection throughout the
research data life cycle. *Information Research*, 27(3), paper 938.
Retrieved from http://InformationR.net/ir/27-3/paper938.html
DOI: **https://doi.org/10.47989/irpaper938**

List of figures

List of tables

List of abbreviations

FAIR principles - Findable, Accessible, Interoperable, and Reusable, enshrine how data should be made available for further research

Plan S – An initiative promoting Open Access publishing launched in 2018 by a coalition of research organizations and funders (cOAlition S)

UNIT - The Directorate for ICT and joint services in higher education and research

Chapter 1. Introduction

Policymakers, including funding agencies, universities and the European Commission, have for several years promoted data sharing and Open data as a desirable and achievable goal for science (Burgelman et al., 2019; cOAlition S, 2019; European Commission, 2016c; OECD, 2007). Embedded in policies and goals of data sharing are multiple assumptions about how easy it will be to establish data sharing in practice:

- It will be possible to seamlessly/frictionlessly transfer datasets from where they were collected, to long-term infrastructures for management, to contexts where they will be reused (OECD, 2007).
- All research datasets have a potential value in reuse for data analytics and the discovery of new knowledge (Wilkinson et al., 2016; Wood et al., 2010).
- All research datasets can and should be open to all, or at least "as open as possible"(European Commission, 2016a).

Making research data available is expected to increase transparency in research[1], avoid data loss, enable researchers to use data collected by others to produce new knowledge, and strengthen the possibility of the reproducibility and verification of research results and advance knowledge (Borgman, 2007, 2015; Tenopir et al., 2019). Data management and sharing will make research more transparent and increase reproducibility (Borgman, 2015; Kitchin, 2014; Tenopir et al., 2017). The potential of new findings by combining reuse of research data and data science is another driver of data sharing (Kitchin, 2014).

---

[1] In this thesis, the term research is used synonymously with the Norwegian term '*forskning*', which includes research in all scholarly traditions from medicine, to the natural sciences and mathematics, the social sciences, humanities, law, LIS, and so forth.

Despite sustained effort, mandates, investments in computational infrastructures, developments in library research data services, and other development efforts, the gap between the high-flown rhetoric and vision of the policymakers on the one hand, and actual reality/practice on the other remains vast (Plomp et al., 2019; Scroggins et al., 2019; Tenopir et al., 2017; Yan et al., 2020).

The reality of data sharing is everything but a homogenic flow of data (Edwards et al., 2011). Digitally produced datasets are fragile and completely technology dependent (Borgman, 2015). Additional friction is introduced as data are transferred between contexts, both by the transmission itself and the new context to which they are introduced. The boundaries data are expected to cross are disciplinary, institutional, and cultural according to the ideals of a global interdisciplinary research data flow. Strategies to reduce data friction, including the adding of quality metadata (Edwards et al., 2011) and implementing the FAIR principles[2] (Wilkinson et al., 2016), are embraced by policymakers as the answer to data management and standardization of infrastructure and data (Bishop et al., 2019; Higman et al., 2019). However, the infrastructures necessary to facilitate data sharing are very complex as they need to respond to the needs of heterogenous data, organizations, stakeholder groups, and technologies, cultures, and policies. Each of these has different interests, work practices, access to resources and perspective on the goal of data sharing. As a result, the work of coordinating these to work together effectively, over the long term and in a global research environment, is a multidimensional challenge. To approach the challenge of creating infrastructures for data sharing recognizing competing perspectives of stakeholders when it comes to values, right and ethics, in addition to the diversity of data, is essential (Borgman, 2015).

Stakeholders drive the development by applying their expertise (Edwards, 2010). The pioneers of open research pointed out how ‘the problems with enacting such a large-scale vision are not technical but social’(Okerson & O’Donnell, 1995, p. 15), this is still valid today (Cooper & Springer, 2019; Yoon & Kim, 2020). Understanding and unpacking motivations and perspectives of the many stakeholders involved in research data sharing is necessary to investigate the knowledge infrastructure for research data (Borgman, 2015;

---

[2] The FAIR principles argue that all research data should be made Findable, Accessible, Interoperable, Reusable

Sands, 2017). Research data services in libraries are vital part of the infrastructure to support data sharing. But they need help, in addressing the multiple dimensions and perspectives competing to set the agenda for the development of data sharing infrastructure. This dissertation will help libraries address these difficuties by contributing to the different stakeholders' understanding of their own role in relation to the others and analysing the translation of knowledge and perspectives that occurs between the different stakeholder groups as collaboration on research data sharing and infrastructure development is carried out.

1.2 Problem statement

How do the perspectives and expertise of key-stakeholders involved in research data sharing affect the collaboration and knowledge transfer amongst these?

The problem statement is divided into the following three research questions:

1) How and why do differences between stakeholders' perceptions affect their ability to collaborate in the work of research data curation?
2) How and why do stakeholders in research data curation perceive their own roles, and the challenges they face, in facilitating research data sharing? How and why do they perceive the roles and challenges faced by other stakeholders?
3) How do stakeholders manage these differences and facilitate knowledge transfer among the key stakeholders involved in research data curation?

Together these three questions build on each other and highlight different aspects of the problem statement, aiming at understanding the relation between the perspectives and expertise of key-stakeholders involved in research data sharing the collaboration and knowledge transfer amongst these.

This study was designed as a three-phased, modified Delphi study. A Delphi study enabled data to be collected on the current development of the knowledge infrastructure for research data through different perspectives to analyse roles, expertise, and knowledge transfer. In addition, testing how qualitative data sharing could take place while respecting the research participants became an additional exploration of data sharing as a research subject.

Borgman writes that ‘the functions of data in scholarship must be examined from the perspectives of diverse stakeholder’ (Borgman, 2015, p. 14). In this study, the different stakeholders’ approaches to data have motivated an investigation of the knowledge infrastructure including the viewpoint of multiple stakeholders all working with the research data, but from different sides of the table. By returning to the same stakeholders three times over a 14-month period, this study has made it possible to start unpacking the perspectives, motivations, and development in the domain. The methods are detailed more fully in chapter four. To understand how translations and negotiation take place, the theories of boundary objects, invisible work, and marginal people were applied (King, 2016; Star, 2010; Star & Strauss, 1999). Boundary objects and expertise were identified through descriptions and in discussion with different stakeholders involved in the development of the knowledge infrastructure for research data, each of which has perspective, priorities, and goals in the infrastructural development. The theories are presented in chapter two together with other central concepts forming the theoretical background for this study.

1.3 The time frame and context

This section briefly presents the context in which this study was conducted. From the autumn of 2017, when invitations to participate in the study were sent out, the political and infrastructural landscape regarding data sharing and open science in Norway experienced numerous changes, with direct or indirect effects on the knowledge infrastructure for research data and its stakeholders.

Figure 1 Timeline of the changes and events in the research data landscape during the project

Figure 1 illustrates major events that have impacted the development of the current project in different ways. In addition, multiple hearings, seminars, and discussions regarding research data have taken place at universities and at the national and European levels, where Norwegian policymakers participate. By drawing the timeline, the window of the data collection, with a first interview, then a survey and finally interviews, was drawn in the context of some of the events that influenced the participants.

In December 2017, the Ministry of Education and Research published a national strategy document giving directions for the sharing and reuse of research data (Ministry of Education and Research, 2017). This strategy emphasizes 'as open as possible, as closed as necessary' (Ministry of Education and Research, 2017). In addition, the strategy pointed out directions for organizational and infrastructural change.

From January 2018, previously consortia-driven organizations such as the BIBSYS consortia became part of the Directorate for ICT and Joint Services in Higher Education and Research (UNIT) (Ministry of Education and Research, 2017). The BIBSYS consortia is a national provider of library systems for the research and higher education sector with 80 member libraires in Norway. BIBSYS went from being a consortia-driven provider of library online public access catalogue (OPAC) and retrieval services, institutional repositories for Open Access publications, repositories for learning resources, and BRAGE, a research data repository service, to being part of the UNIT directory. In the first interviews, the national strategy (Ministry of Education and Research, 2017) along with both policy and organizational implications were brought up by several stakeholder groups.

During 2018, both the General Data Protection Regulation (GDPR) (The Norwegian Personal Data Act, 2018) and Plan-S were launched. The General Data Protection Regulation created concerns among several stakeholders about the responsibilities regarding personal privacy (Lassen, 2019; Molnes et al., 2019). While Plan-S, an initiative promoting Open Access publishing launched by a coalition of research organizations and funders (cOAlition S, 2019), incited a huge debate among researchers regarding open access. The Plan-S debate raised awareness of different Open Access options, but also contributed to an awareness of other issues related to open research, such as the usage of open licences and data sharing (Gjengedal, 2020; Ruud et al., 2020).

Around the time of the last interviews, a report was published giving recommendations for how Norwegian Centre for Research data should be organized in the future (Ministry of Education and Research, 2019). These recommendations did not have immediate consequences but were instead an indicator of what the next steps in the reorganization of infrastructure would be.

In May 2019, the final guide for the implementation of Plan-S was published (cOAlition S, 2019). This was followed by an updated policy on open research from the Norwegian Research Council (The Research Council of Norway, 2020). In the policy on open research from the Norwegian Research Council, the FAIR data principles are described as the international guidelines for reuse of research data (The Research Council of Norway, 2020). The FAIR (Findable, Accessible, Interoperable, Reusable) principles are guiding principles for enhancing reusability of data (Wilkinson et al., 2016). Both the final guide for implementation of Plan-S (cOAlition S, 2019) and the Policy on open research (The Research Council of Norway, 2020) were produced from several rounds of input from debates and hearings involving multiple stakeholders in research. These hearings and debates were discussions with which several participants were familiar or had been involved in.

Both Plan-S (cOAlition S, 2019), 'the new policy' (The Research Council of Norway, 2020), GDPR (The Norwegian Personal Data Act, 2018) and what would happen with the

Norwegian Centre for Research (Ministry of Education and Research, 2019) were the topics the participants brought up in the second interviews in April 2019.

In a press release in February 2021, the Ministry of Knowledge and Research announced further reorganizations (Ministry of Education and Research, 2021), which will likely result in all three infrastructure providers interviewed being part of the same service provider's organization, UNIT, led by another newly established Directorate for Higher Education and Skills. This ongoing reorganization represents a flip in the ownership of the infrastructural development from university-driven and consortia-owned, to a centralized service provider that will be organized under the Directorate for Higher Education and Skills.

1.4 Thesis outline

Chapter two presents the theoretical background and the main concepts in the study including research data, the sociotechnical infrastructure, boundary objects, invisible experts, and marginal people. Chapter three presents new literature on data sharing according to the stakeholders perspectives presented in the different studies and examples of how boundary objects are used in analysing entitles within scholarly communication. Chapter four presents the study design and ethical consideration. Chapter five provides an overview of the three articles including main findings and research questions investigated in each article. Chapter six discusses the research questions in the light of findings presented in the three articles, limitations of the study, reflections on the research design, and provide recommendation for future research. Chapter seven concludes by answering the research questions and describing implications.

Chapter 2. Theoretical background

The theoretical background and conceptualization of relevant terms describe the bases for understanding the infrastructural landscape presented in this thesis. Terms do not have a single, true meaning but are filled with our own interpretations and understandings of a phenomenon. Different understandings of the same concept are part of the challenge when stakeholders discuss research data and infrastructure. This is illustrated with different perceptions of research, data each serving different purposes. Further knowledge infrastructure and boundary objects emerging out of a socio-technical understanding of infrastructure are introduced. Following this, we present the approach to expertise and people in the margins or borderlands between worlds facilitating translation.

2.1 Research data

The term data can be found everywhere; data are both the industry and raw material of the twenty-first century. Access to data and computational power to analyse and predict based on big data have resulted in metaphors such as ‘data is the new oil’ (Palmer, 2006). Artificial intelligence and data science, by creating predictions based on large amounts of data, have directly linked data with new knowledge. Big data, artificial intelligence and data science are defining our age. Much research involves artificial intelligence and data science focusing on exploiting the potential of big data.

Merriam-Webster defines data as ‘factual information (such as measurements or statistics) used as a basis for reasoning, discussion or calculation’ (Merriam-Webster Dictionary, 2021). The term data was first used in English in 1646, already then in the plural form of data as opposed to the singular datum (Borgman, 2015).

Through historical analyses of the usage of data as a term, Borgman concludes that ‘data are neither truth nor reality. They may be facts, sources of evidence, or principles of argument that are used to assert truth or reality’(2015, p. 17). Different definitions serve different purposes; for the purpose of this study, we will follow Borgman’s definition of research data as ‘representations of observations, objects, or other entities used as evidence of phenomena for the purpose of research or scholarship’ (Borgman, 2015, p. 29). Research data are interwoven part of research itself, defined by their usage (when) rather than their substance (what): their form, origin, or rawness. Borgman’s definition highlights how research data are the representation of phenomena or entities, not the phenomena or entities themselves. Further data are defined by the purpose for which they are used, it is not the representation itself that makes the data; it is the application for a given purpose that makes the data. Data are the evidence of phenomena that are created, compiled, or otherwise established in the process of doing research, and what researchers base their conclusions on. Research data are an interwoven part of the research itself. One person’s data can be another’s results, and the distinction between data and information is a blurry line at best.

Another definition of research data can be found in the OECD report from 2007:

> *Factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings (p. 13).*

Like Merriam-Webster’s definition of data, this definition first lists the material as factual records and examples of these and then the relation to research. Similar approaches to defining research data are found in funders’ policies:

> *The term ‘research data’ is defined in this policy to mean the registration/recording/reporting of numerical scores, textual records, images and sounds that are generated by or arise during research projects. (Research Council of Norway, 2017)*

These definitions reflect how data are understood as the material generated during the research project. The project is again funded by the respective agency, which requires the researchers to share ‘as open as possible, as closed as necessary’ (European Commission, 2016a). These definitions from research funders serves the purpose of arguing for sharing according to the funders policies.

Another term commonly used in the context of research is ‘raw data’. This term reflects the idea of data as something pure and unbiased. Bowker states that ‘raw data is both an oxymoron and a bad idea’ (2005). Data are never raw and unbiased but always the result of a decision-making process (Ribes & Jackson, 2013). Research data are collected and created based on the researcher’s knowledge, interests, funding, tools and methods. A software update or new eyes reviewing the analysis could have an impact on research results, this makes the documentation of decisions made along the way essential to ensure data quality. Research data are defined by their usage, not by their form, origin or rawness, they are not ʽpure or natural objects with an essence of their own’ (Borgman, 2015, p 18).

According to Borgman ‘Public good’ is a premise of open science, where nonrival and nonexcludable are the defining elements (Borgman, 2007, p. 35). Nonrivalrous and nonexcludable are also properties assigned to information and data (Floridi, 2010; Kitchin, 2014, p. 11). The nonrival element implies that sharing does not lessen the value. The rival element is often present prior to publication of findings, and exclusivity to publish first is one of the major concerns amongst researchers regarding data sharing (Corti, 2014). For scholarly production from research data to articles, increased usage can increase value. Data and publications that are commonly used might be viewed as the core knowledge of their respective disciplines. Because the importance of researchers and their contributions to a field is commonly calculated based on reuse in terms of referencing, nonexcludable implies that it is difficult and costly to hold the items exclusive while putting them to use. Work on sharing personal health information in Europe shows how challenging it is to keep access exclusive (Beyan et al., 2020). In this example, the importance of the privacy of the data subjects is valued and holding access exclusive becomes the condition. However, where privacy is not a concern, sharing data openly might simply be easier than creating restrictions for access and reuse.

Research data as a concept is fluid, and the same material can be regarded differently by different researchers, one person’s data is another’s technicality or noise (Haider & Kjellberg, 2016). Haider and Kjellberg describe how the handling of data leaves a path of versions regarded as problems, possibilities, noise, or technical issues. They further observe how “the transience of research data that we encountered is expressed in the various temporally structured descriptions employed, and also in the way in which the digital materiality of data in use is constantly changing” (Haider & Kjellberg, 2016). This is supported by a forthcoming study of a multidisciplinary research group describing that what are referred to as data can be completely different depending on the researcher (Dutoit, forthcoming).

Haider and Kjellberg are referring to “the material research” (2016) when discussing the role of data, implicitly acknowledging the existence of immaterial research traditions, some of which exist without research data. While the contextual understanding of data is broad and inclusive, I acknowledge that no-data research exists, as I will elaborate on in the methods chapter (Chapter four).

Acknowledging that different definitions are in use, serving different purposes is important when interacting with different stakeholders, all with their own perception of what data are. These difficulties in defining data poses difficulties for data curation; policy defines research data according to their goals, while researchers or archives might operate with different understandings whereby applying the policy subsequently becomes challenging.

While Borgman clearly makes the usage and purpose of data for research or scholarship a condition, the lines become blurry for data collected for one purpose but not used. An inclusive understanding of purpose and usage is necessary to ensure that data from unpublished results, typically negative findings or research that remains uncompleted for various reasons, are included in our understanding of research data. In a knowledge infrastructure for research data, describing the data is just as important as the data per se (Bowker, 2005); documentation and descriptions recognize the temporality of the data. The when and purpose can be just as essential as more material attributes of data, and the documentation and metadata are the strings tying the data to the research and their origin (Borgman, 2015).

2.2 The sociotechnical infrastructure

This thesis follows a sociotechnical approach to infrastructure which is distinctly different from the commonly understood definition of infrastructure found in the dictionary. According to the Oxford English Dictionary, infrastructure is ʽA collective term for the subordinate parts of an undertaking; substructure, foundation; spec. the permanent installations forming a basis for military operations, as airfields, naval bases, training establishments, etc'. ((OED online, 2018). Whereas a more common understanding of infrastructure is described in lexico.com: ʽThe basic physical and organizational structures and facilities (e.g. buildings, roads, power supplies) needed for the operation of a society or enterprise' (Lexico Dictionaries, 2021). Both these definitions are material emphasizing the physical installations. In the contrast ʽhuman infrastructure' emphasis the organization of actors necessary for work to be accomplished (C. P. Lee et al., 2006). The sociotechnical approach to infrastructure used in this thesis acknowledges both the physical installations and the human and organizational. In data curation, the "people" side of the infrastructure is where the larger gaps are found (German Council for Scientific Information Infrastructures (RfII), 2019).

Within research data management, the terms infrastructure, data infrastructure and e-infrastructure describe the combination of digitally based technology, resources (data) and communication (protocols, access management and networks) and the support people and organizational structures (Kitchin, 2014, p. 23; Whyte, 2012, p. 210). Star and Ruhleder's (1996) describe infrastructure as a relational concept in which technology can become infrastructure within a given context or practice. However, it is not the hardware or software that forms the infrastructure; something becomes infrastructure when used and applied in a context. The question is *when* rather than *what* is an infrastructure, as research data infrastructure is a relational concept; it is the application that makes something infrastructure not the technology within (Star & Ruhleder, 1996). Star and Ruhleder (1996) describe eight dimensions (Table 1) from which infrastructure emerges, each highlighting different qualities of the infrastructure as a socio-technical concept. For the current project the dimensions relating to requests (1. Embeddedness), expertise (4. Learned as part of membership and 5. Links with conventions of practice) and standards (6. Embodiment of standards and 7. Built on an installed base) are particularly relevant, and are contextualized in the two following sections.

Infrastructure emerges with the following dimensions

1. Embeddedness	2. Transparency
3. Reach or scope	4. Learned as part of membership
5. Links with conventions of practice	6. Embodiment of standards
7. Built on an installed base	8. Becomes visible upon breakdown

Table 1 Dimensions of infrastructure from Star and Ruhleder (1996)

‘Embeddedness’ describes how infrastructure is built into other structures, both social and technological. Infrastructures respond to a need and do not exist independent of it. The infrastructures for research data are constructed to respond to political and professional requests; the technologies are adapted to best support these needs and be adaptable to future needs. The key challenge here is getting data curation infrastructure embedded in the day-to-day work practices of researchers. Currently many researchers are unaware of for various reasons are reluctant to use this infrastructure. The embedding of research data management and curation in the researchers’ workflow represents a cultural change in how research data are shared and disseminated.

‘Learned as part of membership’ reflects how communities of practice take the existence of artefacts and organizational arrangements for granted (Star & Ruhleder, 1996), while memberships for outsiders require that they become familiar with the practice within (Star & Ruhleder, 1996). The expertise held by those developing and using infrastructure are part of the infrastructure itself, as highlighted by the concept ‘knowledge infrastructure’ which ‘comprise robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural world’. (Edwards, 2010, p. 17). In his application, Edwards (2010) uses knowledge infrastructure as a concept to describe the sharing of data and scientific knowledge regarding climate data. In the context of research data management and knowledge infrastructures for research data the communities of practice are many and the expert stakeholders can hold membership in one or multiple of these. Understanding the membership in communities as a dimension of infrastructure is central to the approach for investigating infrastructures use in this thesis.

The phrase ‘links with conventions of practice’ emphasizes how infrastructures are developed by users through usage. This aspect of infrastructure reflects the dynamic relationship between the user, technology, and usage. Design shapes how an infrastructure is used, and usage can again influence the design (Bijker & Law, 1992; Lindsay, 2003). Lindsay describes a transition in the usage of computers and how ‘the construction of users, user-representation, and technology is not a static, one-time exercise by the designers’ (2003, p. 30). Research data infrastructures are in a developing phase, where they are moving from being something for a limited group of researchers within certain domains to being adapted by all researchers as part of everyday research infrastructure. Research data infrastructures are no longer reserved for use in traditionally data-intensive domains; however, these are still the most visible practitioners influencing their development (Borgman, 2015).

‘Embodiment of standards’ refers to how infrastructures connect to existing standards and other infrastructures; they do not exist in a vacuum but must work interoperable with the past and the future. Applying open standards to support transparency is also one of the key messages in research data management as communicated through the FAIR principles (Mons et al., 2017). In the FAIR principles, usage of existing standards is encouraged rather than the development of new standards (Wilkinson et al., 2016). The embodiment of standards according to the FAIR principles is essential in mapping the knowledge infrastructure for research data to past and the future.

Finally, infrastructures do not grow out of thin air but are built on something, they are ‘built on an installed base’. The name refers to how existing paths are used for a ‘backwards interoperability’(Star & Ruhleder, 1996). When an organization is developing new services for research data, these are created based on the existing infrastructure, expertise and collaborations within that organization. The libraries existed long before the research data services did – the libraries form the installed base upon which the research data services are built – the research data services use existing library infrastructure (e.g., physical space, staff & their expertise, curatorial methods, technical facilities etc). The installed base is the existing infrastructure on which new infrastructure is built and defines the standards for current and future interoperability. Infrastructures are developed based on existing social and technical structures, applying and often maintain standards. Only then can the infrastructure be adapted to future usage while maintaining the connections with the existing, the installed base.

For research data infrastructures, the term knowledge infrastructure is widely used (Borgman, 2015; Edwards, 2010; Edwards et al., 2013). The socio-technical approach to infrastructure is embedded in the knowledge infrastructure concept where the technical and human are approached under one (Borgman, 2015; Edwards, 2010). What we know and who we trust are learned and based on our social relations (Van House, 2004).

> *There is no 'view from nowhere' – knowledge is always situated in a place, time, conditions, practices, and understandings. There is no single knowledge, but multiple knowledges. (Van House, 2004, p. 40).*

This understanding knowledge as socially situated, based on trust, credibility and practice connects with the concept of infrastructure as defined by Star and Ruhleder. By combining the terms, knowledge infrastructure as a concept highlights the need for knowledge to connect the infrastructures with the future and the past, the invisible work conducted to facilitate this, and the added value created through curation and exchange of research data. In this case, this involves a knowledge infrastructure specifically for the collection, circulation, storage, management, curation, and reuse of research data. Knowledge infrastructure describes the transformation of observations into widely accepted knowledge, which Edwards argues is the reason why we today 'can think globally about climate change'(2010, p. 8). According to Edwards, the knowledge infrastructure becomes the condition for the existence of a whole area of research. It is not the bits and pieces by which that the infrastructure is built up, but rather the continuity of multiple systems and people in a global network of climate data (Edwards et al., 2013). There are no longer clear boundaries between data, infrastructure, and research, it is one knowledge infrastructure.

This understanding approaches infrastructures as ecologies or complex adaptive systems consisting of numerous systems with different origins and different goals interoperating through standards, norms and social practices (Borgman, 2015, p. 33; Edwards et al., 2013, p. 5). A knowledge infrastructure becomes the multitude of systems and people connecting and creating scholarly output through the research data life cycle. The knowledge infrastructure is the ecology in which research is taking place where the ‘digital materiality of data in use is constantly changing’ (Haider & Kjellberg, 2016). The knowledge infrastructures as a concept encourages a holistic perspective on research data as a process rather than a fixed entity. The knowledge infrastructure is the ecology making digital research possible. Knowledge infrastructure as a concept that overlaps with the concept of knowledge commons (Hess & Ostrom, 2007); both concepts can be used for describing the ecologies for resource exchange with the goal of knowledge creation. Knowledge commons, however, is also used as a name to denote institutional data repositories.

2.3 Boundary objects

The knowledge infrastructure is the environment investigated; the different stakeholders are actors within this environment, all holding different expertise and having different perspectives and objectives but who nevertheless have to work together: boundary objects are one way of getting these very different entities to work together effectively. Infrastructures, objects, and communities of practice are all interwoven in an ecology. Boundary objects were developed as a concept and introduced by Leigh Star as ‘those objects that are plastic enough to be adaptable across multiple viewpoints, yet maintain continuity of identity’(Star, 1989). The understanding of and typology of boundary objects were developed in the investigation of researchers’ collaboration with each other and with other professions within a research hospital (Star, 1989) and in a zoological museum (Star & Griesemer, 1989). Both explorations of boundary objects as concepts describe how different stakeholders relate to the same object in different ways: understandings are translated through the usage of boundary objects, which helps facilitate collaboration between different fields.

Star and Griesemer identify four types of boundary objects: repositories, ideal types, coincident boundaries and standardized forms. Each has distinct characteristics and functions (Star & Griesemer, 1989; Star, 1989).

- Repositories are described as ‘ordered files of objects that are indexed in a standardised fashion’ (Star, 1989) memory institutions; libraries, archives and museums are named as some examples of repositories.
- Ideal types are platonic objects that do not accurately describe anything and are ‘abstracted from all domains’(Star, 1989); however, they have a symbolic meaning in the means of communication as an object in between, connecting different domains. Examples of ideal types are maps or early atlases of the brain, none are exact descriptions of an item or location, they are adaptable intermediators of information (Star, 1989).
- Coincident boundaries are ‘common objects that have the same boundaries but different internal contents’ (Star, 1989). Star uses the example of different aggregations of data in the same geographical area (Star, 1989), the same physical location have different contents for professionals of different domains, say a biologist, a geologist and an archaeologist.
- Standardized forms are ‘objects that can be transported over long distance and convey unchanging information’ (Star, 1989). Standardized forms are created for the collection of certain information describing and contextualizing something. Star uses the “Fits sheet” where relatives filled out information about epileptic attacks for the doctor to analyse as one example of a standardized form (Star, 2010). In analysing these, she finds that much information is found not to fit in either of the questions or columns in the form, but rather scribbled on the side, what the caretaker finds relevant to communicate is not necessarily what the doctor finds relevant to ask about, the standardized form is the common object translating between these worlds (Star, 2010).

The term boundary describes the point where different worlds meet, start and end. As stated by social anthropologist Barth, ‘It is clear that boundaries persist despite a flow of personnel across them’ (1996, p. 9). His theory on how identity is created on the boundaries in the encounter with others was developed based on the studying of cultures. Through the notion of boundary objects, this approach is adapted for the investigation of society and technology and how items and objects have different meanings or functions in different cultures. The fact that the objects originate in, and continue to inhabit different worlds reflects the fundamental tension of science: how can findings that incorporate radically different meanings become coherent? (Star & Griesemer, 1989, p. 392). It is the social meaning of the object that is investigated, which is developed through its usage and in dialogue with its users.

One key feature of boundary objects is their plasticity; the same objects are used and understood in different ways. Boundary object as a concept and label are used and applied to something for the purpose of analysing their function. When the boundary object as a label is applied, meaning is read into the image of an object, influencing further development. This is a dynamic process of interpretation following the plasticity of the boundary object. Different understandings of the same concept are highlighted by Star as a reoccurring misunderstanding in dialogue with others:

> *When a named symbol of an action or thing is presented to us during an inquiry dialogue, we easily move from the presumption of a consensus on a shared understanding of the context to the presumption of a consensus-shared meaning that it has for all of us in that inquiry dialogue. (Boland, 2016, p. 236)*

By conceptualizing boundary objects, dialogue, understandings of different perspectives and the translation between perspectives are allowed. Jacob & Albrechtsen (1998) used boundary objects as concepts in the analysis of two digital library systems, here finding that the mediation between documents and users can be regarded as an information ecology enabled by the usage of both tacit and explicit knowledge. In their articles different elements in the digital library systems are categorized according to Stars four boundary object types (Jacob & Albrechtsen, 1998). Corsaro (2018) addresses the importance of boundary objects to coordinate interaction in digital contexts. Arguing that when something is transferred across boundaries such as physical to digital, the boundary objects are what facilitate the translation and transition (Corsaro, 2018).

After elaborating on what boundary objects are, a timely question is also “what is not a boundary object” (Star, 2010). Star describes the limits of when it is reasonable to use the concept of boundary objects as scale and scope (Star, 2010). For scale she uses words as examples; words might fit the description of boundary objects, but it will rarely be meaningful to apply the theories of boundary object to analyse how words carry different meanings depending on context (Star, 2010). Regarding scope, Star argues that boundary objects are useful for studying limited contexts rather than applying them to large general claims, such as 'many people have different interpretations of the American flag' (Star, 2010). Further arguing that this 'does not get us very far analytically in understanding both the materiality and the infrastructural properties of this flag' (Star, 2010). While acknowledging that anything can be a boundary object, Star argues that using a boundary object as a meaningful way of analysing something has its limits in scale and scope. Analysing something as boundary objects is most useful at an organizational level, to understand the materiality and infrastructural properties of objects (Star, 2010).

Wakeford suggests in her application of Star´s theory a 'relational mapping—between the multiple marginality of people and the multiple naturalizations of objects through boundaries and standards' (Wakeford, 2016, p. 82). Drawing inspiration from Wakeford's application of Star, my analysis of the knowledge infrastructure attempts to follow the relations between the people, objects, boundaries, and standards mapping the function of expertise and objects in the infrastructure. The relations between infrastructures, expertise and boundary objects becomes visible in the work on classification by Bowker and Star (1999). When analysing and describing the infrastructures, Bowker and Star make visible the human-made concessions regarding 'the nature and quality of data before hoping to gain any kind of entry into hospital information systems' (Bowker & Star, 1999, p. 313). Agreements on practice and classification are essential for the quality of the data entries as concessions are based on the knowledge and expertise of the actors involved. Using concrete examples from practice in analysing social functions is recurrent in Star's work. Hence, I explore roles and knowledge exchange in the knowledge infrastructure for research data management and knowledge exchange through a sociotechnical infrastructure perspective applying the theories of boundary objects, marginalization, and invisible work building on examples from the three articles compiled in this thesis.

2.4 Invisible experts and marginal people

Knowledge in terms of expertise can be understood as different skills people possess and the way knowledge necessary to perform tasks in a workplace is both structured and unstructured

(Hildreth & Kimble, 2002). Each group has distinct practices with tacit knowledge embedded in the socializations and interactions between members (Nonaka, 1994). Nonaka and Takeuchi (1995) differentiate between tacit and explicit knowledge, as the concrete and invisible types of expertise individuals hold.

Collins and Evans (2008) develop this further in a more complex table of expertise, including terms such as interactional expertise and contributory expertise. Both interactional expertise and contributory expertise are considered to require specialist tacit knowledge (Collins & Evans, 2008). Interactional expertise is the lower level of specialist knowledge used to describe the skills of communication and translating (Collins & Evans, 2008). Collins and Evans describe interactional expertise to be found in between the activity and the theory of the activity as described in books or computers (Collins & Evans, 2008). Interactional expertise is the expertise required to understand and apply the semantic, to master a skill on the level to be able to describe it with a fluent language, without having the practical skill to preform it, slowly accomplished “through conversation with the experts”(Collins & Evans, 2008, p. 32). To perform or conduct an activity with competence, contributory expertise—which is considered the highest level of expertise—is required (Collins & Evans, 2008). Performing, doing, or knowing something is a practical skill often visible when gained, but once it is mastered, it becomes routine. Contributory expertise overlaps with the concept of ‛knowing in practice’ (Orlikowski, 2002) connecting the expertise and tacit knowledge with communities of practice. Orlikowski describe knowing in practice as “an ongoing social accomplishment, constituted and reconstituted as actors engage the world in practice” (Orlikowski, 2002, p. 245).

This categorizing of different types of expertise can be useful for naming and making different types of skills visible. Invisible work is work that is taken for granted as part of a background routine. ‛If one looked, one could literally see the work being done – but the taken for granted status means that it is functionally invisible’ (Star & Strauss, 1999). Invisible work is further described to often ‛include tacit and contextual knowledge, the expertise acquired by old hands, and long-term teamwork’ (Star & Strauss, 1999). Sylvain addresses the role of invisible work in infrastructural development, arguing that it is the ‛industrial design, regulatory arrangements and human labour that have put the Bigtech in their positions of control’ (Sylvain, 2019). Through this argument Sylvain points towards the importance of policy and regulation as parts of the invisible work conducted in infrastructure development. One example of the invisibility of infrastructural work is ʽthe tendency to delegate responsibility to ICTs as an increasingly authoritative intermediary agent' (Floridi, 2015, p. 58). Floridi refers to this as the action-oriented analysis of ethical problems, referring to people's tendency to ʽdismiss an error as only the fault of a computer' (Floridi, 2015, p. 58). Instead of delegating and dismissing ethical concerns as the responsibility of the machine, Floridi argues for information ethics as a way of addressing the importance of moral concerns that are not immediately agent- or action-oriented and anthropocentric (Floridi, 2015, p. 62). ʽWithout information there is no moral action, but information now moves from being a necessary epistemological prerequisite for any morally responsible action to being its primary object ontologically' (Floridi, 2015, p. 64). Information ethics investigates the ethical impact of ICTs on human life. The level of abstraction used for information ethics is helpful in analysing the sociotechnical infrastructure by acknowledging the involvement, decisions, and invisible work of people at all levels of computational development.

In her characterization of knowledge infrastructures, Borgman (2015) uses the term invisible work, building on Star and Strauss when stating that, ʽTo document, organize, and manage scholarly information is essential for others to discover and exploit it. Invisible work is both glue and friction in collaborations, in the development of tools, in the sharing and reuse of data, and many other infrastructure components' (Borgman, 2015, p. 34). Such intermediary work within an infrastructure is conducted by an intermediary or translator in dialog with users and requires high-level expertise. This expertise often remains invisible (Jacob & Albrechtsen, 1998). However, when the intermediary work is removed, it also becomes visible (Jacob & Albrechtsen, 1998). This happens typically when workflows are changing in a knowledge infrastructure. The importance of networks and good communication is exemplified by Star and Strauss as the work of secretaries (Star & Strauss, 1999).

In the context of boundary objects, Star and Griesemer also introduce the concept of marginal people or 'allies' (1989). Between the allies and object, there are passage points where a translation is taking place. When using the term marginality, it refers to memberships in multiple communities of practice, often existing in the margins of these communities. Marginal people are both inside and outside of multiple worlds (Bowker & Star, 1999, p. 300). However, they can be overlooked or excluded because they are not in the centre of either community, but rather with one foot in each. They are the borderlands, the coexistence of two communities of practice in one person (Wakeford, 2016). This marginalization is not synonymous with the traditional usage of marginalized as powerless. By having memberships in multiple communities and speaking the language of different worlds, marginal people can translate the meanings of boundary objects. The ʻability to triangulate from the margins’, to see and think differently about things, is one of the skills assigned to the marginal people (King, 2016, p. 342). The importance of having people who see things from inside and outside at the same time is often unrecognized. The translations performed can be invisible to each of the communities but are essential for boundary objects to move between the different communities. Wakeford argues that:

> *'first, we must think of "communities of practice" in so far as they link together actions, people and artifacts. Second, we have to find the "boundary objects" and the work of translation and communication between communities of practice. Third we should consider "borderlands and "monsters"" (Wakeford, 2016, p. 82).*

While borderlands are the coexistence of communities within a person, monsters are similar but different, defined as ʻobjects which refuse to be naturalized’ (Wakeford, 2016). Both are marginal people, but the coexistence of communities can be either functioning as borderlands or dysfunctional monsters where there are multiple ostracisms. In the context of knowledge infrastructure, the different stakeholder groups represent different communities of practice. Those individuals who operate between these communities are the borderland, the marginal people.

Chapter 3. Literature review

The studies presented here contribute to the analysis conducted in this thesis by providing knowledge of expertise within the knowledge infrastructure for research data. The literature is structured according to two main themes: Perspectives on data sharing expertise; and the application of boundary objects within scholarly communication. Structuring perspectives on data-sharing expertise according to stakeholders addressed in the studies illustrate gaps and tensions between stakeholder groups and the way research data management is addressed. Examples are given to show how boundary objects are used within the domain of scholarly communication to address entities that cross the gaps between stakeholders' domains.

The literature is selected to complement the literature presented in each of the articles and to elucidate the research questions investigated in this thesis. The focus has been on identifying recent research (2017 – spring 2021). The articles were identified through: Searching in the library and information science data base LISA; by searching and browsing journals such as Library and information science research, Journal of documentation, Library tech, Information Research, Jeslib, Portal, and College and Research Libraries, in addition to browsing in the proceedings from ASIST, IConference, and International Digital Curation Conference. Terms used in searching include research data, data sharing, research data management, data steward, boundary objects, knowledge infrastructure, Socio-technical infrastructure. Through reading, annotation, note taking and using keywords, the literature presented in this chapter is identified as contributing to the analysis of the discussion in chapter six.

3.1 Perspectives on data sharing expertise

The sharing of research data and the creation of machine-readable data are creating new workflows; along with this comes the call for new types of expertise. Several studies highlight that balancing different skills needed for data stewardship and prioritising new skills to be acquired and activities to be performed remains a challenge (Darch et al., 2020; Hannah et al., 2020; Henderson, 2020; Scroggins et al., 2019). This chapter is structured according to the stakeholders' perspectives addressed. The literature describes how policymakers embrace the FAIR principles, while practitioners struggle to agree on what the main terms imply in practice. The repository perspective focuses on interoperability and reusability. The gap between high-quality, detailed documentation and interoperable standards makes the interoperability of metadata a complex issue for data repositories. In addition, data repositories struggle with targeting their communities and identifying the user, or re-user of data. Researchers face the challenges of balancing their own agenda of conducting research with other stakeholders' agendas. They also struggle to disseminate the homogeneity of research as domain and to ascertain how traditions of research are impacted by research culture and international collaborations. Data management services in universities are often coordinated by the library and include collaborations with a multitude of stakeholders, each bringing their own domain practices and interests to the table. At the same time, services are requested at multiple levels, making the targeting of curation services and divisions of stewardship roles a complex task to monitor.

3.1.1 Policymakers and the FAIR principles

When investigating the extent to which researchers' data sharing behaviour is influenced by policy pushing for data-sharing, Mallasvik and Martins find the co-existence of different logics; one prizing openness and sharing, the other control and self-interest (Mallasvik & Martins, 2020). This coexistence of different perspectives is echoed by findings from Higman and colleagues pointing at a gap between different goals with data sharing; internal benefits of sharing for research within a research project, are not always related to open data and interoperable sharing, in the way open and interoperable are emphasized by policymakers (Higman et al., 2019). This resonates with Haider and Kjellberg (2016) who draw up a distinction between the ideas of open, available and accessible and data sharing as value embedded and the practice of research data management which resides in the background. While data sharing and open access for research data are different approaches and different conditions for availability and reuse, research data management ostensibly following the FAIR principles can be seen as the enabler (Haider & Kjellberg, 2016; Kraaikamp et al., 2019). Such gaps between stakeholders' perspectives need to be acknowledged and balanced in research data services.

Policymakers are increasingly embracing the FAIR principles as the standard of excellence for the ways by which research data and data-driven science can be improved (Burgelman et al., 2019; European Commission, 2016b). FAIR and Open embed values attracting the attention of policymakers, whereas research data management holds internal benefits for the researcher, project and institution by dealing with practices throughout the life cycle (Higman et al., 2019). In a study of how data managers address the FAIR principles, Donaldson and colleagues find that there is no clear agreement among practitioners regarding the definitions of the main terms, namely findable, accessible, interoperable and reusable (Donaldson et al., 2020).

3.1.2 Repository perspective

Metadata which lie at the core of the FAIR principles is not a straightforward issue and challenges are common in terms of the right metadata standards and level of details in metadata descriptions (Bishop et al., 2019; Kim et al., 2019). For retrieval and interoperability, metadata should be more standardized to clarify potential ambiguity across disciplines (Kim et al., 2019; Mallasvik & Martins, 2020). There is also a need to improve metadata integrity, consistency, and transparency in data archives (J.-S. Lee & Jeng, 2019). On the other hand, extraneous metadata have been found to make assessment of usability more complex for (human) reusers (Bishop et al., 2019). Further proper documentation requirements related to data code and methods is a problem (Yan et al., 2020). The focus on interoperability does not address the issue that researchers also need the right competencies to interact with data (Bishop et al., 2019). Accordingly, the most difficult aspects of FAIR to fulfil are reusability and interoperability (Kim et al., 2019).

The EU imitative European Open Science Cloud (EOSC) is developed with a focus on including different research data infrastructure projects in the same European solution (Limani et al., 2019). However, it does requires multiple types of expertise to implement best practices and open standards across research disciplines to enable interoperability between infrastructures (Limani et al., 2019). Human skills such as listening, translating needs to data, and understanding a collection and its materials are highlighted as some of the many parts of work taking place in data organization (Donaldson et al., 2020). The process of depositing data is based on humans and trust (Kim et al., 2019). Depositing data means that responsibility, including the rights and permissions to manage and access data, is transferred from researchers to data organizations through forms and contracts (Kim et al., 2019). With machine access, the repositories no longer have control over designated community choices (Donaldson et al., 2020). This change in interaction and designated communities is forcing repositories to rethink the way they operate and interact with users. The focus in digital curation is currently on preservation of resources, without clear understandings of users and usefulness (Kouper, 2016). Both Donaldson and Kouper (2020; 2016) suggest that there needs to be more focus on relevance and future usage.

Data-reuse was primary for the purpose of new and comparative analyses. Only rarely were data reused for the purpose of reproducing research (Yan et al., 2020). Experience with data-reuse was found to significantly increase perceived benefits of data sharing, implying that data sharing, and data reuse are interrelated (Yoon & Kim, 2020). The reuse experience as driver for sharing, resonates with Goben and Griffin's observation that the culture of a research group is more significant for a researcher's data sharing behaviour than formal training (Goben & Griffin, 2019)

3.1.4 Researchers' perspectives

Investigations of the researchers' perspectives uncovers several tension between researchers and other stakeholders in the domain (Berman, 2017; Cooper & Springer, 2019; Darch et al., 2020). Moreover, there is no such thing as a typical researcher. Researchers and their practices are homogenous and constantly pushing boundaries by creating new knowledge, new methods, and critically disseminating (Berman, 2017). Perusing data sharing must take into account the epistemic culture of the domain, to avoid unintended negative consequences

(Darch, 2018). Whereas multinational research collaborations, require dialogue between the partners regarding issues such as accreditation, licences and conditions for reuse permissions to share data openly (Zuiderwijk & Spiers, 2019).

Researchers request both outreach services, such as university level support and infrastructure and inward organizational services including data description and versioning (Berman, 2017). Discipline-specific support play an important role in facilitating good data management (Andrews Mancilla et al., 2019). However, there is a risk that good intentions are lost behind an ambiguous term when there is confusion about what data stewardship means (Andrews Mancilla et al., 2019). In research data management, there is a need to develop services that respond both to the demands of the researchers and to the demands of the academic leadership at the university, recognising and addressing the dissonance between these demands (Berman, 2017). These researchers’ data-sharing practices are influenced by institutional policy informing their beliefs and guiding their behaviour, balancing the logic of openness versus that of control (Mallasvik & Martins, 2020).

The identification of emergent data communities and tailored support is essential to allow grass-root initiatives to grow (Springer & Cooper, 2021). Data communities building from the ground up are important; these need the support from larger organizations through the development of technical solutions that make data types more shareable (Springer & Cooper, 2021). The gap between the motivations of the researchers and the infrastructure developers requires addressing divergent motivations to alleviate tensions created between domain researchers and infrastructure developers (Darch et al., 2020). The expectations of the domain researchers must be managed carefully (Darch et al., 2020). The developers have a different commitment and often have multiple projects in addition to constraints based on the installed base on which the infrastructure is developed (Darch et al., 2020).

3.1.3 Data management services in universities

Researchers and universities have been given policy mandates through FAIR, requirements of data sharing and data management plans. Universities have attempted to develop data management services to help fulfil these mandates and policy requirements (Cox, Kennan, et al., 2019a). As policymakers increasingly are requiring data to be made available according to the FAIR principles, researchers are turning to their university libraries for assistance (Darch et al., 2020). University libraries play a leading role in planning and coordinating research data services within their communities (Cox, Kennan, et al., 2019a; Darch et al., 2020; Kim, 2020; Tenopir et al., 2017, 2019). Still, university libraries struggle to fill gaps and balance conflicting interests when providing services and coordinating collaborations (Cox, Kennan, et al., 2019b; Darch et al., 2020).

The development of research data services is constrained by existing structures and cultures, creating the need for new partnerships in collaboration with both internal and external partners (Kim, 2020). Within the universities, the IT Centre and Office of Research are the most frequent partners, while other collaborations within the universities are with university archives, legal offices, and other research support units (Tenopir et al., 2017). The importance of collaborations is also noted by Goben and Griffin (2019) in their analysis of research data management needs in institutions. They find that libraries are taking steps to collaborate effectively in communication support; other positive initiatives include handoff of data by working together with other entities including university IT department, science award recipients and supercomputing centres (Goben & Griffin, 2019).

To fill gaps in expertise between researchers, data management and infrastructures, several university libraries have established digital scholarship centres (Hannah et al., 2020; Mulligan, 2016; Verhaar et al., 2017). Digital scholarship centres are criticized for focusing too much on innovation and leadership, rather than on applying existing expertise in library liaisons (Hannah et al., 2020). The library service and being innovative in the development of new units and services , they are however not taking advantage of the existing structures and expertise within the library (Hannah et al., 2020). Digital scholarship centres are currently unsuccessful in incorporating existing expertise in libraries, notably through liaisons and their subject expertise (Darch et al., 2020; Hannah et al., 2020).

University libraries are critical components of the knowledge infrastructure and referred to as 'an installed base on which research data services are built' (Darch et al., 2020, p. 2). As any installed base, the library comes with conditions and a given set of expertise and focus in their services. Lack of necessary expertise is a challenge; currently libraires are both hiring staff specifically to support research data services and reassigning existing staff (Tenopir et al., 2017). This reorganization in both tasks and staffing require library staff who are knowledgeable and have the aptitude to learn new skills (Tenopir et al., 2017). The absence of IT services and research office in studies of research data services in universities might represent a knowledge gap, as these are involved in the research data services (Cox, Kennan, et al., 2019a). Consultative services are more frequently offered by the libraries than technical research data services (Cox, Kennan, et al., 2019b; Tenopir et al., 2019). The domain of data curation should function as a trading zone of professions, activities and skills, rather than professionalising one data curator (Kouper, 2016). Being an administrator and entrepreneur in addition to including staff with backgrounds such as reference and instruction librarian and data science expertise is an example of multiple skills needed in a domain where skills shortage remains a challenge (Cox, Kennan, et al., 2019b; Henderson, 2020).

3.2 Applications of boundary objects in the literature

Successful data management requires coordination and collaboration of multiple, heterogeneous groups of stakeholders. Boundary objects are those objects shared between multiple heterogeneous groups that facilitate coordination and collaboration between these groups. In analysing the role of objects in collaborations across multiple domains and contexts, boundary objects are useful for seeing how disagreements and compromises can be productive in a dynamic domain. Boundary objects take the focus away from creating consensus, and instead accept the heterogeneity of the stakeholders. The literature presented below provides examples of how boundary objects are used to contextualize elements of open research. The examples of open access (Montgomery & Ren, 2018; Moore, 2017), research evaluation (Åström et al., 2017), policy development, and documentation (Wu & Worrall, 2019) all highlight how collaboration and development can flourish without consensus. Different perceptions of a concept based on context (Montgomery & Ren, 2018) or application (Åström et al., 2017) are relevant when analysing how data sharing is happening across the same dimensions.

In the literature on open research, Moore argues that open access functions as a boundary object, resonating differently between communities of practice (Moore, 2017). This is echoed by Montgomery and Ren (2018) in their analysis of open knowledge in China. Montgomery and Ren included the whole perspective on open research and how this concept resonated differently depending on cultural context. Also withing the bibliometric approach to scholarly communication, boundary objects are used to analyse scientific publications and the intersection between classification of scientific disciplines and research evaluation (Åström et al., 2017). They describe how the classification system for scientific disciplines used in research evaluation originates in a bureaucratic context, and through classification, the scientific publications are translated to something that can be evaluated by bibliometric measures (Åström et al., 2017).

In the context of research data, Donaldson et al. (2020) describe geoscience data as boundary objects because they are 'meant to capture an understanding of an environment or condition at a specific time and place'. Through this description, it is reasonable to assume that Donaldson and colleagues include descriptions (e.g. time and place) in their notion of geoscience data. The understanding of 'data and their documentation' as boundary objects is also put forward by Wu and Worral. Based on a qualitative study of researchers' perceptions of data sharing and ownership, they approach both data with documentation and policy as boundary objects. Arguing that with a common structure and standardization across different research teams, research communities and the funding agencies policy function as a boundary object (Wu & Worrall, 2019).

These applications of boundary objects presented in this section are limited but illustrate how documentation, policy and open research are applied differently by different stakeholders of research (Donaldson et al., 2020; Montgomery & Ren, 2018; Moore, 2017; Wu & Worrall, 2019).

Chapter 4. Study design

This chapter presents the methodology and ethical considerations conducted in the study, including choices and reflections regarding the method, selection of participants, data collection, analysis, processing of personal data and data sharing.

4.1 Delphi method

A Delphi study was conducted to collect data from representatives of four different stakeholder groups: policymakers, infrastructure providers, research support staff and researchers.

A Delphi study involves multiple rounds of data collection with a single group of study participants. All participants are expert stakeholders in the topic of the study (in this case, research data management). Delphi studies are focused on reaching agreement between stakeholders or solving an issue. In a rapidly developing domain such as that of data sharing, the Delphi method offers a way of systematically collecting solution-oriented opinions on a subject or problem. A Delphi study typically contains three phases: 1) the exploration phase, 2) the evaluation phase and 3) the concluding phase (Ziglio, 1996). In each phase, data are collected and analysed, and the intermediate results are used in the development of the next data collection phase. In the exploration phase addresses the subject of the study with a holistic perspective and, if needed, providing additional information elucidating the topic (Ziglio, 1996). The evaluation phase gathers and assesses the views on the different issues investigated in the study in dialogue with the expert groups (Ziglio, 1996, p. 9). As a concluding phase, all the data are brought together and the issues addressed in the study are investigated through an integrated analysis. (Creswell & Plano Clark, 2018)

Delphi studies were initially developed to gather opinions on security issues, policy and development in the USA in the 1950s (Ziglio, 1996) but were later adopted by educational science and marketing. Gupta and Clarke (1996) describe the Delphi method as suitable not only for forecasting but also as a procedure for the verification and perfection of a consensus in decision making when searching for innovative solutions; they further describe the method as useful in situations where there is a lack of consensus or agreement between stakeholders. These are features making a Delphi study a suitable method for my project as the domain of research data management is characterized by the search for innovative solutions combined with a lack of consensus between stakeholders. Within library and information studies and research on academic libraries, in particular, the method has become increasingly popular over the last 20 years (Lund, 2020). Within library and information studies, the Delphi studies used most frequently are electronically adapted versions of the traditional Delphi method and different modified Delphi studies. The Delphi method in library and information studies is most commonly used to identify, predict or evaluate issues and for policy development and refinement (Ju, 2013). My first encounter with the Delphi method was as a participant in a study by Frank and Pharo (2015). The method struck me as a pragmatic and practical strategy for collecting solution-oriented opinions on a subject or problem.

My design can be described as Delphi-inspired by a multiphase-design, mixed-methods study. Creswell and Plano Clark describe the multiphase-design, mixed-methods study as an examination of a problem or topic ‘through an iteration of connected quantitative and qualitative studies that are sequentially aligned, with each new approach building on what was learned previously to address a central program objective’ (2018, p. 100). Similarly, the Delphi method contains different phases, each built on the results from the previous phase. In a classic Delphi study, questionnaires are used in all phases (Gupta & Clarke, 1996; Ziglio, 1996). The benefit of combining qualitative and quantitative data in a Delphi study is the possibility ‘to create scenarios that give concrete numbers’ (Tapio et al., 2011, p. 1623) while having rich illustrative content. Tapio and colleagues describe several Delphi studies combining qualitative and quantitative methods for data collection and how qualitative data makes the comparison of scenarios easier, while the richness of quantitative data makes the scenarios more meaningful: ‘Quantification provokes new aspects of qualitative arguments and vice versa’ (2011, p. 1627).

I have chosen to use interviews in the first and third rounds and a questionnaire in the second round. Interviewing helps identify potential misunderstandings or presumptions at an early stage and allows a broad exploration of a topic in the first round. It would have been possible to conduct the study merely with interviews; the questionnaires, however, provide comparable data highlighting agreements and differences, thus making comparisons easier. The questionnaire in the second phase was developed based on an analysis of the first-round interviews and, therefore, based on concrete input from the participants. At the same time, it was possible to get the participants' feedback on the results from the questionnaire in the concluding phase. Tapio et al. (2011) highlight how the 'Delphi process should be open to questions and topics that emerge in the process and possibly lead to a new understanding of the phenomenon under study' and recommend that only a few of the qualitative questions from the first phase be brought into the quantitative phase. This selection of topics or trends to follow up on in subsequent rounds is described later in this chapter.

4.2 Participants

In a Delphi study, the selection of experts, which from now on will be referred to as participants, must be carefully executed. Ziglio highlights how the 'selection of appropriate experts' (1996, p. 14) must not be a matter of personal preference or availability but must follow a procedure governed by explicit criteria. He further emphasizes knowledge and engagement, capacity and willingness and sufficient time as the three core requirements for the selection of participants (Ziglio, 1996, p. 14). Also, Lund recommends selection based on employment in relevant positions and scholarly publications on topics relevant to the research (2020). In the current study, employment in relevant positions has been the primary criterion used to select participants.

The study is limited to expert stakeholders in Norway, this is one of the many pragmatic choices made to ensure feasibility. Limiting the study to a national context, however, is a natural administrational level in terms of responsibility, funding, and policy. In Europe, there is a tradition of giving responsibility for access and use of research data to national data organizations (Eschenfelder & Shankar, 2020). The division of responsibility for data according to national boundaries is known as ‘the fishing zone agreement’ (Eschenfelder & Shankar, 2020). This territorial division of responsibility for data, like that of fishing zones (thereby the name), has long been the standard in European countries but with close collaboration between the data organizations. According to Eschenfelder and Shankar (2020), the expectations from the data organizations regarding their responsibility for the ‘fishing zone’ extend beyond the caretaking of sensitive data; they also perceive themselves as national service providers with an obligation towards researchers in their countries to provide data services (Eschenfelder & Shankar, 2020).

Researchers in Norway also take part in a multinational community, and Norway is part of Europe. One example is how the infrastructure providers participating in the present study are also part of the network of data organizations investigated by Eschenfelder and Shankar (2020). This illustrates that the word ʽnational’ does not mean detached from the global research communities. Researchers in Norway collaborate with researchers all over the world.

To investigate the division of roles, tasks, and expertise, it was important to have participants from both national service providers and universities. Some issues are solved locally, some are solved nationally, and sometimes, locally developed services become national. Researchers at four universities, Bergen, Oslo, Trondheim and Tromsø (collectively, BOTT) were selected as focal points for this study. The BOTT universities are the oldest in Norway and have a long history of infrastructural and administrative collaboration while also competing and collaborating when it comes to research and funding. All have a high profile of research across a range of academic disciplines, making them rich sites for this study.

Four categories of participants were included in the study:
1) The invited policymakers (PO) were members of the *Research Council of Norway* and the *Norwegian Ministry of Research and Education*, who work with policies for research data sharing and the *Vice-Rectors for Research from the BOTT universities*.

2) Infrastructure providers (IN) were invited from three different national infrastructure and archival services. The data archive at the *Norwegian Centre for Research Data (NSD)*, is historically the archive for social science data dating back to 1971, and part of Consortium of European Social Science Data Archives (CESSDA). NSD has two main services, the data

archive, and the privacy protection service, for the purpose of this study the focus was on the archive. However as elaborated on in Article 3 (L. Kvale & Pharo, 2021) the privacy protection services delivered by NSD are also central for all researchers in Norway working with personal data. The *NIRD archive* is delivered by UNINETT-Sigma2, the high-performance computing collaboration amongst the BOTT universities. Sigma2 is jointly funded by the Research council of Norway and BOTT. Sigma2 coordinates Norway's participation in international collaborations on e-infrastructure such as EUDAT. The third infrastructure provider is the *BIBSYS* consortium which is a national provider of library systems for research and higher education sector, as described in the time-frame BIBSYS became part of the Directorate for ICT and Joint Services in Higher Education and Research (Unit) in 2018. These three infrastructure providers all deliver research data infrastructure at a national level but also target different parts and pieces of the whole research data infrastructure.

3) Research support staff were recruited to cover a range of research support services (library (L), IT and research office (RO)) who work with research data support. The responsibility and organizations of research data services are slightly different at the different universities: In two universities, the library has a primary responsibility but collaborates with IT on technical services and the research office on policy developments. One university has a joint responsibility for research data support but a division of tasks. At the fourth university, there was a collaboration established between the services, but no clear distribution of responsibilities has yet been established. These experts on data management in their respective universities were identified through networking, and often via colleagues forwarding requests within the university to recruit from the different research support services.

4) Researchers were selected based on their position as principal investigators (PI) on projects receiving grants from the EU in 2017 with BOTT as an affiliation. Project funding from the EU required the submission of data management plans, the researchers were therefore expected to have reflected to a certain extent over data management. The researchers were identified through the CORDIS database (European Commission, 2020) and from different disciplinary backgrounds (humanities, sciences and social sciences). During the analysis of the first-round interviews, it became clear that the way the researchers collaborated with other researchers shaped how they saw data management–related services.

in their descriptions of their work with research data, differences were found between the researchers describing a collaboration on the data collection and/or data analysis versus researchers who did not have shared access to research data with research partners during the research process. Researchers were grouped according to how they worked with other researchers on research data group (RG) or individually (RI). This division was carried out after the first phase of data collection and analysis.

Of the 48 invitations sent, 24 participants took part in the study (Table 2), 24 participants being average for Delphi studies within library and information studies (Lund, 2020). The participants showed an interest in the subject and a commitment to participating in the project. All 24 took part in all three rounds.

Role/stakeholder category	**Invited/accepted**	**Participant code**			
RI Researchers working individually	25/8*	RIZ	RIJ	RIL	RIB
RG Researchers working in groups		RGV	RGD	RGA	RGW
PO Policymakers	6/3	POU	POS	POK	
IN Infrastructure providers	5/3	INH	INO	INR	
IT Research support IT	4/3	ITE	ITY	ITI	
RO Research support, Research office	4/3	ROC	ROX	ROT	
L Research support, Library	4/4	LM	LP	LG	LN
Total	48/24				

*Researchers were invited as one group and split into the two groups RI and RG after the first interviews.

Table 2. The participant group

One of the researchers who took part in the phase 1 interviews was omitted from further participation in the study as she did not work with data in her research. Focusing on research data, I was somewhat unprepared for the ‘no-data’ researchers where data remains immaterial. Initially, I wanted to include all the researchers who agreed to participate, to cover the broadest spectrum of research and research data. By making EU funding a criteria, it was possible to invite researchers across domains, with different experiences in research data sharing. During the first interview, however, it became clear that one researcher had no experience with research data. For this study experience with research data being the criteria for participation, this researcher was therefore excluded from the study. This researcher is not included in Table 1 because the first interview was not completed. For more details, please consult transcript RIF1.md and the affiliated readme file in the data deposited in Zenodo (Kvale, 2021). It is important to acknowledge that there are ‘no-data’ researchers, where the research[3] does not rely on representation of objects: immaterial research. The two immaterial researchers encountered in this study were researcher RIF who worked within informatics and theoretical mathematics, and philosopher RIJ. RIJ had extensive experience from ethical boards along with a strong research interest in personal privacy, personal data, and information ethics; however, she did not use research data in her own research.

4.3 The three phases

The study had three phases of data collection (Fig. 2). Phase one, the exploration phase, took place in January and February 2018. Phase two, the evaluation phase, was conducted in August and September of the same year. The third concluding phase was conducted in March and April 2019. The time lapse between the different rounds of data collection allowed time for the study to follow the changes and developments taking place in the *knowledge infrastructure*. The different phases and the time between them then become an advantage because they allowed the investigator to identify and follow transformations. The analysis of each phase was used as the basis for the development of the next phase and for preparation of a concluding integrated analysis after the final round looking at all the data collected (Creswell & Plano Clark, 2018). Two or three rounds are most commonly used in Delphi studies (Ju, 2013). Three rounds were found suitable for this study as this allowed for corrections and verification, while limiting the number of rounds to avoid overloading the participants. In each round, the participants were sent the questions before the data collection.

---

[3] Reminder that the use of the term research in this thesis refers to both science and scholarship.

Several of the participants had not looked at the questions before the interviews; others had read through, while some had prepared their responses.

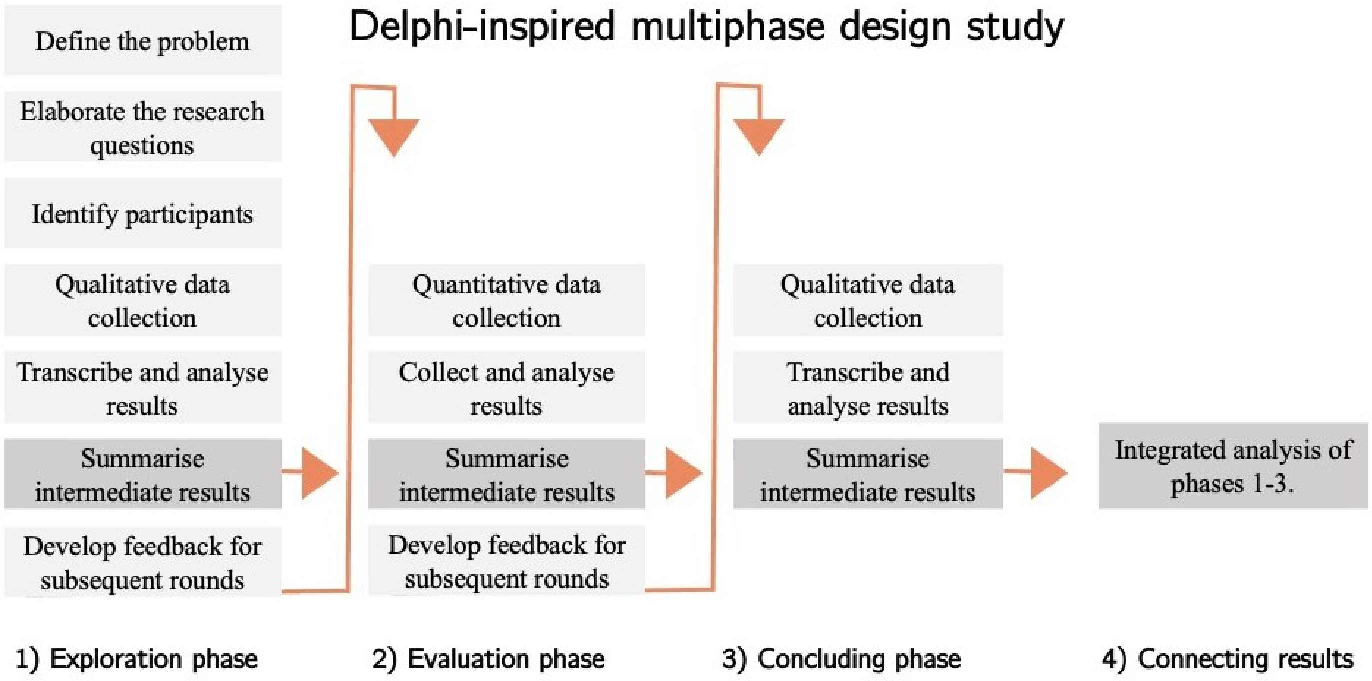


Figure 2. The research design (L. Kvale & Pharo, 2021)

4.3.1 The exploration phase

Ziglio uses the term ‘exploration phase’ (1996, p. 9) to describe the first phase of a Delphi study. Through open-ended interviews that were conducted face to face, the opinions and expectations of the different stakeholders were explored. The participants were asked, among other things, how they work with research data, what challenges they encounter in their work with research data and how they would describe their ideal system for data sharing. The focus in the interview was to discover topics within research data management and sharing that caught the attention of the different stakeholders and for which they were looking for clarifications. Most of the questions were common across the interviews. Exceptions were experience with data journals, which only the researcher was asked about. None of the researchers had experience with a data journal; subsequently this line of inquiry that was dropped in the next round. Whereas thoughts on the role of UNIT and the policy document of 2017 were brought up with policy makers, research support and infrastructure providers, they were not discussed with the researchers. Appendix D shows the complete interview guide followed in all the 24 interviews. The first interviews were all planned to be conducted in person; however, because of bad winter weather, two interviews with research support staff were conducted over Skype. The interviews were approximately one hour long, resulting in about 24 hours of audio. The interviews were transcribed, amounting to 216 transcribed pages. The interview transcripts were then sent back to the participants for an accuracy check.

A qualitative analysis of the exploration phase was performed using NVivo. The qualitative analysis was based on meaning interpreting according to the hermeneutic tradition of text interpretation (S. Kvale, 2007). During the transcription, codes in the form of themes and concepts brought up throughout the interviews were written down. These codes were then structured hierarchically as overall themes and more detailed codes in Nvivo as a preliminary qualitative codebook. All the transcripts were then read twice and marked with codes with the associated topic. The text was then extracted according to different topics, re-read and annotated with condensations of the central themes. Examples of this process including qualitative codes, coded text and the condensation of themes are found in table 3, while figure 3 illustrates the different concepts and approaches used in the coding in the exploration phase, highlighting the topics selected for further inquiry. The complete qualitative codebook used in the exploration phase, and the final version used for coding in the concluding phase and the integrated analysis is found in appendix G.

Figure 3 Extract of preliminary qualitative code book highlighting topics which were explored further in the next phases (dark grey), and presented in the articles (orange)

Examples of coding from the exploration phase

1	Roles, responsibilities, and collaboration / Data steward
2	*When we uploaded something to Norwegian Centre for Research Data the organization was very much focused on how the technical solutions would work, but they did not understand the content. Then you end up storing things in a nice technical way, with high quality metadata etc. but still the data are meaningless to a researcher on the other side. […] Therefore~~,~~ professionalization of data management personnel, is a role that will appear, a person understanding the data, but also with an understanding of how to structure them. It is my belief that much of the library function will be directed towards data management in the future. (RGA1)*
3	Need for specialized data stewards who combine the domain expertise with curation expertise
1	Data management plan Roles, responsibilities, and collaboration / Data steward
2	*It is crucial to have a data steward, a new professional profile that supports the researcher in the process of defining this data management plan. At the end, the creator of the data management plan will be the end user, but they cannot do it alone, they need guidance from a person who has been educated to do this, who can combine the ethical aspect and the FAIR principles paradigm, and the technical challenges. (INH1)*
3	Need for specialized data stewards who combine the domain expertise with curation expertise
1	Ethics / Privacy
2	*When we get to Google and platforms, and the fact that anonymity no longer exists, the response must be embedded privacy, and this is a problem we as a society simply must solve. We need to approach this in one way or another, at least we cannot just stick our fingers in the ears and believe that it will go away. The best thing to do is to spend research funding on the development of good infrastructure. (POK1)*
3	In the "Google age" there is no such thing as anonymity. Need to invest in the development of platforms with privacy protection embedded.
1	Advantages / Better research / Transparency Practical experience with data

	Roles, responsibilities, and collaboration / Own role
2	*I have collected lots of data in my time, and still do. I wear both the researcher's hat and the librarian's hat, and I am of course very pro data sharing. Because this is a matter of getting access. If I use my own research experience: If you read an article you are interested in, and you think that the data described are interesting and might be relevant for you, then it is nice to get access. (LM1)*
3	Mixed identity, build on research experience in the work in the library Data sharing for transparency and the advancement of research

Table 3. Examples of themes and subthemes (1) with coded text (2), and meaning condensation (3), exploration phase.

4.3.2 The evaluation phase

The second phase of a Delphi study is designed to evaluate views on issues investigated. Based on the qualitative coding of the transcripts in the exploration phase, six themes were selected to follow up on in the subsequent round of the study (Table 4). The participants answered a questionnaire where they were asked to share their opinions of different statements. These statements were based on the analysis and "condensation of central theme" (Table 3) in the first round of interviews. A pre-testing of the questionnaire to avoid ambiguous questions was performed. The questionnaire was created and sent out electronically. The questionnaire is found in appendix F. For the two first questions, the respondents were asked to rate their level of agreement or disagreement with several statements using a six-point Likert scale, including both neutral and no opinion options. Questions three and four explored the stakeholders and function of a data management plan. The participants were asked to select five reasons to make a DMP, ranging from one to five based on importance (Q4), and to list the different stakeholders of the DMP, also by importance[4] (Q3). Question five (Q5) asked the participants to estimate the cost of data management and how these costs should be covered. In question six (Q6), the participants were asked to elaborate on data stewards by drawing up their ideal, fictive data person. This

---

[4] The responses to question four regarding the goal of the data management plan showed that the descriptions on how to answer this question had been unclear; approximately half of the participants answered unintended. To obtain coherent responses, I decided to repeat the question during the interviews.

question was inspired by user experience (UX) methods and the use of fictive personas to illustrate user and users' needs (Hartson & Pyla, 2012). In question seven (Q7), the participants were asked to rate the importance of the different functions of infrastructure and built on the suggestions from the first interviews regarding an ideal infrastructure solution for research data. In question eight (Q8), the participants were asked for suggestions on how the infrastructure can facilitate transparency and verification in research. In question nine (Q9), the respondents were encouraged to add additional comments and suggestions.

Theme	**Question**
Motivation for data sharing	Q1 a-h
Privacy and ethical challenges	Q2 a-j
Data management plan	Q3 and Q4 a-m
Costs and cost recovery	Q5 a) and b)
Data person/data steward	Q6 a-h
Infrastructural functions	Q7 a-t and Q8

Table 4. Overview of teams explored in the questionnaire

In the evaluation phase, three different analytical approaches were applied on different types of questions. For open-ended question responses, content analysis was conducted. The different responses were gathered and grouped thematically, see figure 4 [5]. The questions asking for level of agreement using a Likert scale or for ranking importance of roles or functions were analysed quantitatively; the low number of participants provided indicative tendencies rather than statistical validities. Responses were grouped according to stakeholder category to identify potential conflicting views on the issues, see figure 5 for example. For the question regarding relation, namely for whom the data management plan is intended, the data visualization software Gephi was used. Gephi is built on visual analytics and highlights relational patterns in the responses (Loth et al., 2019). Figure 6 is a visualization of how the participants pointed at different stakeholders of the DMP. This visual analysis of relations

---

[5] Another example is found in table 2 Data stewardship skills (L. Kvale, 2021b)

shows how different stakeholder groups see the DMP as having a function for multiple people in different roles (L. Kvale & Pharo, 2021).

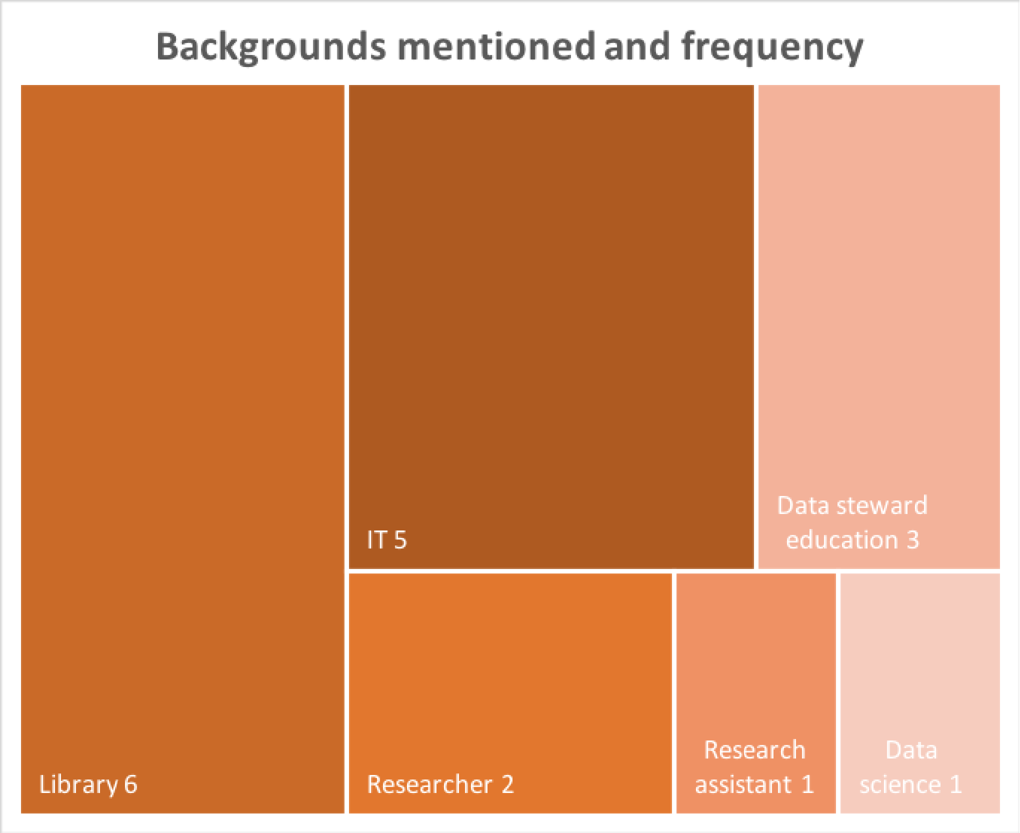


Figure 4 Preferred background for data stewards (L. Kvale, 2021b)

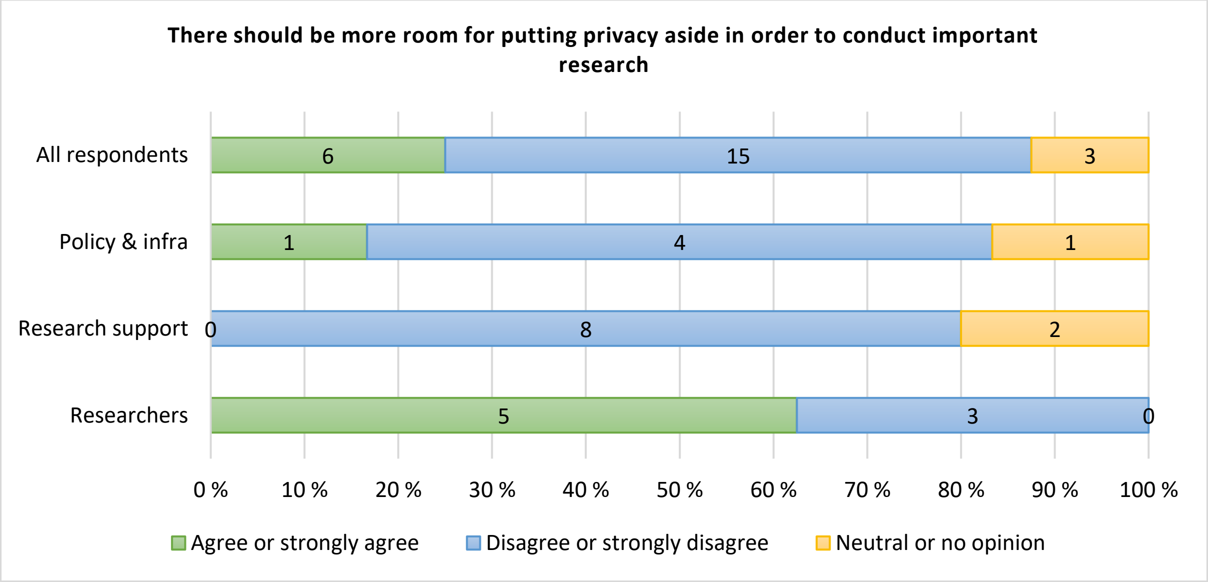


Figure 5 Comparison of responses from the different stakeholder groups (L. Kvale & Darch, in press)

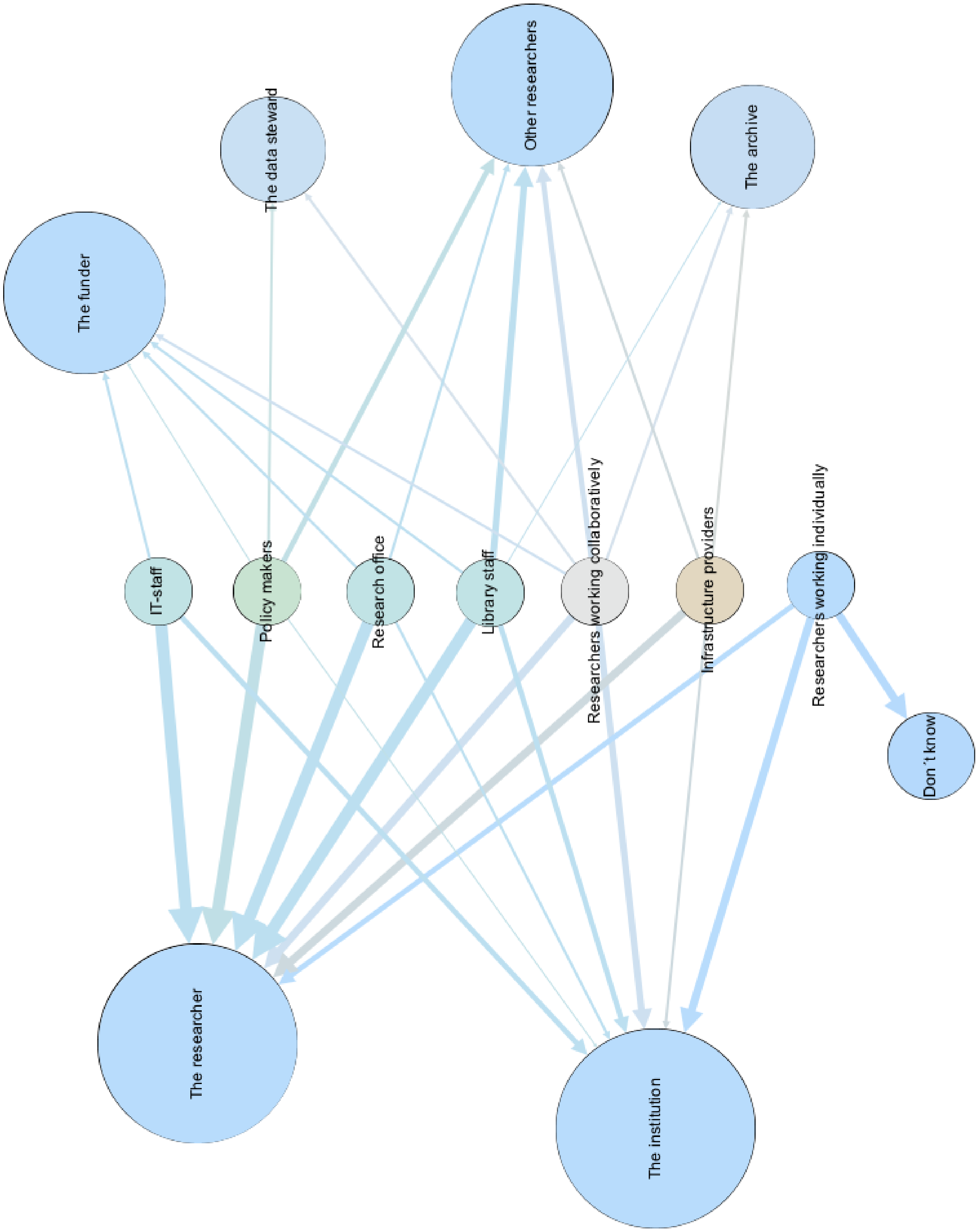


Figure 6. Relational visualization made with Gephi: Who is the data management plan primarily for? (L. Kvale & Pharo, 2021)

4.3.3 The concluding phase

The third round involved interviews that were approximately half an hour long, resulting in 12 hours of recordings and 98 pages of transcripts. The interviews were performed partly via Skype and partly in person. The interview guide was structured according to four broad themes: data stewards, data management plans, ethics including personal privacy and ethical aspects of data sharing, and changes or developments regarding both research data and open science in general (the interview guide is found in appendix E). In addition, the participants were encouraged to give feedback on the research design and method. In the interviews, the participants were presented with preliminary results (se appendix E for details) and asked to offer additional comments and suggestions. Three of the themes; persona, data management plan and personal privacy are covered in the three articles included in this dissertation.

Before each interview, I prepared by reading through the first interview. Notes were taken before the interviews, after the interviews and during transcription. The preliminary codebook from the exploration phase, was developed into a final codebook. In the final codebook, codes were grouped according to the themes explored in the final interview with qualifiers describing if it was experiences or reflections that were shared, and code terms related to the subject (Figure 7 and Table 5). The transcripts were coded according to the tradition of meaning interpretation and meaning condensation (S. Kvale, 2007). Qualitative coding of the transcripts and extraction of text for condensation of themes were conducted using XML and queries using Python script. A transition to XML and Python was the result of extensive litigation with NVivo regarding lacking interoperability between operating systems and to open formats. The findings were summarized based on the condensation of themes and used as a basis for the integrated analysis.

Examples of coding from the concluding phase

1	Cultural change /Experience /Data sharing
2	*Today it is relatively standard that when something is published you should upload the data, unless you work with confidential data, and you upload read-me files, code, and descriptions of how the data are used. But as mentioned, until now there has been no quality control on this; that is something left to be done by the journal [where RIB is editor], and I believe other journals are doing the same thing. (RIB2)*

3	Awareness and quality control regarding data submitted as appendices because the journal is hiring a data manager, a development from first to second interview.
1	Persona /Reflections /Suggestions Ethics /Personal privacy /Reflections
2	*There are two good arguments for having a centralized role; one is the law, the law is the same and it is complicated enough to make the lawyers agree and if you are going to have and IT system in the bottom, you need one interpretation of the law. That is why it is very inconvenient to leave it to the lawyers at the universities to govern, then it's hell developing this in an IT system (INO2)*
3	Argue for centralizing research data services nationally; from a developer's perspective, one interpretation of the law is practical.
1	Technical development /New infrastructure
2	*It is my impression that they [UNIT] are getting a hold on this. What's nice is that the director has worked in other sectors before and reports that our sector is ahead on things because so many things have been solved bottom-up and according to the tradition of pooling the resources. He said this is admirable and very effective, so that now they have a good base to build on. (POK2)*
3	Bottom-up initiatives and collaborative projects created as responses to needs are put together as the base for the new infrastructure organization.
1	KI development /Organization /New infrastructure
2	*I still find it troublesome that the people delivering services are also going to be a directorate controlling other organizations. My impression is based on the work done with the plan of action, and that they struggle to find the right level regarding what they are going to pursue, what the universities will pursue and what should be done elsewhere. So I hope they are able to take on the role of coordinating. (LG2)*
3	Concerns regarding unclear mandates of the new infrastructure organization as both directory controlling and prioritizing services and provider of services.
1	Persona /reflections /experience
2	*I have a case you could use where all the positions are filled now, you could probably take a look at the CV of the people who got the jobs and compare a bit. What I know is*

	that they decided on two hard-core programmers and then two more outgoing, more of a librarian type rather than only the programming skills. (IT2I)
3	IT is hiring data stewards with complementing skill sets to cover a range of needs
1	Ethics /Public trust /Research essere
2	*What is the value of people trusting research? Well, two things: it is good for people and good for research. For people because when decisions are made, placing trust in research, and then in particular in a democracy, but in any government, but in particular in a democracy, is that people know and have a justified trust in high quality research – that is a «sine qua non» to be able to make fairly intelligent decisions and right choices. In this way the public trust in research is crucial for the citizens. And then it is the same for researcher – to be able to conduct research, have the access, have access to data, have funding, then the researchers must have the trust of the public. It is such an important matter, more like a means, this would require further analysis. (RIL2)*
3	Society is built on a trust in research, and research would not exist without trust from society. Why trust in research is important is beyond my scope and near the core of what research is.

Table 5. Examples of themes and subthemes (1) with coded text (2), and meaning condensation (3), concluding phase

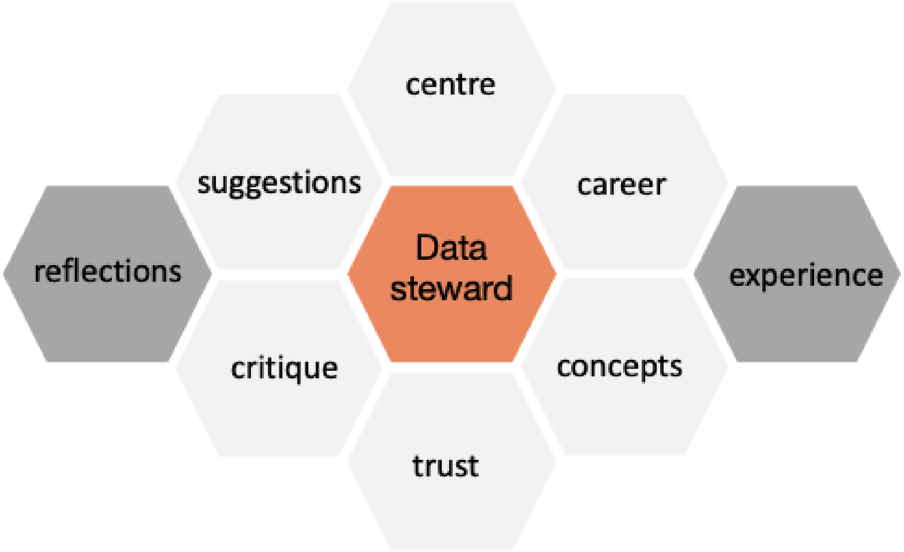


Figure 7 Examples of theme, codes and qualificator used in the qualitative coding in the concluding phase and the integrated analysis

Code	Description
Reflections	Sharing of thoughts or reflections on the subject
Practical experience	Referring to own experience on the subject
Consent	Thought or experiences regarding the use of consent
Public trust	Thought or experiences regarding public trust in research
Cost profit	Thought or experiences regarding the cost and profit aspect of data archiving
Integrity	Thought or experiences regarding research integrity
Research essere	Thought or experiences regarding the ethos of research, what research are or strive to be
Research ethics	Thought or experiences regarding research ethics
Personal privacy	Thought or experiences regarding aspect of privacy protection
Internationalization	Thought or experiences regarding internationalization in research and data sharing
Embedded privacy	Thought or experiences regarding the usage of embedded privacy in privacy protection
Privacy vs. research	Thought or experiences regarding the balancing of the respect for privacy with conducting high quality research
Concepts	Concepts or to explore further are introduced by the participant

Table 6 Qualitative codes within the ethics theme (L. Kvale & Darch, in press)

4.3.4 Integrated analysis

After first analysing the results from the third round, a separate integrated analysis modelled after convergent design (Creswell & Plano Clark, 2018, p. 222) was performed. The purpose of the integration of the different results in the integrated analysis was to expand the understanding and develop an integrated and comprehensive result (Creswell & Plano Clark, 2018). In the integrated analysis, the themes in the codebook developed for the third phase (Table 5 and Figure 5) were used in reading through transcripts and form the first phase and the qualitative responses from the survey again looking for common concepts across the results. This allowed for an integrated extraction of results from approximately 36 hours of interviews. The responses from the survey were used as additional quantitative context. The quantitative and qualitative results were compared, looking for ways in which "the results confirmed, disconfirmed, or complemented each other" (Creswell & Plano Clark, 2018, p. 222). The results presented in the three articles are from all three rounds of data collection.

4.4 Ethical considerations

This part describes reflections and considerations made regarding data collection, data sharing and language.

4.4.1 Data collection and data sharing

Approval to collect personal data, was applied for and given by the NSD privacy protection in 2017 (appendix A), initially with the idea of anonymizing the data material. An informational letter about the project and a consent form to participate in the study (appendix B) were sent to the participants and signed prior to the first interviews, late in 2017. Based on this consent, all the collected data are part of the study and included in the three articles.

The limitation on consent can be challenging to balance with sharing of research data. I have experienced this friction when creating consent forms for the participants of the current study. The solution for tis was to supplement the I initial consent regarding participation in the study, with a second consent regarding data sharing after the data collection was completed.

In the first consent form, the participants were also asked if they agreed to share anonymized data when the project was finished. As the first round of interviews were transcribed, with directly identifiable information removed, it became evident that the data were not anonymous. It did not make sense to aim at anonymizing the interviews within a small community of domain experts . One would have to remove all the context, such as the profession and experience shared by the participants. When reading the transcripts, I would recognize several participants based on the language and metaphors they used. The Delphi method itself focuses on the importance of anonymity for participants to speak freely (Gupta & Clarke, 1996; Ziglio, 1996, p. 22). However, because of the combination of qualitative material and small-scale conditions, several of the participants could be identifiable because of their functions. While there are clear advantages of anonymity, 'It is not, however, always an advantage to conduct a Delphi exercise under absolute anonymity, and there can be clear advantages to reducing or eliminating anonymity' (Rotondi and Gustafsen 1996, p. 39). The debate regarding anonymity shows that it is hard to achieve, and not always the appropriate approach to privacy (Barocas & Nissenbaum, 2014). According to Nissenbaum and Barocas, it is important to discuss what privacy is to be protected and the context rather than focusing on finding solutions to ensure anonymity.

To balance sharing and privacy without focusing on anonymity, informing participants, and allowing them to review and select what to share was chosen as strategy. After the data collection was completed, the participants were sent transcripts from both interviews, their responses to the survey and an individual link to a second consent form regarding data sharing (appendix C). A drafted version of this consent was shared with privacy advisors at NSD for input on how the research participants could be asked to share non-anonymous data.

The participants were each given the transcripts from their interviews for review and asked what they would give permission to publish as part of the research data package in a relevant repository. Sequences that the participants wanted removed before the data were published were highlighted. In the two articles using quotes that participants did not want included in the data package, care has been taken to protect the participants' identity. In Kvale and Pharo (L. Kvale & Pharo, 2021), quotations are used only to refer to a participant group. While in the third article by Kvale and Darch (L. Kvale & Darch, in press), the participant codes have been scrambled and changed into numbers. Both strategies aim to protect the identity of participants, while clearly conveying their opinions and input on the issue investigated. Consenting to participation in a research project and consenting to open publishing of interview transcripts are two different things, and by asking for two different consents, this distinction has also been made explicit to the participants.

In total, 21 participants gave permission to publish all or parts of the data fully in an open repository. Despite having received several reminders, three participants did not respond to the consent form regarding data sharing (appendix C), and their data are not published in the data packaged. In the consent form regarding data sharing, the following phrasing was used:

> *I understand that despite removal of directly identifiable information such as name and workplace, it could still be possible to identify me based on the information in the interviews. (Appendix C)*

According to the principle of 'the right to be forgotten' research participants have the right to withdraw their consent at any time (GDPR, 2016; The Norwegian Personal Data Act, 2018). A clear and explicit consent for data sharing does however make open sharing of human subjects research data possible (The Research Council of Norway, 2021). Data from the study, including interview guides, questionnaires, consent forms, approval and so forth are available on Zenodo (L. Kvale, 2021a).

In the texts, I have chosen to refer to all the participants as she or her to make identification more difficult. Therefore, *she* and *her* have been pragmatically (and slightly feministically) preferred as the neutral form. There is a balanced participation from both genders in the study, and the gender of the participants is not of relevance in the results.

4.4.2 Language

The interview guides, questionnaire and consent forms were distributed in both English and Norwegian. Most participants responded in Norwegian; however, none of the participants were native English speakers, and some gave part of the responses in English and parts in Norwegian. For the purpose of the publication of articles, Norwegian quotes have been translated into English. The complete transcripts are in the language in which the interview was conducted. The categories used for the coding were initially developed in Norwegian but were translated for the purpose of sharing.

Chapter 5. Article summaries

The following are brief overviews of each of the three articles included in the present thesis and their publication status.

When selecting journals, two criteria have held priority: relevance and open access. *College and Research Libraries, International Journal of Digital Curation and Information Research* were selected based on their relevance to the different papers. All three journals are diamond open access journals: journals that publish without charging authors and readers (Bosman et al., 2021).

Table 7 summarizes the research questions and main finding of each article. All three articles are based on the theoretical background described in Chapter 2 and use data from the integrated analysis from the Delphi study described in Chapter 5. Each article addresses a separate topic, with independent research questions, the thread is the people aiming at translating needs between stakeholder groups. This is addressed in chapter 6 in the light of the research questions for this compilation.

Article 1 *Using Personas to Visualize the Need for Data Stewardship* (A1) was published in *College and Research Libraries* (C&RL) volume 82/3. C&RL is the official scholarly research journal of the Association of College & Research Libraries, a division of the American Library Association. C&RL is the top-ranked open access journal within library and information studies according to the Norwegian publication registry.

Article 2 *Understanding the Data Management Plan as a Boundary Object from a Multi-Stakeholder Perspective* (A2) was published in the *International Journal for Digital Curation* (IJDC) volume 16/1. The article was first accepted as a conference paper at the International Conference for Digital Curation (IDCC) in 2020, a preprint version is published in the proceedings, IJDC volume 15, while the revised and final version is published in volume 16. IJDC is published by the University of Edinburgh for the Digital Curation Centre. Article 2 is co-authored with supervisor Professor Nils Pharo.

Article 3 *Privacy protection throughout the research data lifecycle* (A3) has been submitted to *Information Research* (IR). IR was established in 1995 and is one of the longest-running OA journals within library and information studies; it is published by the University of Borås and hosted by Lund University Libraries. A3 was first submitted as a poster at IConference 2020, where it was also a nominee for best poster (L. Kvale & Darch, 2020). Article 3 is co-authored with co-supervisor Associate Professor Peter T. Darch.

Title, journal	Knowledge gaps	Main findings
Using Personas to Visualize the Need for Data Stewardship College and Research Libraries, C&RL	Who are the data stewards in the universities? a. What roles should data stewards play? b. What services should data stewards provide as part of these roles? c. What skills do data stewards need to carry out these services?	A1 presents a set of fictional personas for research data support that were developed to visualize the different types of research data management support requested in the universities. The analysis and results are based on experience and demands by the different stakeholders involved in the study.
Understanding the Data Management Plan as a Boundary Object from a Multi-Stakeholder Perspective International Journal of Digital Curation, IJDC	1.What perspectives on DMPs are held by different stakeholder groups? 2. How do these perspectives help, or hinder DMPs as tools for supporting data management?	Different perspectives among stakeholders of what the DMP should be, and different practice approaches, yet they present a common goal. These differences needs to be considered if the DMP is to work as a document translating between different groups supporting the longevity of data.
Personal Privacy Throughout the Research Data Lifecycle Information Research, IR	1. What perspectives on privacy are held by stakeholders in the curation of research data on human subjects 2. How do stakeholders' perspectives on privacy shape their data curation actions?	Researchers, research data management support staff and data organisations must reconcile divergent motivations and resolve tensions throughout the data life cycle. Through dialogue and negotiation, all stakeholders involved in data sharing should aim to respect the research subjects' own understandings of privacy.

Table 7. Overview of the three articles in the thesis

Using Personas to Visualize the Need for Data Stewardship (A1)

The article identifies four primary challenges for providing data stewardship at universities: 1) Placement of responsibility: researchers must retain their responsibility for data throughout the research cycle. When depositing to a data archive, responsibility can be transferred. 2) Communication: lines of communication between support levels must be established to avoid closed subcultures and exchange best practices between domains. 3) Knowledge of data and methods: knowledge of research is essential; however, there is also a need for local and specialized expertise within an increasing number of domains. 4) Joint research support effort: research data management requires several different types of expertise that are traditionally spread among different research support departments at universities. The creation of a general research data support team or centre with connection to the research office, university IT department and library is crucial to cover all aspects of data management.

Based on experience and requests from experts in different areas of data management involved in the Delphi study, three different fictional personas for data stewardship were developed, along with a research data support centre proposed as a competency hub for data stewards and a go-to place for researchers. The research data personas were developed to illustrate the range of skills required to support data management within universities and were inspired by user experience methodology.

Outreach, education and problem-solving are only some of the keys to the creation of a functional service for data management. One solution can never fit all, and although a general team will be able to solve and support a wide range of issues, many larger research communities need dedicated staff with specific knowledge of the issues and concerns that are relevant for their research data. Although data management is gradually becoming a current practice within several data-intensive communities, it is also needed among researchers who are producing and collecting small heterogeneous datasets, which are referred to as the long tail of research data.

Understanding the Data Management Plan as a Boundary Object from a Multi-Stakeholder Perspective (A2)

A2 explores how the data management plan (DMP) may perform the role of a boundary object for the different stakeholders involved in research data sharing.

The stakeholders of this study agreed that a DMP is written to improve data management by making researchers plan for sharing their data internally within research groups and externally (FAIR) by creating procedures for documentation and collection at an early stage. Although sharing a common goal, the stakeholders had different perspectives on what the DMP is. Four perspectives reflect the stakeholders' different views: 1) the curating and fulfilling requirements perspective; 2) the sharing and open science perspective; 3) the stewardship perspective; and 4) the protocol and procedures perspective.

The tension between different perspectives, the research reality and the higher goals can be resolved by a common understanding of the DMP as a document that is not the product of a consensus but an everyday translation between worlds and communities. The need for consensus as a basis for cooperation is a common misunderstanding. By introducing the concept and theories of boundary objects, it is possible to explain how cooperation can continue unproblematically without a consensus.

By writing a DMP, researchers plan for their data to move from collection through analysis and to sharing, as was agreed upon in the goal of the DMP. In this sense, DMP creation facilitates translation between the different worlds of the different stakeholders. Disagreement on the degree of standardization and the degree of automation, however, are two obstacles that were identified. The struggle for the different stakeholders is to find the right balance in the guidance and automated decision making, leaving leeway for the creator of the DMP to describe the right level of detail for the data. By resolving this, the DMP can represent a translation corresponding with that of a standardized form, hence making it a document to be understood and interpreted in the different worlds independently of disciplines, institutions, or national boundaries.

Privacy protection throughout the research data lifecycle (A3)

The findings show multiple tensions between maintaining research subjects' right to privacy and advancing research through data sharing. This paper identifies and analyses three particular sources of tension: 1) Maintaining trust with research participants; 2) Managing divergent views of privacy in international and intercultural research collaborations; and 3) Interpreting and applying policy. The divergent motivations and perspectives on privacy held by different stakeholders complicate these tensions. Researchers, research data management support staff, and data organizations must resolve these motivations and tensions throughout the data lifecycle, from collection to archiving and eventual sharing. Through dialogue and negotiation, all stakeholders involved in data sharing should aim to respect the research subjects' own understandings of privacy.

Personal privacy protection in research involves respecting research participants, requiring awareness of roles, attributes, and transmission principles. In digital research, multiple stakeholders are involved in data management, all of whom must demonstrate sensitivity towards data privacy and research participants. If and when data sharing is to take place, respecting the research participants and their perception of what information is sensitive and private must be prioritized.

The requirements of open research and international research collaborations make balancing personal privacy with data sharing a complex task for researchers. Providing expertise and guidance on how to best balance these requirements is part of research support and something Research Data Management support should offer. To facilitate making the sharing of data 'as open as possible and as closed as necessary', we must acknowledge that different stakeholders in data sharing have different perspectives on how personal privacy and data sharing should be balanced. Increasing quality and transparency of research must be the primary motivation for the sharing and reuse of data and must be carefully balanced with the privacy of research participants when human subjects are involved.

Chapter 6. Discussion

The three articles address different challenges in research data management; the common thread is the focus on expertise, on people with the ability to interpret and translate the needs between stakeholder groups. To facilitate collaboration on data sharing amongst internal and external partners is key. As highlighted by one of the participants, working together is necessary for success in providing research data services in universities (L. Kvale, 2021b).

> *I believe it is important with such a holy trinity that the IT- department, the library and the administrations [research office] could become if they work together. (RGA in Kvale 2021)*

This quote pinpoints the challenge of complex task of research data sharing. Multiple communities of practice within the university are required to collaborate with each other and with external partners to deliver the services required by the researcher and policymakers. People with knowledge and expertise was a recurrent topic throughout the interviews, resulting in the first article (L. Kvale, 2021b), specifically addressing data stewardship's roles and skills.

This chapter brings together findings presented in the three articles exploring the role of expertise, boundaries, interpretation and translation of knowledge and perspectives in research data sharing as a domain. In addition to findings previously presented in the three articles, some additional findings not previously presented are introduced to supplement and support the arguments of the discussion.

The discussion is structured in six sections: 6.1–6.3 address each of the three research questions: 6.1 argues that additional agendas risk obscuring the focus on research quality as the goal of data curation in interpretation and application of policy. Section 6.2 shows how a combination of practice, learning and recognition is needed to succeed in providing research data services. This is best achieved through maintaining active memberships in multiple communities. Section 6.3 addresses the dynamic development of standards which trough experience and application lead to new understanding. This development of standards facilitates knowledge transfer connecting the past with the future. Sections 6.4–6.6 address limitations of the study, the research design, and provide recommendations for future research.

6.1 Differences between stakeholders' perceptions and development in research data curation

This section addresses RQ 1: How and why do differences between stakeholders' perceptions affect their ability to collaborate in the work of research data curation?
Three examples of changing views towards aspects in the infrastructure are presented and discussed. The three sections address aspects where differences in the stakeholders' standpoints were detected between the two interview rounds, namely; 6.1.1 Researchers' awareness of data management; 6.1.2 Identities and services in the breakdown and build-up of organizations; and 6.1.3 Agendas embedded in the data management planning tools. Having a clear understanding of own contributions and expertise and common goals makes collaborating easier. The creation of new organizational entities and cultures can create new frictions between goals, tasks, power relations and expertise. The timeline (figure 1) in chapter 1.3, provides the context by visualizing the time scope of the study in which development and changes in approaches took place.

6.1.1 Researchers awareness of data management

Between the two interviews, a change was detected regarding awareness and expectations concerning reproducibility and research data management (L. Kvale & Pharo, 2021). Based on debates within their communities, the stakeholders describe increased interactional expertise developed in discussing the principle of open research amongst peers:

> *"It's evolving. Some groups have high data management standards, and you try to adapt to it. [...] in data [practice] we are influencing each other internationally"* (RGW in Kvale & Pharo, 2021).

Emerging standards of data management are developed and shared within communities of practice as described above. Such changes are not visible, they are part of the good research practice for the researchers and developed from the bottom-up. The literature supports this; participation and learning based on performing activities and engagement in communities are the main point of reference for best practice (Berman, 2017; Cooper & Springer, 2019; Darch et al., 2020; Thompson, 2017). Data management expertise is embedded in the practice of the different research communities (Cooper & Springer, 2019; Darch, 2018; Thompson, 2017). Following best practice and development within research groups with high data management standards is a path of learning and source for acquiring contributory expertise for both researchers aiming at improving own data practice and data support services.

In the final interviews, PlanS was highlighted when participants were asked if they experienced any cultural change towards open science. While PlanS has raised strong feelings among researchers regarding academic freedom (*Innspill Til Forskningsrådet Om Veilederen Til Plan S [Input to the Research Council Regarding the Guidelines for Plan S]*, 2019; Graver, 2018), it also inspired a debate regarding journal-quality research dissemination (Svarstad, 2019). The focus on quality control in publishing addressed in discussions in the aftermath of PlanS created an awareness regarding quality control on research data: *'Within many areas there has been an increased focus on reproducibility, and then it is important to actually have the data'* (RIB2 in Kvale & Pharo, 2021). RIB further described how a journal with which she is involved is now hiring a data editor or data steward.

> *You can submit these things [data submitted as an appendix to an article], and it can be complete nonsense, and no one checks it. But that is changing, for instance in [journal title], a leading journal in economics, where I am on the board. We are now hiring a data editor to quality check all data submitted as appendices. (RIB2)*

Increased awareness regarding data quality resulting in the hiring of a journal-specific data steward connects the tradition of the journal editing with data management through the goal of quality improvement. Investments in quality checks and curation of data illustrate an ongoing change regarding data and reproducibility awareness (Cooper & Springer, 2019; Darch et al., 2020).

In the final interview, the participants were asked directly whether they experience any changes regarding data sharing. Several researchers described changes in the perception of the refusal to provide data access to others.

> *Yes, I believe there is [a cultural change regarding data sharing], in the sense that now there are much stronger requirements to publish, and that must in some way reflect a cultural change. So, I believe it less acceptable now to adhere to your data and refuse others access. It was never very acceptable, but it is even less acceptable now. (RIZ2)*

Awareness and opinions on data sharing between the two interviews indicate an ongoing change where increased data sharing is the result of increased solidity and transparency in research as goals. The need for a cultural change regarding research data management in academia has been addressed in several studies (Andrews Mancilla et al., 2019; Collins, 1998; Dalton et al., 2020). Such a change does not happen top-down through policy, but it does require the different communities to adapt their practices and find appropriate standards withing their domains (Berman, 2017; Cooper & Springer, 2019; Darch et al., 2020). While Plan-S does not address research data, it triggered a debate amongst researchers regarding quality control in research output. The findings discussed above suggest that the Plan-S debate has contributed to heightened awareness about data quality and data sharing as the foundation of research outputs.

6.1.2 Identities and services in the breakdown and build-up of organizations

During the period of this study, there has been a continuous reorganization of the infrastructure providers and the larger higher educational infrastructural ecology in Norway. Existing organizations and communities have been split and have merged very rapidly (Ministry of Education and Research, 2016, 2021). In this process, existing service providers compete in being valued and finding their place in the new infrastructural landscape.

The establishment of UNIT in January 2018 was a central event discussed by several of the participants in the first interviews (se chapter 1.3). For the stakeholders involved, this created a controversy regarding the identity and skills in the new organization.

> *In the national strategy, there is much talk about the new [UNIT] having a coordinating role regarding [research data], but the point is that they have barely any research data experience [...]. I know that both the Norwegian Centre for Research Data and Sigma2 threw themselves into this based on the report written at the university, saying that some of the technical solutions should be external (ROT1)*

The universities were used to having a direct say in infrastructure development: '*many things have been solved bottom-up and according to the tradition of pooling the resources' (POK2).* With the reorganization, the infrastructure organizations were suddenly one strong stakeholder, hierarchically placed above the universities as decision-makers regarding infrastructure development. The directory represents a new administrational level between the universities and Ministry of Education and Research.

Furthermore, one of the three infrastructure providers, the Norwegian Centre for Research Data, was evaluated and put under constant pressure to maintain its long-lasting identity and tradition as a data archive. In addition to the consequences of reorganization, the literature described how automatization of depositing and retrieval of data are creating a new distance between archives and their communities (Donaldson et al., 2020; Kim et al., 2019; Kouper, 2016). As a response, they chose to prioritize proving their worth and contributing as a separate entity with specialist expertise on research data curation. One example of this was the suggestion to offer expertise on data curation to the sector (L. Kvale, 2021b), in addition to increased involvement in data management training on a national level. As suggested in the quote below, communication is challenging, not only in the distinction between their two roles, as personal privacy service provider and a data archive for their users, but also with what is possible according to the privacy law.

> *I have heard researchers multiple times claiming that The Norwegian Centre for Research Data told them to delete their data, and I have never said this to anyone. But still, this is the perception. We have a recurring*

> *communication challenge in making the individual [researcher] familiar with the legal system. (IN2 in Kvale & Darch, in press)*

The importance of expertise regarding personal privacy and data sharing throughout the research data life cycle is highlighted in article 3, where researchers share the experience of receiving letters asking them to delete data (L. Kvale & Darch, in press). At the same time, findings suggest that research data services should be provided closer to the researchers (Darch et al., 2020; L. Kvale, 2021b). An interest in and understanding of research with its embedded values, ethics, different methodological traditions and paradigms is essential for providing both personal privacy services (L. Kvale & Darch, in press) and data stewardship services to researchers (L. Kvale, 2021b). Reuse and evaluation of the data's relevance for further use are aspects of research data management requiring further knowledge (Donaldson et al., 2020; Kouper, 2016). However, lines for collaboration and development of services require trust, and as the restructuring of services has made several stakeholders uncertain about their own contributions. It is crucial to enable people to look for, and continue, external partnerships as the reorganization are implementing a new identity and responsibilities.

6.1.3 Agendas embedded in data management planning tools

Two infrastructure providers have developed different tools for data management plans. However, during the period of the data collection, infrastructure developers have changed the focus on data management plans from meeting formal requirements created by policymakers, to focusing on how the DMP could become an embedded part of the research planning thereby useful for researchers (L. Kvale & Pharo, 2021). The literature describes machine-actionable plans with automated assessments of funders' requirements (Cardoso et al., 2020). In addition, the development of tools is impacted by infrastructural developers who are eager to explore the possibility of harvesting metadata for the archives of their organizations (L. Kvale & Pharo, 2021). This multiplicity of data purposes in a data management plan, and the potential of using them for control, automated assessments, and metadata harvesting, is seen as problematic by research support staff.

> *'There are some entities who think they should use the DMP for all other types of purposes, for their own advantage' (LG in Kvale & Pharo, 2021)*

Machine-actionable data management plans with automated assessment would be in line with what is described as the users adapting to the technology rather than adapting the technology to the user (Floridi, 2015). Hiding decision-making processes within algorithms is problematic for many reasons including transparency and accountability (Wachter et al., 2017). For prominent researchers, the ability to think outside the box is part of what makes them distinct; preselection and automated decisions then become a straitjacket rather than an advantage. The importance of infrastructure maintenance is highlighted by the realization that what the researchers need might not be what the engineers are most keen on developing (Bowker, 2005). The work of interoperability between data repositories as described in the ideal of the European Open Science Cloud requires advanced understanding of metadata quality (J.-S. Lee & Jeng, 2019; Limani et al., 2019).

Ubiquitous expertise is difficult to capture in computers (Collins & Evans, 2008). Recognition of the fact that a certain understanding of implicit rules and traditions is needed for data management support has not yet gained full acceptance among developers. Lacking flexibility is already a well-known frustration among researchers in their encounters with the research administrative bureaucracy (L. Kvale & Darch, in press). ʿSticklers by the rules' and ʿzombies' are some of the harsh vocabulary used by frustrated researchers (L. Kvale & Darch, in press). Automated decision-making and artificial intelligence risks are rapidly becoming the zombie-bureaucrats that are described by one of the researchers cited in Kvale and Darch. Still, even if it is called 'intelligence', there is no reflection and reason in an algorithm (Floridi, 2015). In data management planning, the requests highlight the need for reflection, automated decision-making based on a black and white yes or no; readings of guidelines and legalization are the contrary. This illustrates how balancing internal goals and agendas of their own organizations with those of the researchers remains a complex task for the infrastructure providers (L. Kvale & Pharo, 2021). Infrastructure organizations are under constant pressure to prove their intrinsic value to their funders while digitalization is also transforming their user communities, as echoed by the literature (Donaldson et al., 2020).

6.1.4 Summary of discussion of RQ1

The three examples provided here individually illustrate how development in the knowledge infrastructure is interconnected with users and the expertise and interest of the involved stakeholders.

As described by Star and Ruhleder, infrastructure is built on an "Installed base"(Star & Ruhleder, 1996). The installed base also includes the existing organizations with their professional identities and communities of practice. As the discussion in this chapter shows, political decisions to restructure all infrastructural services in the higher education sector has influenced priorities within the different organizations. Restructured aspects experienced by infrastructure providers are planned breakdowns, based on organizational ideals (Ministry of Education and Research, 2016, 2021). Star and Ruhleder (1996) writes how the infrastructure 'becomes visible upon breakdown' while Latour (1987) describes the breakdowns as moments when it is possible to investigate technology as when the black box is reopened for repair. Following Edwards (2010, p. 432), one could argue that the robustness of a well-functioning knowledge infrastructure lies in maintenance and continuous development and that black boxing hence becomes outdated. The expertise required is what Clarke describes as unrecognized anticipation work that goes unappreciated (Clarke, 2016, p. 104).

The knowledge infrastructure is dynamic and is never finished or closed (Edwards, 2010). In Norway new infrastructural services are currently being developed top-down, building on existing solutions. Its development follows the political winds of open science and data science as it gradually becomes part of the researcher's workflow with data, ensuring transparency and research quality. Still the effort in creating new, strong organizations is at risk of becoming the goal about which the infrastructure providers seem uncertain. For the infrastructure to be relevant to the different user communities, the goal of the infrastructure organizations must align with the goals of the researchers, namely, to produce high quality research. In the process of reorganization's additional goals, including intrinsic value, own interpretations of funders' requirements and rationalization are at risk of obscuring the goal of quality research. To ensure data quality and comply with the FAIR principles, stakeholders must keep the focus on data curation.

6.2 Stakeholders' roles, and challenges faced in facilitating research data sharing

In this section, RQ 2 is addressed: How and why do stakeholders in research data curation perceive their own roles, and the challenges they face, in facilitating research data sharing? How and why do they perceive the roles and challenges faced by other stakeholders? The question is discussed from two perspectives: stakeholders with multiple identities and data stewards as professionals on the boundary of domains.

Research data management practices are developed at the crossroads between the different research support services, policy, infrastructure, and research. The process of data management closely relates to invisible work (Sands, 2017) as strategies to improve information flow and reduce obstacles to data sharing.

6.2.1 One stakeholder, multiple identities

In this study, the participants were invited based on their work with research data in their primary occupation. The interviews did, however, reveal that several of the participants identified with more than one of the stakeholder groups. The IT staff referred to experience with research as part of the success: 'How I work with data depends on what role I have, as I used to be a researcher' (IT2 in Kvale & Darch, in press) 'Part of the success in my department is due to half of the staff having a PhD, so that we can communicate with the researchers ' (ITI in Kvale, 2021). One of the library staff described how she worked with data collection, illustrating how she was actively participating in conducting research: 'I was part of a data collection project in France' (L1 in Kvale & Darch, in press). These are but a few examples of how stakeholders combine roles and identities in their work. Figure 8, illustrate the number of participants who were invited as an individual stakeholder, but who repeatedly referred to other roles during the interviews. During the interviews, nine frequently referred to their background as researchers. Moreover, two of the research support staff offered infrastructure at a national level and therefore identified themselves as both infrastructure provers and research support staff. Three researcher's policy experiences were identified: two had experience from ethical review boards (L. Kvale & Darch, in press), one was involved in multiple refence groups regarding open research policies. It is potentially useful for librarians, policymakers, infrastructure providers to have a background as researchers It is potentially useful for librarians, policymakers, infrastructure providers to have a background as researchers. By having a background as a researcher, it is easier to relate to the perspectives of research.

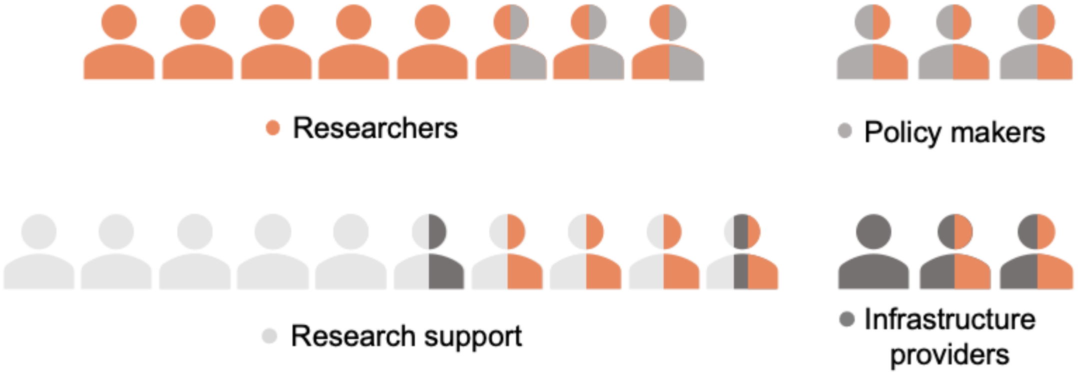


Figure 8. Illustration of the multiplicity of identities held by the stakeholders

In their work on boundary objects, Star and Griesemer (1989) refer to people belonging to multiple groups as 'marginal people', arguing that multiple memberships can be an advantage in terms of translation. Translation as the ability to communicate needs across domains requires interactional expertise (Collins & Evans, 2008). Having expertise in multiple domains is an advantage for work with research data management, but it requires extra effort to maintain the multiple memberships.

> *Trying to figure out how to manage one's boundary crossing that joins and separate social worlds, in which the marginal man lives but does not quite gain full acceptance. Should I assimilate, return, transcend? (Griesemer, 2016)*

Although the participants belonging to multiple groups use different strategies to balance two (in one case three) identities and values, assimilation is most common When it comes to data management plans and encountering privacy law and ethical committees (L. Kvale & Darch, in press; L. Kvale & Pharo, 2021), there are clear differences amongst stakeholders, despite those having two or more roles. This tells us that marginalized researchers primarily adapt to their new roles and values in work with policy, infrastructure, or research support. Still, having experience as a researcher was regarded as an advantage and a qualification for the work with research data. Speaking the same language, mutual trust and not adding additional agendas to that of solidity are underscored as important qualities of data stewards; these are complemented by recruiting trained researchers for data stewardship and infrastructure development (L. Kvale, 2021b).

Of the nine marginal researchers, eight referred to their research experience in the past tense, illustrating how maintaining their identity as a researcher while at the same time performing work in research support, policy or infrastructure is challenging. One example of transcending and maintaining an active membership in multiple groups was found among the library staff L1[6] cited in Kvale and Darch (in press) – a participant describing her active participation in research.

> *I was part of a data collection project in France where we also had partners from Japan. And when the participants talked about what food they like [...] this was considered sensitive information by the Japanese researchers and could not be made available. (L1 in Kvale & Darch, in press)*

This library staff holds an academic library liaison/subject specialist position with dedicated research time and mentions actively conducting research within the field of her subject expertise. While providing research data support and data stewardship services, she is also a researcher. In digital scholarship centres that commonly host research data management services in libraries, such library liaison positions are rarely embedded (Hannah et al., 2020). Involvement and embedding of the subject expertise in data management services is regarded as essential for successes (Darch et al., 2020; Hannah et al., 2020).

An identity as a researcher is perceived as valuable for interactional expertise, namely through understanding and translating the needs of the researcher in their current position (L.

---

[6] In Kvale and Darch all participants were given new codes in order to protect the identity of some participants who did not wish identification.

Kvale, 2021b; L. Kvale & Darch, in press; L. Kvale & Pharo, 2021). More possibilities for balancing and transcending between multiple identities and stakeholders roles would benefit research data management services.

6.2.2 Stewards as professionals on the boundary of domains.

The tendency described in the literature (Kim, 2020; Tenopir et al., 2017, 2019) and confirmed in this study is that the library normally has the overall responsibility for research data management services (L. Kvale, 2021b). ‛The plan is that I shall be one of the driving forces behind this on behalf of the library, preparing the whole organization for research data sharing’ (L4 in Kvale & Darch, in press). Still, research support personnel from all four universities pointed out collaboration between the different groups as central when delivering research data services (L. Kvale, 2021b). Collaboration brings out different cultures, ideals and motivations among the research support services.

> *Compared to colleagues [in the library] who have this glowing attitude towards sharing, wearing Open Access buttons and such, well, I feel I am not quite there, with such a strong sense of dedication. (Research Office)*

Researches are not convinced by the idealistic arguments of open access “evangelists” to share data (L. Kvale, 2021b).

> *“If you are not very into these things and you have a project where you just want your research done, then you probably don’t want an ideologist marching into your stuff with a banner, preaching this and that” (RIJ)*

Still the researchers are positive towards sharing data for creating new and more solid research. The multiplicity of agendas in research support take the focus away from solidity (Darch et al., 2020; L. Kvale, 2021b). Researchers are committed to their research, while developers have other commitments in terms of both multiple projects (Darch et al., 2020) and their own organizations as discussed above in chapter 6.1. Closer collaboration between existing research support services would require a breakdown of existing organizational and cultural boundaries within the different support functions, and would demand that the research office, university IT and university library form a ‛holy trinity’ (RGA in Kvale, 2021). Responding to this would be to bring together expertise from different traditions of research support with solid research as the main goal along with the creation of a new additional entity on the margins of current communities (L. Kvale, 2021b).

The emerging communities of data stewards could potentially contribute to the development of a new common ground for research support (L. Kvale, 2021b). Several of the participants in research support positions in this study already provide different types of data stewardship services, including guidance, training, data storage and archiving and curation. In this work, they are also sharing the borderland with researchers (L. Kvale, 2021b) . Documentation, including descriptive processes, decisions, errors, transformations, and evaluations conducted with research data as part of research is time-consuming for researchers. These are necessary processes to document and make the evaluations and conclusions visible when data are to change hands. Data stewards with a background from research, infrastructure providers and policymakers all operate on the borderland to research (Teperek et al., 2018). They all do their best to simplify and assist the researchers in this documentation and handover of data. Their work in translating between different domains often goes unrecognised, however, and the doors allowing one to return into the research milieu are not kept open.

Many of the decisions made regarding research data are based on invisible knowledge (L. Kvale & Darch, in press; L. Kvale & Pharo, 2021). This is the case for both ‘the ethical considerations conducted by researchers on when, what and where to share’ (L. Kvale & Darch, in press) and data management where ‘researchers look to their international research communities for best practice and request support functions in their research institutions’ (L. Kvale & Pharo, 2021). This invisible knowledge has been built up through experience and dialogue with fellow researchers. Maintaining such knowledge is challenging and done through active memberships in the respective research communities. Creating dialogue with library liaisons on how to succeed in combining identities and calling for academic recognition of data stewardship are two strategies for connecting data stewards with multiple communities.

6.2.3 Summary of discussion of RQ 2

Scroggins and colleagues highlight the importance of the domain expertise of the researchers: 'No matter how well code is documented, no paper trail can substitute for the rich domain expertise and tacit knowledge of those who conducted the science' (Scroggins et al., 2019). Data stewards recruited from research to operate in a borderland of research and support are one example of marginal people in the knowledge infrastructures. Data stewards are valued for the skills of translating and performing invisible work. Of improving the flow or decrease the number of obstacles when the research data are moving between different worlds: the research group, other researchers, the archive and so forth. Handling multiplicity calls for recognition of the effort in maintaining memberships in multiple communities of practice (Wakeford, 2016). Recognizing these efforts would also facilitate for an exchange of experiences with maintaining multiple memberships. There is also the possibly of merging the two roles by making research a part of the work of data steward, looking to the roles of the library liaisons (Hannah et al., 2020). Creating teams in which data stewards with different expertise work together is another way of approaching the need for multiple skills. Both approaches may be necessary to fill the current data stewardship gap.

Griesemer questions whether it is possible to assimilate, return or transcend (Griesemer, 2016). For this to be a real option, it is necessary to facilitate for the maintenance of multiple memberships. Creating communities of practice for learning and exchanging experiences would be a first step in recognizing data stewardship as a domain in the borderlands of multiple distinct disciplinary traditions (Teperek et al., 2018). Marginalized researchers seek recognition both as researchers and as data stewards; this requires strategies for maintaining multiple memberships over time and career tracks and credit recognized in more than one domain. As mentioned in 6.2.2, dialogs with library liaisons and working for academic recognition for data stewardship are possible strategies for providing date stewards with bonds to multiple communities. Exploration of ways for marginalized researchers and the emerging group of data stewards to achieve recognition and keep their expertise as researchers up-to-date is necessary as the data stewardship domain is growing. Just as research is a lifelong process of learning and exploring new knowledge, so will data stewards need to balance practice with learning.

6.3 Facilitation of knowledge transfer in research data curation

The third research question, RQ 3, builds on the previous discussions of managing multitudes in collaboration and own roles by asking: How do stakeholders manage these differences and facilitate knowledge transfer among the key stakeholders involved in research data curation?

In collaborations, common platforms and agreements of standards and terms reduces obstacles in the knowledge exchange. Often the different stakeholders perceive and apply the entities necessary to facilitate data sharing differently. This is not necessarily a problem if there is an agreement on a common goal and if the understanding of the entities are plastic enough take the role as boundary objects (Star & Griesemer, 1989). Several entities in the infrastructure share characteristics with boundary objects: Common grounds for collaboration, Policies, Data management plans, Standardization documentation, and Data repositories. In addition, Marginal researchers, and Research data have many of the same qualities and functions in knowledge exchange.
In this section, the entities will be analysed as boundary objects in order to show how they function to facilitate collaboration and knowledge transfer between the different stakeholders and development of the infrastructure.

6.3.1 Common grounds in research collaborations

Common grounds on ethics and privacy in research collaborations are platforms of agreement on directions for further collaborative work proposed in article 3 (L. Kvale & Darch, in press). While creating common understandings of complex concepts such as privacy and research ethics may appear labour-intensive and abstract, they work as points of reference and common platforms in collaborations. In international research collaborations, countries represent boundaries which researchers must relate to in terms of culture and legislation (L. Kvale & Darch, in press). Article 3 highlights the importance of a common ethical and personal privacy ground in international research collaboration (L. Kvale & Darch, in press). These are complex concepts requiring an alignment of the differences in abstraction of norms. Instead of presuming consensus-shared meanings, common grounds seminars could facilitate the creation of common approaches to ethics and privacy within research groups. This again would results in the ‘deletion of local contingencies and global rules’ (Star, 1989).

Shared understandings are often presumed, however local contingencies in how complex concepts such as research ethics and personal privacy are perceived and applied within a project could reduce potential obstacles to sharing at a later stage.

6.3.2 Policies connecting domain by pointing direction

Policy documents shape the development work in a similar way by facilitating communication and connecting different domains (Star, 1989; Wu & Worrall, 2019). Policy documents create rules and directions from the top for how a domain should be developed. Policies are sometimes criticized for being implemented without sufficient funding and without the proper infrastructure in place.

> *The funders have started to say that if you are not FAIR and Open science with your data, you will not receive funding, and the infrastructures are behind, because it is very popular to create policies without money attached (ITI1)*

Aligning development of infrastructure with policy is challenging for the stakeholders involved. Policies address two ends of the infrastructure development by outlining the expectation for what should be developed and at the same time create expectations amongst the researchers that these solutions are already in place. In this way, research support services and infrastructure developers are caught in-between with expectations to deliver promptly what they might lack the time and funding to develop.

As illustrated with Plan S and reorganization of infrastructure providers in chapter 6.1, policies are created to push the development in given directions but are commonly interpreted according to the needs of the various stakeholders as with the example of development of data management plan tools discussed in 6.1.3.

6.3.3 Data management plans

The data management plan plays a role for multiple stakeholders in multiple ways (L. Kvale & Pharo, 2021). The data management plan connects the stakeholders as a document laying the groundwork for how data collection and dissemination is to take place with curation, sharing and reuse as goals (L. Kvale & Pharo, 2021). The current interest amongst infrastructure developers in automation of the data management plans risks placing the researchers on the side lines rather than focusing on the researchers as the primary users of a data management plan, i.e., as a document for actually planning their homogenous ways of conducting research. There is no single answer as to what research is and how it should be conducted, nor is there one correct answer to how research data should be prepared for sharing. The fact that something can be automated, measured and controlled according to one fixed standard is not a valid argument for automating. Instead, there is the risk that the data management plan will become merely a checklist to submit, "a bureaucratic exercise" (L. Kvale & Pharo, 2021), rather than a tool to prepare the research data for sharing and for increasing research quality and transparency. Researchers need autonomy and guidance in developing a data management plan so that it enhances reproducibility and enables data sharing (L. Kvale & Pharo, 2021).

6.3.4 Standardization documentation

During data collection and analysis, the writing of documentation as an unstructured or low structured detailed description of what the data are and the transformations that are taking place plays a key role in facilitating data reuse (Wu & Worrall, 2019). Documentation typically holds a low degree of standardization across disciplines and methods to describe the variations of data and research conducted. The role of documentation is to facilitate the translation from the researcher to reuse in different contexts. Documentation is commonly stored alongside data in repositories; the level of documentation and metadata in line with the FAIR principle, however, is complex (Bishop et al., 2019; Kim et al., 2019; Yan et al., 2020). The expertise to develop documentation, including describing the data so that the conditions, time and place are specific enough for the data to travel, could be described as invisible work (Star & Strauss, 1999). The standards for documentation and metadata must be both flexible enough to grasp the richness of the data and standardized enough to allow interoperability. The development of standards is happening within research communities, as discussed in 6.1.1., and within repositories. Consequently, different stakeholders talk past each other when addressing metadata because metadata are necessary on multiple levels according to different standards in order to disseminate the data they describe.

6.3.5 Data Repositories

As data are transferred to the data archive or data repository, general metadata according to the standard required by the archive are added (Kim et al., 2019; J.-S. Lee & Jeng, 2019). Similar to libraries, 'repositories are built to deal with problems of heterogeneity caused by differences in unit of analysis' (Star, 1989). Research data repositories do not end in a string of events but are rather an element in the life cycle for research data with a separate repository-central lifecycle for data curation. Archival metadata are commonly highly structured for retrieval; ideally, also standardized across repositories, allowing discovery tools to search across multiple repositories and fulfil the FAIR principles. Data archives provided by data organizations/infrastructure providers align research data and documentation in a new context as entities within the collection. The reuse of data requires the repositories to facilitate both the archiving and the retrieval, which at first glance is similar to that of libraries; however, research data are complex objects not easily classified and standardized (Bowker & Star, 1999). Definition of scope, communities, and standards are shaping the data repositories as units with a certain level of homogeneity in their multiplicity (Donaldson et al., 2020). Different data repositories operate with different communities and with defined scopes as memory. Curation and quality assurance are key operations conducted by repositories to receive recognition from designated communities. In addition, they facilitate research data in crossing disciplinary boundaries, to be reused and re-interpreted for new purposes.

6.3.6 Marginal researchers

As discussed in 6.2, the number of stakeholders recruited from research to other stakeholders' functions involved in research data sharing, creates a diaspora of trained researchers operating in the borderlands of research: 'A new social world of others like themselves' (Star & Griesemer, 1989). The creation of communities of practice as worlds where they are valued for their research experience and identity is one strategy to acknowledge the existence of researchers in the borderlands (See chapter 6.2). Their contribution to being marginal yet multiple in identities brings together interactional expertise from multiple domains. When researchers fear that the outsider will add additional agendas (L. Kvale, 2021b), the role of the marginalized researcher becomes to disseminate the voice of researchers and translate the needs and goals of research to the outside worlds.

6.3.7 Research data

As described in the theoretical background, research data have a fluid meaning but have the characteristics of boundary objects as they bring to light "the multiplicity of meanings and the evocative powers of the artefact as a symbol' (Boland, 2016). Symbolic powers of research data are part of what makes defining research data so complex. Research data are symbolically described as "the lifeblood of research" (Borgman, 2012) and are socially created through the meanings researchers assign to them, making them artefacts of research. Research data are temporal, as argued with the *when of research data* (Borgman, 2015; Haider & Kjellberg, 2016). Just as research data must be defined and conceptualized depending on application and context, it is also very much a matter of boundary crossing. Recognizing that research data means different things to different stakeholders is a first step in collaboration on the development of infrastructure for research data sharing.

6.3.8 Summary of discussion RQ3

Collaboration takes place amongst stakeholders with divergent perspectives on data sharing. Stakeholder groups work towards multiple secondary goals, including political goals and positioning of own organization (chapter 6.1 and Cox et al., 2019). Still, the previous subsections in chapter 6.3 illustrate how collaboration and knowledge exchange amongst stakeholders is based on an agreement on the importance of data curation for research quality and transparency through sharing. Still, there is no consensus on how and why. The different entities described are evolving back and forward between attempts at standardization and residual categories. Standards are never static; our classification and understanding of the world are ever changing (Bowker & Star, 1999). Knowledge and cultural perceptions define existing hierarchies and understandings of the world (L. Kvale & Darch, in press). Mapping and patching so that knowledge is not lost as categories change is a continual process of curation taking place in repositories (Bowker & Star, 1999). By analysing infrastructure development in the context of the hermeneutic circle, where added experience and knowledge lead to new interpretations, the knowledge aspect of an infrastructure can be visualized. The process of adding knowledge becomes visible both in the evolving understanding of data management plans and the dialogue necessary for policy development; here stakeholders increase their knowledge on the issue and apply this in the development of entities. Infrastructural development is a constant cycle of creating and dissolving as new knowledge and new technology are added.

6.4 Methodological limitations

The Delphi method has been allegedly unscientific and criticized for forcing consensus through group pressure (Woudenberg, 1991). According to Fletcher & Marchildon (2014), the application of strategies from qualitative research in a modified Delphi strengthens the quality of the study. The current study did not focus on creating a consensus, but rather on understanding the directions of infrastructural development.

Other methods including survey, focus groups and ethnography were considered. The homogeneity of the stakeholders, however, created a risk that surveys would become either superficial and general or not property understood by all participants. Large surveys, however, are popular in the domain and provided results that are easier to generalize. Focus groups might have been an interesting approach, putting stakeholders in each location together in dialogue about how they approach data sharing; on the other hand, focus groups would silence the critical voices and the disagreements between the different participants. Ethnography would not have given the frames of study that I needed to complete the data collection but could potentially have provided a similarly rich data material.

Time and capacity are two of the main constraints when conducting a multiphase study (Creswell & Plano Clark, 2018, p. 103). A PhD project has its clear limitations in both. I have been lucky, however, in having done this over many years while working part-time at the University of Oslo Library. Data collection, transcribing and preparation for the next data collection have been time-consuming, and time to reflect on methodological choices and the analysis has been essential.

6.5 Reflections on the research design

In the final consent form, the participants were asked openly about how they felt about participating in the study. The question was added to allow the participants to provide open feedback on the research process and method. Based on notes taken along with the interviews, I found mixed experiences amongst the participants. Some of the participants experienced methodical choices as difficult. Some participants requested that the interview guide be shared ahead of the interview and perceived it a workload for others; whereas some said that reviewing the interview transcripts made them feel vulnerable reading their own voice in a transcript, others appreciated this possibility to review. Others experienced that they were asked to have opinions on issues they had not reflected on, while some participants appreciated the sharing and were positive towards the experience of reflecting on issues they had not previously thought about. I wanted clearer data on the participants experience, to also evaluate how my emphasis on data sharing affected the participants.

In the consent form regarding data sharing, three types of feedback were received: 1 Opinions on issues they had not reflected on; 2 discomforts in sharing interview transcripts; 3 own learning through participation. Each of these three reflects aspects of the method noted during the interviews and requires reflections when designing a study.

6.5.1 Opinions on issues the participants had not reflected on

While the participants were recruited as experts on research data in different roles, this was not the primary occupation of all, and some questions were outside the scope of their domain or interest. During the interviews some of the researchers made disclaimers regarding parts of their responses such as 'I just have to say that I don't have a carefully weighed opinion on this; this is not something I have spent time reflecting on – I just reply whatever pops into my head as we speak' (RIZ). The interviews were therefore carefully designed not to push responses, and instead moved on to other questions or flipped the order of the questions to start with what the participants were most comfortable discussing. The analysis focused on using experiences as starting points and made it a point to respect statements like the one above. One researcher said she sometimes felt she was asked to have opinions on issues she had not reflected over:

> *It has been a bit demanding because many of the questions were new to me—and I barely understood the scope of some of them. So I felt sometimes that the interviewer wanted me to have opinions on things I did not quite understand and therefore did not have qualified opinions about. (Researcher)*

Similar feedback was also given before or after the final interviews by some of the other researchers. It is challenging to select participants in a Delphi study; care was taken to avoid recruiting biasedly only researchers who were outspoken on the issues of data sharing. Extra attention regarding the participants' opinions was necessary during the analysis, which also shows the importance of giving the participants the response option 'no opinion'. *Knowledge infrastructure* for research data is a complex matter, and as one researcher pointed out laughingly when I asked about DMP, 'We do not all go around thinking about data management plans 24/7' (RIL2). She had not heard of a DMP prior to our first conversation. In the end, the differences in expertise and perception amongst the participants contributes to describe the rich landscape of data sharing.

6.5.2 Discomforts in sharing interview transcripts

The second type of feedback pertained to the review of the transcripts and sharing of non-sensitive data. This was given from participants in different stakeholder groups:

> *The interviews worked well and have hopefully also been useful. One recommendation for later projects is to modify transcripts so that they consist of complete sentences; this would not reduce the truth content in the reporting (which is to be approved by the interviewee anyway) and would probably increase the understanding and ensure more unambiguous interpretation. (Researcher)*

Transcriptions as text format often consist of incomplete sentences, one researcher suggested that these should have been edited to complete sentences, arguing that this would give better interpretations. While what she suggests would have required a complete rewriting and locking the text interpretation to something that could blur uncertainty and associations the subjects expressed, I acknowledge that some editing and proper proof-writing would have made the transcripts easier to read.

> *It is interesting to be in the information position and relate to one's own statements. Even if the information I have provided is in no way sensitive or provoking, I feel some hesitation in accepting publication of the interview material, even in pseudonymised form. This is interesting and will be a useful experience to carry along with in the work with support services for those working with qualitative data. (Research support)*

The vulnerability the participants felt when reading their own transcripts adds to the debate over personal privacy and data sharing addressed in the third article (L. Kvale & Darch, in press). This discomfort regards participants' personal identity and self-concept in the context of data sharing and needs addressing from the research support functions to ensure that participants privacy is protected. Some of the participants passively did not consent to sharing data, this was done by not responding to the consent form for data sharing. Care was taken not to push for responses, as no answer was interpreted as the easiest way to say no. Identification and self-image are two different aspects of privacy that require equal weight, the first often receives more attention and is easier to measure, self-image, however, is equally or possibly more important to research participants.

6.5.3 Own learning

The third aspect was the learning and development of different perspectives. This is not only a result of the development in the data sharing domain, or new questions brought up through the interviews, but also a consequence of new knowledge from the participants. One research support staff participation writes the following:

> *Interesting theme and interesting question, which initiated reflections and ideas. Considering my work participating in the study, it felt relevant and useful. (Research support)*

This participant brought out how the experience of being a research participant provided a new perspective on data sharing. This aspect of participating in a Delphi study was shared by several of the participants and corresponds with what Gupta and Clarke (1996) describe as the learning that takes place in simulations, with the Delphi study being a research instrument. In addition to this learning through participation, the learning was described as conceptual understanding of issues as infrastructural development occurring. Comparing understandings and knowledge in the two rounds of the interviews helped display not only contextual and infrastructural changes but also developments in the participants' views or understandings of data management issues. As reflected on in the last interview by another participant:

> *I have become more and more uncertain about what a DMP is [...]. I am sure I had a simpler understanding of the problem the last time we spoke, and I am a bit frustrated regarding my lacking capability to get anywhere. (INO in Kvale & Pharo, 2021)*

The participant reflected on her own learning during the year. Opinions and understandings of a phenomenon change over time; what might appear simple can suddenly become overwhelmingly complex as one learns to understand it in more detail. This reflection corresponds with way it is described by Gadamer (2004) in the hermeneutic circle. Our pre-understandings shape our understanding and 'this constant process of new projection constitutes the movement of understanding and interpretation' (Gadamer, 2004, p. 270). This development is constantly happening as new knowledge is gained and part of the knowledge infrastructure development remains invisible without the longitudinal approach. While the time frame of a PhD project is too short to create true longitudinal data, the development in the domain is happening so fast that some changes, such as the one noted by INO become visible. The learning perspective illustrate how the development of the infrastructure continues and expertise amongst the stakeholders grows, even though this study of the KIs for research data is completed.

6.6 Recommendations for further research

A line of inquiry which requires further exploration is the role of researchers as data stewards and ways of balancing and adapting identities. This thesis scratches the surface of the identity of professionalized data stewards but there are still many questions to answer regarding these roles.

Furthermore, the exposure the participants felt when reading their own transcripts is an important addition to the debate of personal privacy which would require further addressing and unpacking from a research data management perspective and related to ensuring participants privacy.

A final topic I hope to see addressed in future research is how research data sharing relates to the identities of research. This topic was brought up by several participants when motivations for data sharing came up. An investigation of data sharing in relation to what research is would require levels of unpacking and analysing that go far beyond my philosophical capacity. I believe a better understanding of this aspect is relevant for the current debate and policy development regarding data sharing and would love to see this properly addressed.

Chapter 7. Conclusion

Through the findings discussed in the previous chapter, the conclusion explores **How do the perspectives and expertise of key stakeholders involved in research data sharing affect the collaboration and knowledge transfer amongst these** through the lens of the three research questions

RQ 1: How and why do differences between stakeholders' perceptions affect their ability to collaborate in the work of research data curation?

The development of solutions for research data is continuous, based on input from different stakeholders. Both the Plan-S example, development of data management plan tools and the reorganization of service providers illustrate a highly policy-driven development of solutions promoting open research through a streamlining of the research services. When policy documents are interpreted by infrastructure providers, researchers and research support, their application of policy targets different aspects. Technical development fluctuating between automation and reflection in a data management plan illustrates the continuous balancing of the standardized and the temporal. Still, standards and best practice for data management are developed within communities of practice as invisible bottom-up parts of research practice. Further alignment of goals, putting research quality in the centre, would benefit the development by indicating one direction, rather than multiple directions. The interpretive translation of needs is a continuous proses of learning and knowledge exchange between different expert groups. In this process the voice of the homogenic mass of researchers needs amplification if we do not want to invest in solutions that are irrelevant for the research communities.

RQ 2: How and why do stakeholders in research data curation perceive their own roles, and the challenges they face, in facilitating research data sharing? How and why do they perceive the roles and challenges faced by other stakeholders?

Current development of the knowledge infrastructure changes existing roles as researchers operate in related domains such as infrastructure providers, research support and policy developers; the marginalized role creates challenges for researchers operating in the borderland of research and support. This also goes for emerging professional roles such as the data stewards. Data stewards are commonly recruited and evaluated based on their experience with research, creating a new identity in the form of data stewards. Further, a breakdown of existing boundaries between research support services could simplify the dialogue with these services for the researchers; the data stewards place a central focus on the creation of a 'new' research support identity.

RQ 3: How do stakeholders manage these differences and facilitate knowledge transfer among the key stakeholders involved in research data curation?

Development within the domain of research data happens with input of knowledge from expert users and reinterpretation. Continuous development causes the different elements in forms of boundary objects to be recreated within the infrastructure, moving between attempts at standardization and residual categories. This process is taking place in a social context where technology is changing society and society is changing technology; one cannot be separated from the other, as the technological changes are ecological. A search for one perfect infrastructure for data sharing would fail as it approaches infrastructure as something to be finished rather than kept up-to-date and developed, linking the past with the future. Different stakeholders apply entities in their different contexts with a certain level of standardization; through application within the different communities the standards evolve, the entities become residual categories before they again research a new level of standardization. This process of developing standards and sharing goals contributes towards knowledge transfer amongst the stakeholders involved; it does, however, require lines of communication between the different communities.

7.2 Implications

Expertise and the need to rethink research data support services with a focus on the identity of data stewards as domain specialists and data management experts requires the creation of communities and incentives for recognition of multiple memberships as addressed in 6.2.3. In addition, it requires a re-thinking of research data services based on a common goal of creating better research. Research data are expected to cross disciplinary, institutional, and cultural boundaries according to the ideals of a global interdisciplinary research data flow. Some research data are to be made openly available, while much research data should be preserved and shared without being open due to rightful ethical or legal constraints. A close connection to domain expertise and a focus on research data sharing applying and developing interactional expertise are also needed. By abstracting the function of different items as boundary objects, it is easier to address the different functions they have for different stakeholders group, and instead of working towards a consensus, the differences can be encompassed as a continuous development and improvement of standards necessary for knowledge exchange.

References

Andrews Mancilla, H., Teperek, M., Van Dijck, J., Den Heijer, K., Eggermont, R., Plomp, E., Turkyilmaz-van der Velden, Y., & Kurapati, S. (2019). On a Quest for Cultural Change—Surveying Research Data Management Practices at Delft University of Technology. *LIBER Quarterly*, *29*(1), 1. https://doi.org/10.18352/lq.10287

Åström, F., Hammarfelt, B., & Hansson, J. (2017). Scientific publications as boundary objects: Theorising the intersection of classification and research evaluation. *IR Information Research*, *22*(1), 10.

Barocas, S., & Nissenbaum, H. (2014). Big data´s end run around anonymity and consent. In J. Lane, V. Stodden, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks For Engagement* (1 edition, pp. 44–75). Cambridge University Press.

Barth, F. (1996). *Ethnic groups and boundaries* (pp. 75–82). Waverland Press.

Berman, E. (2017). An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Integrated Findings to Develop Research Data Services. *Journal of EScience Librarianship*, *6*(1), e1104. https://doi.org/10.7191/jeslib.2017.1104

Beyan, O., Choudhury, A., van Soest, J., Kohlbacher, O., Zimmermann, L., Stenzhorn, H., Karim, R., Dumontier, M., & Decker, S. (2020). Distributed Analytics on Sensitive Medical Data: The Personal Health Train. *Data Intelligence*, *2*, 96–107. https://doi.org/10.1162/dint_a_00032

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society: Studies in sociotechnical change*. The MIT Press.

Bishop, B. W., Hank, C., Webster, J., & Howard, R. (2019). Scientists' data discovery and reuse behavior: (Meta)data fitness for use and the FAIR data principles. *Proceedings of the Association for Information Science and Technology*, *56*(1), 21–31. https://doi.org/10.1002/pra2.4

Boland. (2016). The concept of boundary object. In G. C. Bowker, S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 229–241). The MIT Press.

Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. The MIT Press.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, PC., & Proudman, V. (2021). *OA Diamond Journals Study. Part 1: Findings*. Zenodo. https://doi.org/10.5281/zenodo.4558704

Bowker, G. C. (2005). *Memory Practices in the Sciences*. The MIT Press.

Bowker, & Star. (1999). *Sorting things out: Classification and its consequences*. The MIT Press.

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European

Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*, *2*, 43. https://doi.org/10.3389/fdata.2019.00043

Cardoso, J., Proença, D., & Borbinha, J. (2020). Machine-Actionable Data Management Plans: A Knowledge Retrieval Approach to Automate the Assessment of Funders' Requirements. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, & F. Martins (Eds.), *Advances in Information Retrieval* (pp. 118–125). Springer International Publishing. https://doi.org/10.1007/978-3-030-45442-5_15

Clarke, A. E. (2016). Anticipation work: Abduction, simplification, hope. In G. C. Bowker, S. Timmermans, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 85–119). The MIT Press.

cOAlition S. (2019). *Accelerating the transition to full and immediate Open Access to scientific publications*. https://www.coalition-s.org/wp-content/uploads/PlanS_Principles_and_Implementation_310519.pdf

Collins, & Evans. (2008). *Rethinking Expertise*. University of Chicago Press.

Collins, H. M. (1998). The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves. *American Journal of Sociology*, *104*(2), 293–338. https://doi.org/10.1086/210040

Cooper, D., & Springer, R. (2019). *Data Communities: A New Model for Supporting STEM Data Sharing*. Ithaka S+R. https://doi.org/10.18665/sr.311396

Corsaro, D. (2018). Crossing the boundary between physical and digital: The role of boundary objects. *IMP Journal*, *12*(2), 21.

Corti, L. (2014). *Managing and sharing research data: A guide to good practice*. SAGE Publishing.

Cox, A. M., Kennan, M. A., Lyon, E. J., Pinfield, S., & Sbaffi, L. (2019a). Progress in Research Data Services. *International Journal of Digital Curation*, *14*(1), 126–135. https://doi.org/10.2218/ijdc.v14i1.595

Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019b). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, *75*(6), 1432–1462. https://doi.org/10.1108/JD-12-2018-0211

Cox, A. M., Pinfield, S., & Rutter, S. (2019). Academic Libraries' Stance toward the Future. *Portal: Libraries and the Academy*, *19*(3), 485–509. https://doi.org/10.1353/pla.2019.0028

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). SAGE Publishing.

Dalton, E. D., Tenopir, C., & Björk, B.-C. (2020). Attitudes of North American Academics toward Open Access Scholarly Journals. *Portal: Libraries and the Academy*, *20*(1), 73–100. https://doi.org/10.1353/pla.2020.0005

Darch, P. T. (2018). Limits to the Pursuit of Reproducibility: Emergent Data-Scarce Domains of Science. In G. Chowdhury, J. McLeod, V. Gillet, & P. Willett (Eds.), *Transforming Digital Worlds* (Vol. 10766, pp. 164–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-78105-1_21

Darch, P. T., Sands, A. E., Borgman, C. L., & Golshan, M. S. (2020). Do the stars align?: Stakeholders and strategies in libraries' curation of an astronomy dataset. *Journal of the Association for Information Science and Technology*, asi.24392. https://doi.org/10.1002/asi.24392

Donaldson, D. R., Zegler-Poleska, E., & Yarmey, L. (2020). Data managers' perspectives on OAIS designated communities and the FAIR principles: Mediation, tools and conceptual models. *Journal of Documentation*, *76*(6), 1261–1277. https://doi.org/10.1108/JD-10-2019-0204

Dutoit, M. (forthcoming). *We exchange data all the time, Researchers' perspectives on data sharing and data policy*. OsloMet - Oslo metropolitan university.

Edwards. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. The MIT Press.

Edwards, Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). *Knowledge infrastructures: Intellectual frameworks and research challenges*. http://deepblue.lib.umich.edu/handle/2027.42/97552

Edwards, P., Mayernik, M. S., Batcheller, A., Bowker, G., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 667–690. https://doi.org/10.1177/0306312711413314

Eschenfelder, K. R., & Shankar, K. (2020). Of Seamlessness and Frictions: Transborder Data Flows of European and US Social Science Data. In A. Sundqvist, G. Berget, J. Nolin, & K. I. Skjerdingstad (Eds.), *Sustainable Digital Communities: 15th International Conference, iConference 2020, Boras, Sweden, March 23–26, 2020, Proceedings* (Vol. 12051, pp. 695–702). Springer International Publishing. https://doi.org/10.1007/978-3-030-43687-2

European Commission. (2016a). *Guidelines on FAIR data management in H2020*. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. (2016b). *Guidelines on FAIR data management in H2020*. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. (2016c). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. European Commission.

European Commission. (2020). *CORDIS : Projects and Results*. https://cordis.europa.eu/projects/en

Regulation (EU) 2016/679, Pub. L. No. 32016R0679, 119 (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng

Fletcher, A., & Marchildon, G. (2014). Using the Delphi Method for Qualitative, Participatory Action Research in Health Leadership. *International Journal of Qualitative Methods*, *13*(1), 1–18.

Floridi, L. (2010). *Information: A Very Short Introduction* (Illustrated Edition). Oxford University Press.

Floridi, L. (2015). *The Ethics of Information*. Oxford University Press.

Frank, E. P., & Pharo, N. (2015). Academic Librarians in Data Information Literacy Instruction: A Case Study in Meteorology. *College & Research Libraries*, crl15-828.

Gadamer, H.-G. (2004). *Truth and method*. Continuum. http://site.ebrary.com/id/10224795

German Council for Scientific Information Infrastructures (RfII). (2019). *Digital competencies – urgently needed! Recommendations on career and training prospects for the scientific labour market* (p. 56). German Council for Scientific Information Infrastructures (RfII). http://www.rfii.de/?p=4015

Gjengedal, K. (2020). Plan S rullar vidare. Frå 2021 skal den sikre open tilgang til forskning ein gong for alle. [Plan S roles on. From 2021 it will ensure open access to research once and for all]. *Forskerforum*. https://www.forskerforum.no/plan-s-rullar-vidare-fra-2021-skal-den-sikre-open-tilgang-til-forskning-ein-gong-for-alle/

Goben, A., & Griffin, T. (2019). *In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys*. 22.

Graver, H. P. (2018). Plan S er et angrep på forskningens frihet [Plan S is an attack on academic freedom]. *Khrono*. https://khrono.no/forskningsradet-hans-petter-graver-open-access/plan-s-er-et-angrep-pa-forskningens-frihet/237503

Griesemer. (2016). Sharing Spaces, Crossing Boundaries. In G. C. Bowker, S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 201–218). The MIT Press.

Gupta, U. G., & Clarke, R. E. (1996). Theory and applications of the Delphi technique: A bibliography (1975–1994). *Technological Forecasting and Social Change*, *53*(2), 185–211. https://doi.org/10.1016/S0040-1625(96)00094-7

Haider, J., & Kjellberg, S. (2016). Data in the making: Temporal aspects in the construction of research data. In *New big science in focus: Perspectives on ESS and MAX IV* (Vol. 8, pp. 143–163). Lund Studies in Arts and Cultural Sciences, Lunds universitet. https://lup.lub.lu.se/search/ws/files/7710905/New_big_science_in_focus_webb.pdf

Hannah, M., Heyns, E. P., & Mulligan, R. (2020). Inclusive Infrastructure: Digital Scholarship Centers and the Academic Library Liaison. *Portal: Libraries and the Academy*, *20*(4), 693–714. https://doi.org/10.1353/pla.2020.0033

Hartson, H. R., & Pyla, P. S. (2012). *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.

Henderson, M. (2020). Why You Need Soft and Non-Technical Skills for Successful Data Librarianship. *Journal of EScience Librarianship*, *9*(1), e1183. https://doi.org/10.7191/jeslib.2020.1183

Hess, C., & Ostrom, E. (2007). *Understanding knowledge as a commons: From theory to practice*. MIT Press.

Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: The intersections of research data management, FAIR and Open. *Insights the UKSG Journal*, *32*, 18. https://doi.org/10.1629/uksg.468

Hildreth, P. M., & Kimble, C. (2002). The duality of knowledge. *Information Research*, *8*(1). http://informationr.net/ir/8-1/paper142.html

Innspill til Forskningsrådet om veilederen til Plan S [Input to the Research Council regarding the guidelines for Plan S], (2019) (testimony of Akademiet for yngere forskere, Det kongelige norske videnskabers selskab DKNVS, Forskeforbundet, & Universitetet i Oslo UiO). https://www.forskningsradet.no/siteassets/malgruppene-med-caser/forskningsorganisasjoner/innspilltilforskningsradetomveilederentilplans7-2.pdf

Jacob, E., & Albrechtsen, H. (1998). The Dynamics of Classification Systems as Boundary Objects for Cooperation in the Electronic Library. *Library Trends*, *47*(2), 293–312.

Ju, B. (2013). Incorporating nonparametric statistics into Delphi studies in library and information science. *IR Information Research*, *18*(3). http://www.informationr.net/ir/18-3/paper589.html#.XsvFlS9h1QI

Kim, J. (2020). Academic library’s leadership and stakeholder involvement in research data services. *Proceedings of the Association for Information Science and Technology*, *57*(1). https://doi.org/10.1002/pra2.304

Kim, J., Yakel, E., & Faniel, I. (2019). Exposing Standardization and Consistency Issues in Repository Metadata Requirements for Data Deposition. *College & Research Libraries*, *80*(6), 843–875. https://doi.org/10.5860/crl.80.6.843

King, J. L. (2016). Triangulation from the margins. In G. C. Bowker, S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 339–344). The MIT Press.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publishing.

Kouper, I. (2016). Professional participation in digital curation. *Library & Information Science Research*, *38*, 12.

Kraaikamp, E., Grootveld, M., Hollander, H., & Roorda, D. (2019). On Open Access to Research Data: Experiences and reflections from DANS. *The Grey Journal*, *15*(2), 11.

Kvale, L. (2021a). *Data from a three-phase Delphi study used to investigate Knowledge Infrastructure for Research Data in Norway, KIRDN_Data v.2* [Data set]. OsloMet - Oslo Metropolitan University; Zotero. http://doi.org/10.5281/zenodo.5582714

Kvale, L. (2021b). Using Personas to Visualize the Need for Data Stewardship. *College & Research Libraries*, *82*(3). https://doi.org/10.5860/crl.82.3.332

Kvale, L., & Darch, P. (in press). Privacy protection throughout the research data life cycle. *IR Information Research*.

Kvale, L., & Darch, P. T. (2020). Dealing with privacy – Personal privacy from a research data management perspective. *IConference 2020 Proceedings*. iConference 2020, Borås. http://hdl.handle.net/2142/106564

Kvale, L., & Pharo, N. (2021). Understanding the data management plan as a boundary object through a multi-stakeholder perspective. *International Journal of Digital Curation*, *16*(1). https://doi.org/10.2218/ijdc.v15i1.729

Kvale, S. (2007). *Doing interviews*. SAGE.

Lassen, K. (2019, March 29). Vi bør passe bedre på persondataene våre. *Forskning.no*. https://forskning.no/a/1317418

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, 483–492. http://dl.acm.org/citation.cfm?id=1180950

Lee, J.-S., & Jeng, W. (2019). *The landscape of archived studies in a social science data infrastructure: Investigating the ICPSR metadata records*. 10.

Lexico Dictionaries. (2021). *Infrastructure English Definition and Meaning*. Lexico Dictionaries | English. https://www.lexico.com/en/definition/infrastructure

Limani, F., Latif, A., Borst, T., & Tochtermann, K. (2019). Metadata Challenges for Long Tail Research Data Infrastructures. *Bibliothek Forschung Und Praxis*, *43*(1), 68–74. https://doi.org/10.1515/bfp-2019-2014

Lindsay, C. (2003). From the shadows: Users as designers, producers, marketers, distributors, and technical support. In *How users matter: The co-construction of users and technologies* (pp. 29–50). MIT Press.

Loth, A., Vogel, N., & Sparkes, S. (2019). *Visual Analytics with Tableau*. John Wiley & Sons, Incorporated.

Lund, B. D. (2020). Review of the Delphi method in library and information science research. *Journal of Documentation*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JD-09-2019-0178

Mallasvik, M. L., & Martins, J. T. (2020). Research data sharing behaviour of engineering researchers in Norway and the UK: Uncovering the double face of Janus. *Journal of Documentation*, *77*(2), 576–593. https://doi.org/10.1108/JD-08-2020-0135

Merriam-Webster Dictionary. (2021). *Data*. https://www.merriam-webster.com/dictionary/data

Ministry of Education and Research. (2016). *Kunnskapssektoren sett utenfra* [Rapport]. regjeringen.no. https://www.regjeringen.no/no/dokumenter/kunnskapssektoren-sett-utenfra/id2469493/

Ministry of Education and Research. (2017). *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata [National strategy for access and sharing of research data]* (p. 48). Kunnskapsdepartementet.

Ministry of Education and Research. (2019). *Gjennomgang av NSD [Review of The Norwegian centre for research data]* (p. 96).

Ministry of Education and Research. (2021, February 4). *Regjeringen slår sammen virksomheter til nytt direktorat for høyere utdanning og kompetanse [The government merges organisation for a new directory of higher education and skills]* [Press release]. Regjeringen.no; regjeringen.no. https://www.regjeringen.no/no/aktuelt/direktorat/id2831466/

Molnes, T. E., Egeland, T., Munthe, L., Haraldsen, G., Jahnsen, L. F., Corthay, A., Kolstad, A., Holte, H., Hovig, E., Olweus, J., Kristensen, V., Levy, F. O., Inngjerdingen, M., Loge, J. H., Tønnfjord, G., Bretthauer, M., Kalager, M., Wyller, T. B., Røise, O., … Langmoen, I. (2019, January 7). Hvis vi ikke får dele kunnskap, er det god grunn til å snakke om dødelig personvern. *Aftenposten*. https://www.aftenposten.no/meninger/debatt/i/VRnber/naar-personvern-truer-folkehelsen-32-forskere-ved-oslo-universitetssy

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/ISU-170824

Montgomery, L., & Ren, X. (2018). Understanding Open Knowledge in China: A Chinese Approach to Openness? *Cultural Science Journal*, *10*(1), 17–26. https://doi.org/10.5334/csci.106

Moore, S. A. (2017). A genealogy of open access: Negotiations between openness and access to research. *Revue Française Des Sciences de l'information et de La Communication*, *11*. https://doi.org/10.4000/rfsic.3220

Mulligan, R. (2016). *SPEC Kit 350: Supporting Digital Scholarship (May 2016)*. Association of Research Libraries. https://doi.org/10.29242/spec.350

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14–37.

OECD. (2007). *OECD Principles and guidelines for access to research data from public funding*. OECD. http://www.oecd.org/dataoecd/9/61/38500813.pdf

OED online. (2018). Infrastructure, n. In *Oxford English Dictionary Online*. Oxford University Press. http://www.oed.com/view/Entry/95624

Okerson, A., & O'Donnell, J. J. (Eds.). (1995). *Scholarly journals at the crossroads: A subversive proposal for electronic publishing: an Internet discussion about scientific and scholarly journals and their future* (p. 242). Office of Scientific & Academic Publishing, Association of Research Libraries.

Orlikowski, W. J. (2002). Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science*, *13*(3), 249–273. https://doi.org/10.1287/orsc.13.3.249.2776

Palmer, M. (2006). *Data is the New Oil*. ANA Marketing Maestros. https://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Plomp, E., Dintzner, N., Teperek, M., & Dunning, A. (2019). Cultural obstacles to research data management and sharing at TU Delft. *Insights*, *32*(1), 29. https://doi.org/10.1629/uksg.484

Research Council of Norway, M. D. (2017). *The Research Council of Norway's Policy for Open Access to Research Data* (p. 7).

https://www.forskningsradet.no/contentassets/e4cd6d2c23cf49d4989bb10c5eea087a/the-research-council-of-norways-policy-for-open-access-to-research-data.pdf

Ribes, D., & Jackson, S. J. (2013). Data Bite Man: The Work of Sustaining a Long-Term study. In *"Raw Data" Is an Oxymoron* (pp. 147–166). MIT Press.

Rotondi, A., & Gustafsen, D. (1996). Theoretical, Methodological and Practical Issues Arising out of the Delphi Method. In *Gazing into the Oracle—The Delphi Method and its Application to Social Policy and Pulic Health* (pp. 34–55). Jessica Kingsley.

Ruud, K., Raade, J., & Aspaas, P. P. (2020, August 20). Plan S: Snøballen som må rulle. *Khrono*. https://khrono.no/a/508905

Sands, A. E. (2017). *Managing Astronomy Research Data: Data Practices in the Sloan Digital Sky Survey and Large Synoptic Survey Telescope Projects*. UCLA.

Scroggins, M. J., Pasquetto, I. V., Geiger, R. S., Boscoe, B. M., Darch, T., Cabasse-Mazel, C., Thompson, C., Golshan, M. S., & Borgman, C. (2019). *Thorny Problems in Data (-Intensive) Science* (p. 10) [Publications]. UCLA: Center for Knowledge Infrastructures. https://escholarship.org/uc/item/31b1z69c.

Springer, R., & Cooper, D. (2021). Data Communities: Empowering Researcher-Driven Data Sharing in the Sciences. *International Journal of Digital Curation*, *16*(1), 7. https://doi.org/10.2218/ijdc.v15i1.695

Star, & Griesemer, J. R. (1989). Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001

Star, S. L. (1989). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. *Distributed Artificial Intelligence*. https://doi.org/10.1016/B978-1-55860-092-8.50006-X

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, *35*(5), 601–617.

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134.

Star, & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work (CSCW)*, *8*(1–2), 9–30.

Svarstad, J. (2019). – Hvis Forskningsrådet ønsker å bevare tillit, bør de ta en pause og lytte [If the Research Council wishes to maintain trust they should take a break and listen]. *Forskerforum*. https://www.forskerforum.no/hvis-forskningsradet-onsker-a-bevare-tillit-bor-de-ta-en-pause-og-lytte/

Sylvain, O. (2019). Recovering tech´s humanity. *Columbia Law Review Forum*, *119*(7), 252–282.

Tapio, P., Paloniemi, R., Varho, V., & Vinnari, M. (2011). The unholy marriage? Integrating qualitative and quantitative information in Delphi processes. *Technological Forecasting and Social Change*, *78*(9), 1616–1628. https://doi.org/10.1016/j.techfore.2011.03.016

Tenopir, C., Allard, S., Pollock, D., Hughes, D., Lundeen, A., Sandusky, R. J., & Baird, L. (2019). Academic Librarians and Research Data Services: Attitudes and Practices. *IT Lib: Information Technology and Libraries Journal*, *1*. https://itlib.cvtisr.sk/wp-content/uploads/docs/24_academic%20librarians.pdf

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. J., & Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, *27*(1), 23–44. https://doi.org/10.18352/lq.10180

Teperek, M., Cruz, M. J., Verbakel, E., Böhmer, J., & Dunning, A. (2018). Data Stewardship addressing disciplinary data management needs. *International Journal of Digital Curation*, *13*(1), 141–149. https://doi.org/10.2218/ijdc.v13i1.604

Lov om behandling av personopplysninger, LOV-2018-06-15-38 (2018). https://lovdata.no/dokument/LTI/lov/2018-06-15-38

The Research Council of Norway. (2020). *Forskningsrådets policy for åpen forskning [Policy for open science]* [Policy]. Norges forskningsråd. https://www.forskningsradet.no/siteassets/tall-og-statistikk-seksjonen/apen-forskning/nfr-policy-apen-forskning-norsk-ny.pdf

The Research Council of Norway. (2021). *Hvordan skal vi dele forskningsdata? -Utredning og anbefalinger om lisensiering og tilgjengeliggjøring [How should we share research data—Recommendation regarding licensing and access to research data]* (p. 67). Forskningsrådet. https://www.forskningsradet.no/siteassets/publikasjoner/2021/rettighets--og-lisenssporsmal-ved-tilgjengeliggjoring-av-forskningsdata.pdf

Thompson, C. A. (2017). *Data expertise and service development in geoscience data centres and academic libraries*. University of Illinois at Urbana-Champaign.

Van House, N. A. (2004). Science and technology studies and information studies. *Annual Review of Information Science and Technology*, *38*(1), 1–86. https://doi.org/10.1002/aris.1440380102

Verhaar, P., Schoots, F., Sesink, L., & Frederiks, F. (2017). Fostering effective data management practices at Leiden university. *LIBER Quarterly*, *27*(1), 1–22. Scopus. https://doi.org/10.18352/lq.10185

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99.

Wakeford. (2016). Dont´t go all the way. In G. C. Bowker, S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 69–83). The MIT Press.

Whyte, A. (2012). Emerging infrastructure and services for research data management and curation in the UK and Europe. In G. Pryor (Ed.), *Managing research data* (pp. 17–46). Facet.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D. R., Los, W., & et al. (2010). *Riding the wave—How Europe can gain from the rising tide of scientific data* - (p. 38) [Final report of the High Leverl Expert Group on Scientific Data]. European Union. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change*, *40*(2), 131–150. https://doi.org/10.1016/0040-1625(91)90002-W

Wu, S., & Worrall, A. (2019). Supporting successful data sharing practices in earthquake engineering. *Library Hi Tech*, *37*(4), 764–780. https://doi.org/10.1108/LHT-03-2019-0058

Yan, A., Huang, C., Lee, J., & Palmer, C. L. (2020). Cross-disciplinary data practices in earth system science: Aligning services with reuse and reproducibility priorities. *Proceedings of the Association for Information Science and Technology*, *57*(1). https://doi.org/10.1002/pra2.218

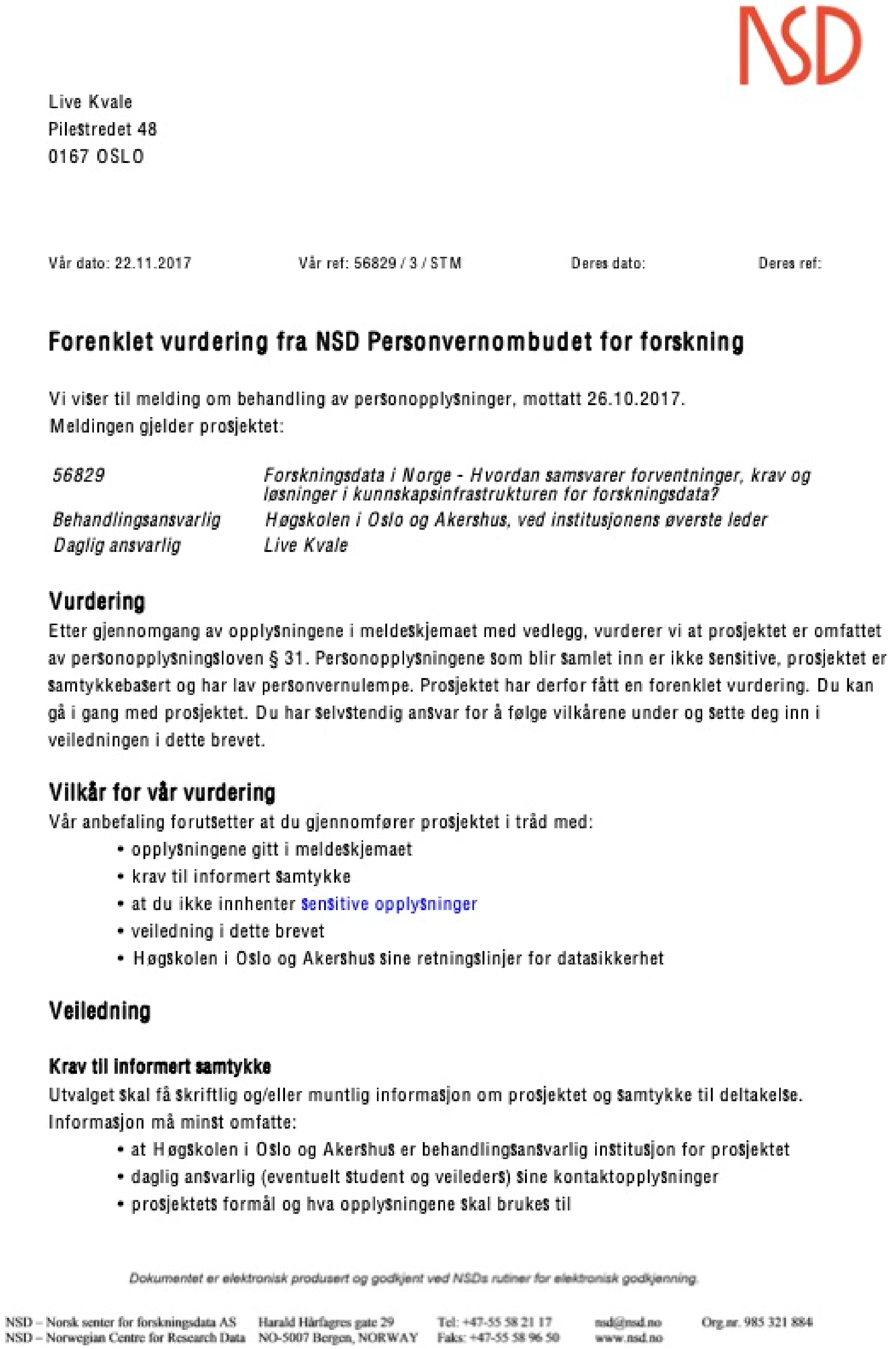
Yoon, A., & Kim, Y. (2020). The role of data-reuse experience in biological scientists' data sharing: An empirical analysis. *The Electronic Library*, *38*(1), 186–208. https://doi.org/10.1108/EL-06-2019-0146

Ziglio, E. (1996). The Delphi Method and its Contribution to Decision-Making. In M. Adler & E. Ziglio (Eds.), *Gazing into the Oracle—The Delphi Method and its Application to Social Policy and Public Health* (pp. 3–33). Jessica Kingsley.

Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, *49*, 228–241. https://doi.org/10.1016/j.ijinfomgt.2019.05.024

Appendices

Appendix A: Approval personal data processing



NSD

Live Kvale
Pilestredet 48
0167 OSLO

Vår dato: 22.11.2017 | Vår ref: 56829 / 3 / STM | Deres dato: | Deres ref:

Forenklet vurdering fra NSD Personvernombudet for forskning

Vi viser til melding om behandling av personopplysninger, mottatt 26.10.2017.
Meldingen gjelder prosjektet:

56829	*Forskningsdata i Norge - Hvordan samsvarer forventninger, krav og løsninger i kunnskapsinfrastrukturen for forskningsdata?*
Behandlingsansvarlig	*Høgskolen i Oslo og Akershus, ved institusjonens øverste leder*
Daglig ansvarlig	*Live Kvale*

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet med vedlegg, vurderer vi at prosjektet er omfattet av personopplysningsloven § 31. Personopplysningene som blir samlet inn er ikke sensitive, prosjektet er samtykkebasert og har lav personvernulempe. Prosjektet har derfor fått en forenklet vurdering. Du kan gå i gang med prosjektet. Du har selvstendig ansvar for å følge vilkårene under og sette deg inn i veiledningen i dette brevet.

Vilkår for vår vurdering

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

- opplysningene gitt i meldeskjemaet
- krav til informert samtykke
- at du ikke innhenter sensitive opplysninger
- veiledning i dette brevet
- Høgskolen i Oslo og Akershus sine retningslinjer for datasikkerhet

Veiledning

Krav til informert samtykke

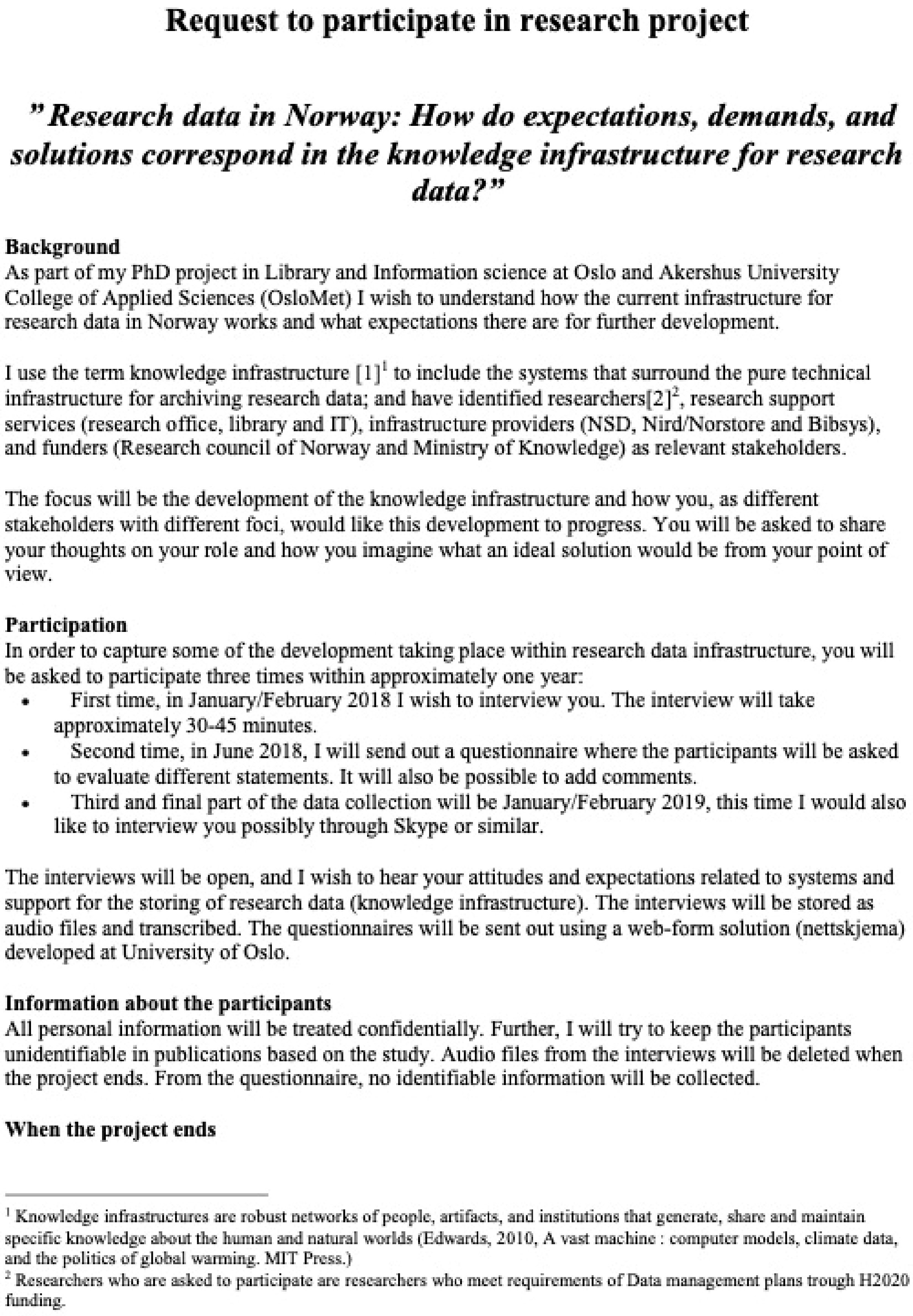
Utvalget skal få skriftlig og/eller muntlig informasjon om prosjektet og samtykke til deltakelse.
Informasjon må minst omfatte:

- at Høgskolen i Oslo og Akershus er behandlingsansvarlig institusjon for prosjektet
- daglig ansvarlig (eventuelt student og veileders) sine kontaktopplysninger
- prosjektets formål og hva opplysningene skal brukes til

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Appendix B: Information letter and consent form one



Request to participate in research project

"Research data in Norway: How do expectations, demands, and solutions correspond in the knowledge infrastructure for research data?"

Background
As part of my PhD project in Library and Information science at Oslo and Akershus University College of Applied Sciences (OsloMet) I wish to understand how the current infrastructure for research data in Norway works and what expectations there are for further development.

I use the term knowledge infrastructure [1][1] to include the systems that surround the pure technical infrastructure for archiving research data; and have identified researchers[2][2], research support services (research office, library and IT), infrastructure providers (NSD, Nird/Norstore and Bibsys), and funders (Research council of Norway and Ministry of Knowledge) as relevant stakeholders.

The focus will be the development of the knowledge infrastructure and how you, as different stakeholders with different foci, would like this development to progress. You will be asked to share your thoughts on your role and how you imagine what an ideal solution would be from your point of view.

Participation
In order to capture some of the development taking place within research data infrastructure, you will be asked to participate three times within approximately one year:

- First time, in January/February 2018 I wish to interview you. The interview will take approximately 30-45 minutes.
- Second time, in June 2018, I will send out a questionnaire where the participants will be asked to evaluate different statements. It will also be possible to add comments.
- Third and final part of the data collection will be January/February 2019, this time I would also like to interview you possibly through Skype or similar.

The interviews will be open, and I wish to hear your attitudes and expectations related to systems and support for the storing of research data (knowledge infrastructure). The interviews will be stored as audio files and transcribed. The questionnaires will be sent out using a web-form solution (nettskjema) developed at University of Oslo.

Information about the participants
All personal information will be treated confidentially. Further, I will try to keep the participants unidentifiable in publications based on the study. Audio files from the interviews will be deleted when the project ends. From the questionnaire, no identifiable information will be collected.

When the project ends

---

[1] Knowledge infrastructures are robust networks of people, artifacts, and institutions that generate, share and maintain specific knowledge about the human and natural worlds (Edwards, 2010, A vast machine : computer models, climate data, and the politics of global warming. MIT Press.)

[2] Researchers who are asked to participate are researchers who meet requirements of Data management plans trough H2020 funding.

The project is estimated to end primo 2021. With your consent, I wish to be able to store the transcribed interviews and the results from the questionnaire in a relevant open data archive. You will be given the possibility to review and "black out" elements or claims you do not wish to be shared or withdraw consent prior to data publication. Audio files from the interviews will be deleted when the project ends.

Voluntary participation

Participation in the study is voluntary, and you can withdraw you consent without giving any reason at any time.

Below is a consent form, if you wish to participate I ask you to sign by filling in your name and returning the form to me.

The study is reported to Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS. http://www.nsd.uib.no/personvernombud/en/index.html

Please get in touch if you have any questions.

Best Regards
Live Kvale
Stipendiat
Institutt for arkiv-, bibliotek-, og informasjonsfag
Fakultetet for samfunnsvitenskap
Høgskolen i Oslo og Akershus
Tlf: +47 40 04 74 17
live.kvale@hioa.no

Consent to participate in the study

If you wish to participate, sign by filling in name and date below.

I have received information about the study and am willing to participate.

(Signed by participant, date)

If you wish to give permission to the sharing of transcripts of the interviews and responses given in the questionnaire after anonymization and final consent from you, sign by filling in name and date below.

I consent to sharing of anonymized interview and questionnaire data in a relevant data archive.

(Signed by participant, date)

Appendix C: Consent form two

Consent form knowledge infrastructure for research data

What is your e-mail address?
Your e-mail address will only be used to link your response to your participant code. The e-mail adress will be deleted once your response is exported and linked with your participant code.

Please select preferred language
Norsk
English

Etter å ha lest og analysert intervjuene i flere runder, ser jeg at flere av intervjuene inneholder informasjon som gjør deltageren relativt enkelt identifiserbar for noen med kjennskap til deltageren, eller gjennom sammenstilling med annen lett tilgjengelig informasjon.
Dersom du ikke er komfortabel med at du kan være mulig å identifisere ber jeg deg trekke tidligere gitt samtykke (under).
Resultatene fra spørreskjemaet er i mindre grad identifiserbare, siden resultatene er aggregert i grupper (finansiør, forsker, bibliotek, infrastruktur osv.). Hver gruppe består av 3-4 deltagere, totalt er det 24 deltagere i prosjektet.

After having read and analyzed the interviews multiple times, I see that several interviews contain information that makes some participants identifiable for someone with knowledge of the participant or by combining it with other available information.
If you are not comfortable with the possibility to be identified I advise you to redraw previously given consent to share the interviews (below).
The results from the questionnaire are less identifiable, as the results are aggregated in groups (funder, researcher, librarian, infrastructure provider etc.) Every group contain 3-4 participants, with 24 participants in total.

Dersom du svarer ja på spørsmål 1 og 2, velger du hva som kan publiseres i spørsmål 4.

If you answer yes to question 1. and 2., you select what material can be published in question 4.

1. Jeg har hatt mulighet til å se over materialet som deles og fjerne opplysninger jeg ikke ønsker å dele. (J/N)

1. I have had the possibility to review the material shared and remove any information that I do not wish to share openly. (Y/N)

2. Jeg er innforstått med at en publisering i relevant arkiv også innebærer at materialet ikke vil være mulig å fjerne på et senere tidspunkt (J/N)

2. I understand that publication of the data implies that it will not be possible to remove the material post publication (Y/N)

3. Jeg er innforstått med at til tross for at direkte personidentifiserende informasjon som navn og arbeidssted er fjernet, så kan det være mulig å identifisere meg basert på informasjonen som ligger i intervjuene. (J/N)

3. I understand that despite removal of directly identifiable information such as name and workplace, it could still be possible to identify me based on the information in the interviews. (Y/N)

Jeg samtykker til at følgende innsamlede data publiseres:
Intervju gjort vinteren 2018 (J/N)
Spørreskjema av høsten 2018 (J/N)
Intervju av vinteren 2019 (J/N)
Dette samtykket (pseudonymisert) av våren 2019 (J/N)

4. I consent to publication of the following data:
Interviews from winter 2018 (Y/N)
Questionnaire from autumn 2018 (Y/N)
Interviews from spring 2019 (Y/N)
This consent (pseudonymized) from summer 2019 (Y/N)

Hva syns du om å være deltager i denne studien?
Metoden jeg har brukt i studien, og da særlig bruken av samtykke for å dele pseudonymiserte intervjudata, blir en viktig del i min avhandling. Dersom du har noen refleksjoner om det å være deltager i studien vil jeg gjerne høre om det, bruk feltet under.

What did you think of participating in this study?
The method I used in this study, and in particular the use of consent for sharing pseudonymized interview data, will be a central part of my dissertation. If you have any reflection on your experience of being a participant in the study, I would like to hear about this, please use the space below.

Tusen takk for din tid og for deltagelse i denne studien.

Thank you for your time and participation in this study.

Appendix D: Interview guide one

Intervjuguide

Husk at om de ikke svarer så ikke svar for de!!! La de få tid!!! Evt gå videre!!!!!

Introduksjon

- Hvem jeg er
- Hvorfor jeg er der/ Forklare hensikten med intervjuet
- Hva jeg vil/ Antyde hva jeg er interessert i få vite noe om
- Ingen eksamen, lov å si at dette vet jeg ikke noe om, eller har jeg ikke noen mening om – for meg er også det verdifulle svar.

Hva som skal skje med materialet/ Anonymitet og datadeling

- Lydopptaket vil bli transkribert (lydopptaket slettes ved prosjektslutt)
- Ønsker å kunne lagre transkriberingen i et åpent arkiv
- Mulighet til å lese over og evt «svarte ut» elementer som ikke ønskes delt, eller kommentere om det ønskes vil bli gitt så snart transkriberingen er gjort.
  Alt som blir sagt håndteres fortrolig inntil du har hatt muligheten til å se over.
- Informere om at det er. mulig å trekke seg fra intervjuet, eller når som helst senere i studien.

1. **Innledning:**
   1.1 Kan du beskrive på hvilken måte du jobber med forskningsdata?
   1.2 Hvilke utfordringer møter du?

2. **Forskningsdata:**
   2.1 I konteksten nye krav om deling og lagring av forskningsdata, hvordan vil du beskrive dine holdninger?
      - Stikkord: resurs, tidkrevende, merarbeid, muligheter

3. **Krav:** Det stilles stadig nye krav til hvordan forskningsdata skal håndteres.
   3.1 «Så åpne som mulig så lukkede som nødvendig» står det om forskningsdata i den nasjonale strategien for tilgjengeliggjøring og deling av forskningsdata[7]. Kan du reflektere litt rundt hva som menes med dette?
   3.2 Hvordan påvirker krav om datahåndtering ditt arbeid?

---

[7] https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-tilgjengeliggjoring-og-deling-av-forskningsdata/id2582412/sec1

3.3 Det refereres mye til FAIR [8] prinsippene som innebærer at data skal være: Findable, Accessible, Interoperable and Reusable hva tenker du om disse i forhold til måten du jobber på (og dine forskningsdata)?

4 Hvor mye og hvor lenge:

Et tilbakevendende spørsmål er gjerne hva skal bevares, og for hvor lenge.

4.1 Har du tanker eller forslag til hvordan dette kan løses?

- Bruke gjerne eksempler fra eget fagområde og tenk gjerne på hvem som bør ha ansvar for hva her.

5 Kunnskap og opplæring:

5.1 Til hvilke ressurser eller kanaler henvender du deg for å være oppdatert på krav til lagring og deling av forskningsdata?

- Stikkord: nettsider, konferanser, kolleger, offentlige dokumenter osv.

5.2 I hvilken grad vil du si at resursene du trenger er tilstrekkelig godt organisert og tilgjengelige?

5.3 Hvem har ansvaret for at forskere har tilgang på tilstrekkelig kompetanse på organisering av data med tanke på tilgjengeliggjøring?

- Stikkord: Fagmiljøer, dataarkivene, forskerutdanningen, bibliotek, forskningsrådgivere, IT

6 Roller og ansvar

6.1 Hva er din rolle i kunnskapsinfrastrukturen[9] for forskningsdata?

- dataprodusent – rådgiver -teknikker – finansiør, annet

6.2 Hvordan ser du din rolle i forhold til andre som jobber med forskningsdata?

6.3 Hvilke andre er mener du er viktige aktører i en kunnskapsinfrastruktur for forskningsdata?

- Utdyp gjerne hvilke ansvarsområder de ulike har og hvorfor du ser det slik.

7 Inndelt etter funksjon, om den som intervjues har flere funksjoner stilles spørsmål fra flere kategorier

7 a) Forsker

7.1 Utfra hvordan du jobber, hva er dine infrastruktur behov for forskningsdata?

---

[8] https://www.nature.com/articles/sdata201618

[9] Knowledge infrastructures are robust networks of people, artifacts, and institutions that generate, share and maintain specific knowledge about the human and natural worlds (Edwards, 2010, A vast machine : computer models, climate data, and the politics of global warming. MIT Press.) For å se hvordan begrepet benyttes anbefaler jeg rapporten Knowledge «Infrastructures: Intellectual Frameworks and Research Challenges» http://hdl.handle.net/2027.42/97552

7.2 Har du laget en datahåndteringsplan for ett eller flere av dine prosjekter? Fortell gjerne hvorfor (ikke) og del eventuelle erfaringer med å lage en slik plan.

7.3 Om du er kjent med datatidsskrifter/data journals og data papers som måte å publisere data på, hva tenker du om dette?

7 b) Finansiør

7.1. Hvilken linje bør Norge ligger på i forhold til krav om deling? Utdyp gjerne hvorfor

- Stikkord: Foregangsland, på linje med EU, litt etter EU, Norden, UK, andre?

7.2 Premiering av de som er flinke til å gjøre sine data tilgjengelig for andre kommer stadig opp som tema, hva er dine tanker om dette?

7.3 Om du er kjent med datatidsskrifter/data journals og data papers som måte å publisere data på, hva tenker du om dette?

7.4 Hva ser du som hensikten med en datahåndteringsplan?

7 c) Forskerstøtte

7.1 Hvordan er oppgavefordelingen og samarbeid mellom dere og nasjonale infrastrukturtilbydere?

7.2 Hva er det viktigste dere gjør for å lette arbeidet med forskningsdata for forskerne ved deres institusjon?

7.3 Hva ser du som hensikten med en datahåndteringsplan?

7 d) Infrastruktur

7.1 Hva vil du si er viktigst i tjenestene dere leverer?
7.2 Hvordan er oppgavefordeling og samarbeid mellom dere og forskerstøtte ved institusjonene?
7.3 Hvilke infrastrukturbehov dekker dere ikke i dag?
7.4 Dekkes disse behovene av andre aktører, eller er det planer for å løse disse behovene?
7.5 Hva ser du som hensikten med en datahåndteringsplan?

8 Fordeler og ulemper

8.1 Hvilke ulemper ser du med deling av forskningsdata?
8.2 Hvilke fordeler ser du med deling av forskningsdata?

9 Infrastruktur

9.1 Hvordan vil du beskrive den tekniske infrastruktur/lagringsløsninger for forskningsdata i Norge slik den er i dag?
9.2 (spørsmål til finansiør og infrastruktur) Har det nye tjenesteorganet en rolle?

10 **Drømmescenario**

10.1 Hvis du skulle få skissere opp en drømmeløsning for forskningsdata hvordan skulle denne være?

10.2 Hvilke krav bør stilles de som benytter denne for lagring av data? (metadata, lisenser, publisering)

10.3 Skulle forskere være pålagt å benytte denne?

10.4 For at data skal være forståelige for andre og mulig å finne igjen trengs metadata/beskrivelser av dataene, hvordan bør dette løses?

- Hvem har ansvaret for å registrere metadata?

10.5 I hvilken grad mener du at det er nødvendig med kontroll/gjennomgang av data og metadata etter at de er lagt inn?

10.6 Hvordan bør sitering foregå? (hva og hvem skal krediteres)

10.7 Hvordan ser du for deg at support bør fungere?

10.8 Ser du det som mulig at dagens løsninger kan møte disse ønskene?

10.9 Hva mener du er de viktigste funksjonene i en slik infrastruktur, og hvorfor?

Informere om oppfølging

Mulighet til å se over det transkriberte intervjuet (sendes ut senest juni)

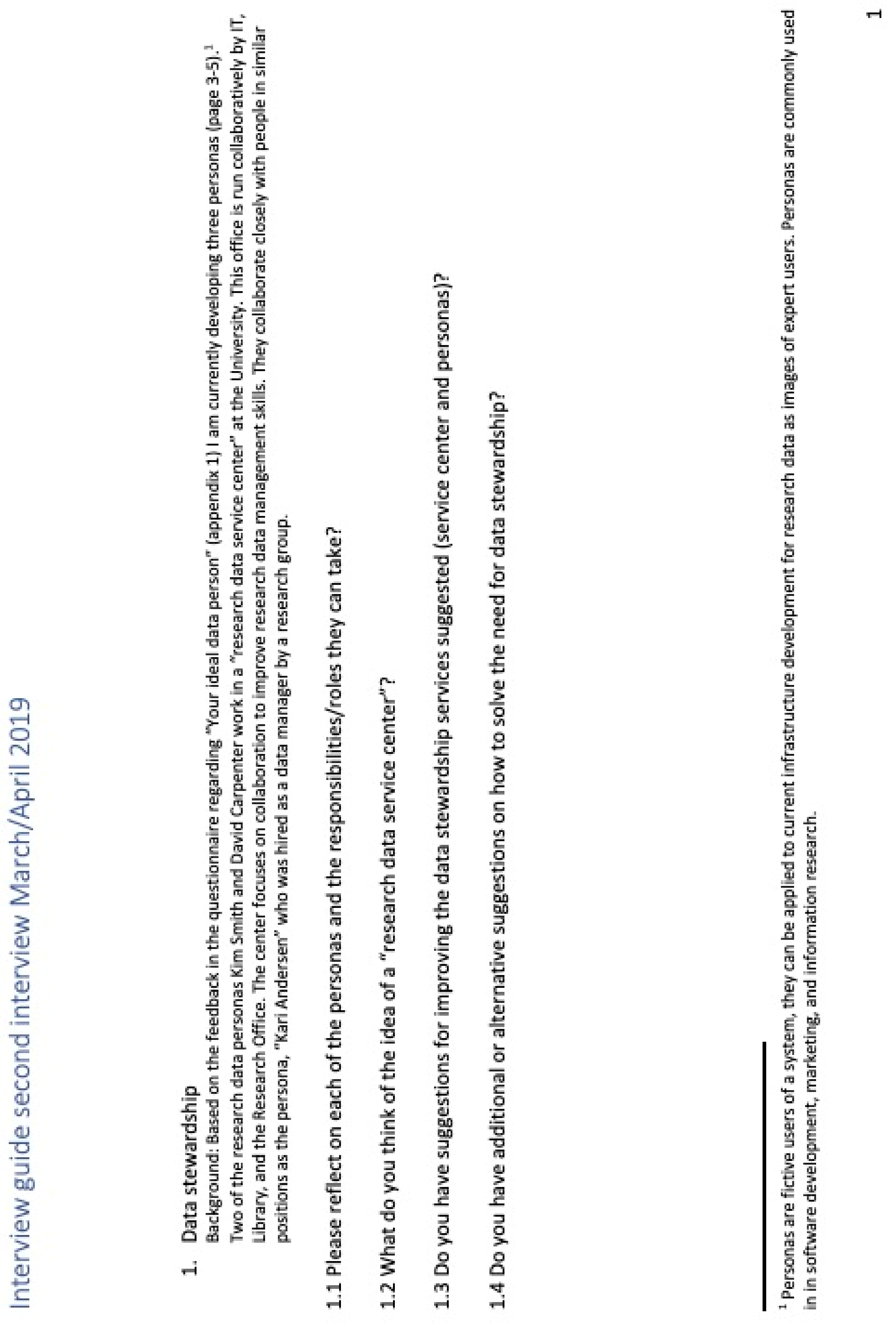
Spørreskjema med påstander aprox. Juni

Nytt intervju våren 2019 - skype?

Ta kontakt om du ønsker å tilføye noe

Er det noe du kommer på i etterkant at du gjerne skulle ha sagt så er det bare å sende meg mail eller ringe.

Appendix E: Interview guide two



Interview guide second interview March/April 2019

1. Data stewardship
   Background: Based on the feedback in the questionnaire regarding "Your ideal data person" (appendix 1) I am currently developing three personas (page 3-5).[1]
   Two of the research data personas Kim Smith and David Carpenter work in a "research data service center" at the University. This office is run collaboratively by IT, Library, and the Research Office. The center focuses on collaboration to improve research data management skills. They collaborate closely with people in similar positions as the persona, "Kari Andersen" who was hired as a data manager by a research group.

1.1 Please reflect on each of the personas and the responsibilities/roles they can take?

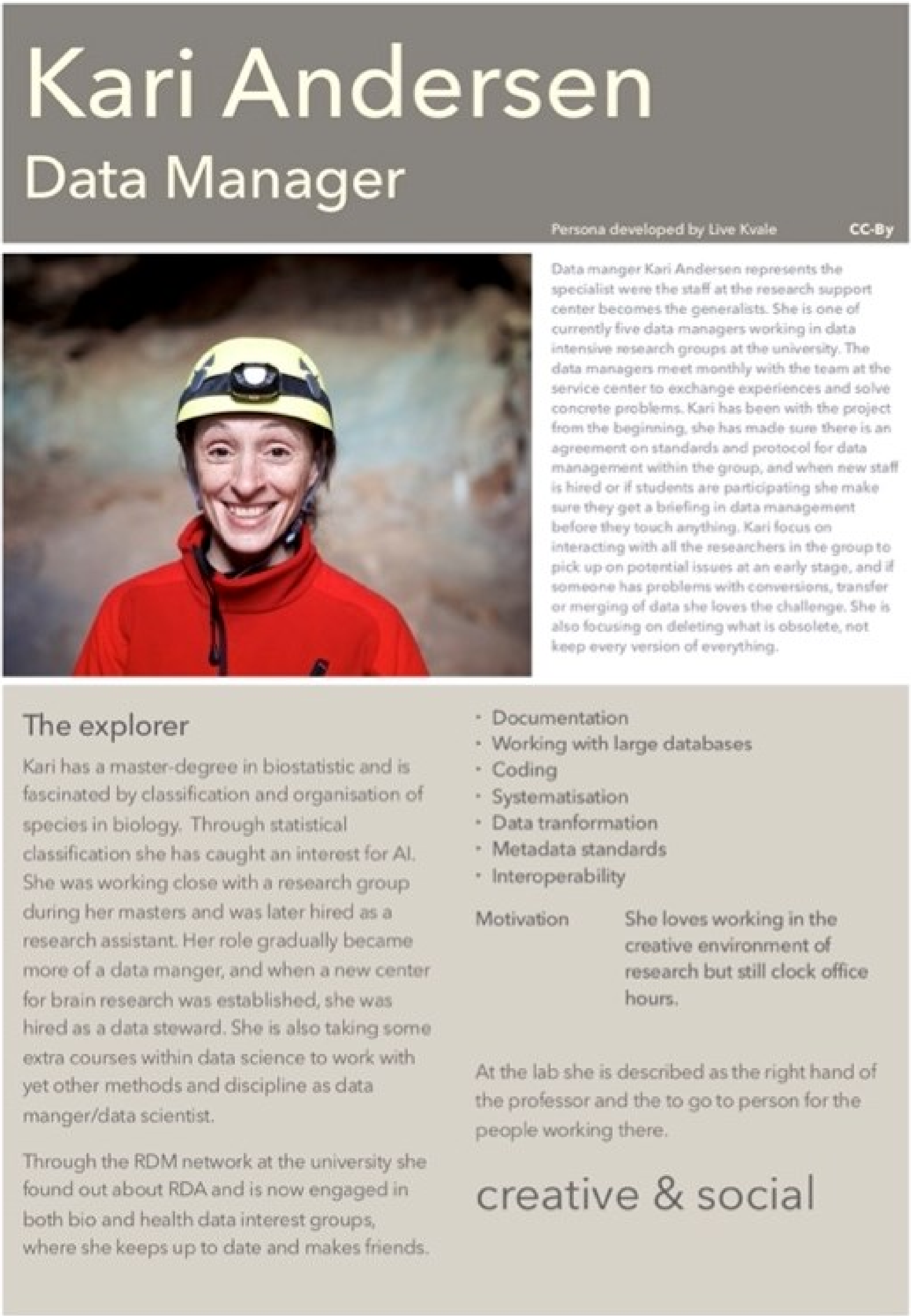
1.2 What do you think of the idea of a "research data service center"?

1.3 Do you have suggestions for improving the data stewardship services suggested (service center and personas)?

1.4 Do you have additional or alternative suggestions on how to solve the need for data stewardship?

---

[1] Personas are fictive users of a system, they can be applied to current infrastructure development for research data as images of expert users. Personas are commonly used in in software development, marketing, and information research.



Kari Andersen

Data Manager

Persona developed by Live Kvale

Data manger Kari Andersen represents the specialist were the staff at the research support center becomes the generalists. She is one of currently five data managers working in data intensive research groups at the university. The data managers meet monthly with the team at the service center to exchange experiences and solve concrete problems. Kari has been with the project from the beginning, she has made sure there is an agreement on standards and protocol for data management within the group, and when new staff is hired or if students are participating she make sure they get a briefing in data management before they touch anything. Kari focus on interacting with all the researchers in the group to pick up on potential issues at an early stage, and if someone has problems with conversions, transfer or merging of data she loves the challenge. She is also focusing on deleting what is obsolete, not keep every version of everything.

The explorer

Kari has a master-degree in biostatistic and is fascinated by classification and organisation of species in biology. Through statistical classification she has caught an interest for AI. She was working close with a research group during her masters and was later hired as a research assistant. Her role gradually became more of a data manger, and when a new center for brain research was established, she was hired as a data steward. She is also taking some extra courses within data science to work with yet other methods and discipline as data manger/data scientist.

Through the RDM network at the university she found out about RDA and is now engaged in both bio and health data interest groups, where she keeps up to date and makes friends.

- Documentation
- Working with large databases
- Coding
- Systematisation
- Data tranformation
- Metadata standards
- Interoperability

Motivation	She loves working in the creative environment of research but still clock office hours.

At the lab she is described as the right hand of the professor and the to go to person for the people working there.

creative & social

David Carpenter

Data Curator

Persona developed by Live Kvale

Research data management service coordinator "Kim Smith" and data curator "David Carpenter" work in a "research service center" at the University. This office is run collaboratively by IT, Library, and the Research Office, Kim and David are two of three full time employees in the center. The center also employs data science students and PhDs working there part time. Staff from the Research Office, Library and IT are also associated with the center, making it an interdisciplinary environment focusing on collaboration and research data management. The center has been established to solve the issues of RDM support, but also deals with other related research skills, such as citation, data analysis software, support on statistics etc. They collaborate closely with the data protection officer and with a network of data managers such as "Kari Andersen" who is hired by a research group.

The analyst

David has a PhD in computational linguistics and many years of data intensive research. Recently he has gotten an accredited course in data stewardship offered by GoFAIR office. David has a scientific oriented, analytical mindset. He was engaged several years in data driven research but he got more interested in the challenges related to ontologies metadata definitions and less interested in the scientific topic and the final publication. Thanks to his research background he is able to read through the lines and understand the researchers and that priorities of a researcher. David is good at convincing them that a by-product of a proper data management planning will be more citing and therefore more accreditations!

- Systematisation
- Making data FAIR
- Metadata, documentation and provenance
- Data archives and archiving
- Coding and programming
- Data mining
- Formatting and data transformation

Motivation:	David enjoys translating between disciplines, understanding their needs and solve problems

David loves research and the university as a work environment but he prefers working with the data rather than publishing.

accurate & structured

Kim Smith
RDM service coordinator

Persona developed by Live Kvale

Research data management service coordinator "Kim Smith" and data curator "David Carpenter" work in a "research service center" at the University. This office is run collaboratively by IT, Library, and the Research Office, Kim and David are two of three full time employees in the center. The center also employs data science students and PhDs working there part time. Staff from the Research Office, Library and IT are also associated with the center, making it an interdisciplinary environment focusing on collaboration and research data management. The center has been established to solve the issues of RDM support, but also deals with other related research skills, such as citation, data analysis software, support on statistics etc. They collaborate closely with the data protection officer and with a network of data managers such as "Kari Andersen" who is hired by a research group.

The idealist

Kim Smith is responsible for the data management services at the university. Through a series of workshops held at the center she has given researchers and master students their first "plan for you data before it is lost" course. She also advice on privacy and copyright issues, while she does not have a background in law, experience has made her able to advice on many of the issues that occur, when in doubt she consults the data protection officer. Kim is also responsible for reviving and approving DMPs even if the workload is shared and the planes reviewed collaboratively at the center. Trough DMP review she is able to identify potential challenges at an early stage. Kim likes to teach, she hold several of the RDM training courses offered at the university and is active in RDA policy groups.

- Communication and interpretation
- Policy expert
- Research ethics & personal privacy
- IP- law
- DMP
- Metadata

Motivation	Builds unique knowledge in the organisation and contribute to making research transparent and verifiable

Kim genuinely believe open science will make the world a better place and that proper data management can solve the reproduction crisis.

structured & strategic

Due to inconsistent responses would like you to answer this question from the questionnaire again. Alternative n. and o. have now been added:

Please select what you find to be **the five** most important reasons to make a data management plan (DMP)

Please select **no more than 5** of the statements below and **range them 1-5 by importance**

a. To create awareness in the research community for the need for data stewardship.
b. By being in control of their own data management the research gets better and more efficient.
c. DMP gives the universities an overview of the ongoing research projects
d. When the researchers think about what they are doing at an early stage, they can make intelligent choices for their data.
e. The DMP provide the archives with information (metadata) they need when data is deposited.
f. DMP gives the archive the possibility to plan for data that are going to be deposited there.
g. It makes the researcher think about how to make their data and metadata FAIR (Finable, Accessible, Interoperable, Reusable).
h. A DMP creates awareness and agreements on data procedures within a research group, in particular important when several partners with different data needs are involved.
i. A DMP shows what data is produced in a project when the project is finished.
j. A DMP is primarily a tool for those who take care of the data after it has left the project.
k. A DMP is a planning with a commitment to making the data as open as possible.
l. A DMP makes researchers aware of the value of what they are collecting.
m. Other reasons
n. I don’t see why DMPs are important.
o. I am not familiar with data management plans.

2. Data management plans

2.1 How can the data management plan best serve the needs of the different users visualized in figure 2?

In my results, researchers working in collaborations and researchers working individually, report different needs for data management planning. Research teams refer to the DMP as an agreement on standards and formats within the group, whereas researchers working individually do not report this need.

2.2 How can the plan itself reflect different needs?

If you are familiar with the NDS data management plan or Sigma2 Easy DMP:

2.3. To what extent would you say that "NDS data management plan" and "Sigma2 Easy DMP" reflect the users (fig 2) and their needs?

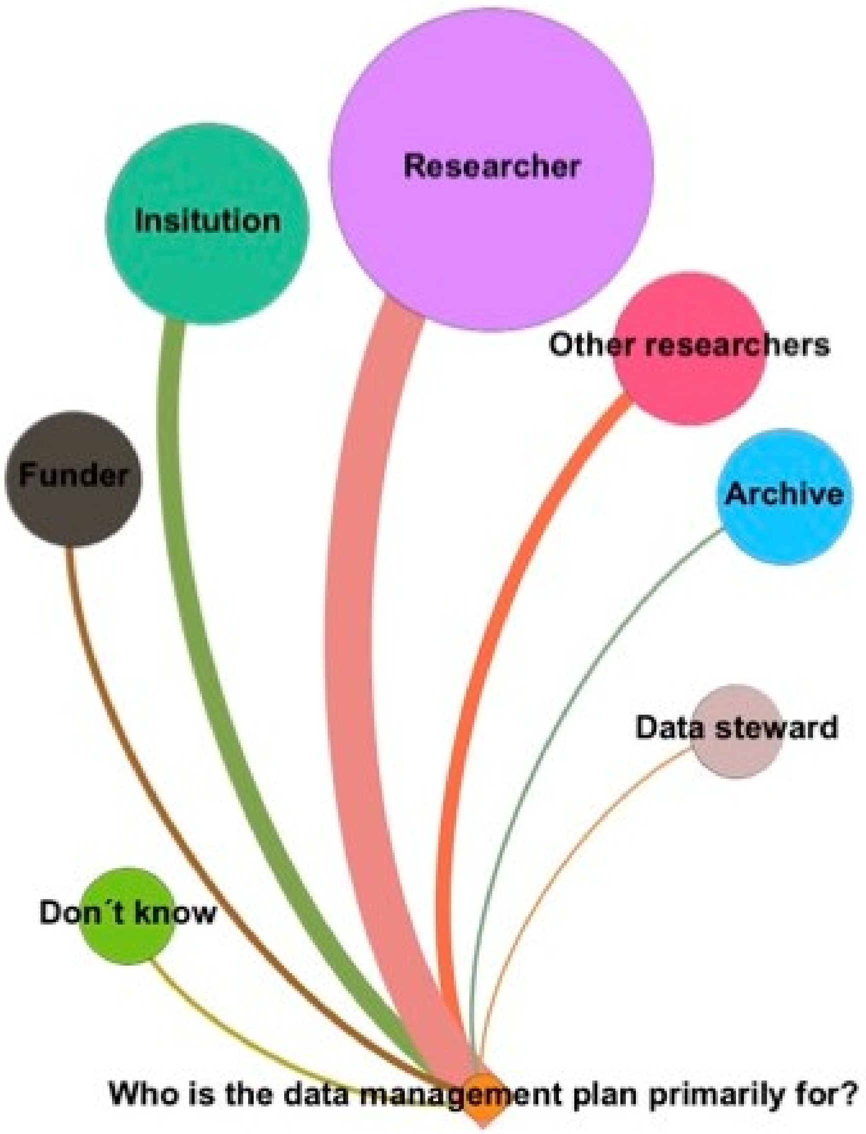


Figure 2

3. Personal privacy and ethics of data sharing

3.1 Please reflect on the relation between the following issues in the contexts of data sharing:

1-The need for high quality scholarship
2-Personal privacy
3- Public trust in research
4-Maximizing profit of investments in research

3.2 How are research ethics and privacy keeping up to date with technology?

3.3 How do you see the balance between research practice, personal privacy law and research ethics? (in the context of fig 3)
(Q2b: 5 of 8 researchers questioned agree that there should be more room for putting privacy aside in order to conduct important research.)

3.4 How do National or regional norms and regulations for privacy work in practice in a global research environment?

3.5 Q2D (fig 4) has two parts, anonymity and embedded privacy could you reflect on both aspects of this question?
(The 4 researchers who work individually does not agree, 3 disagree, 1 neutral)

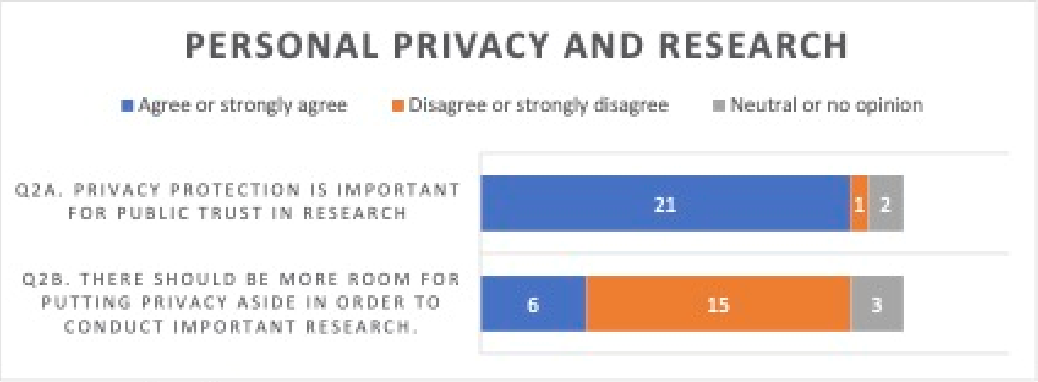


Figure 3

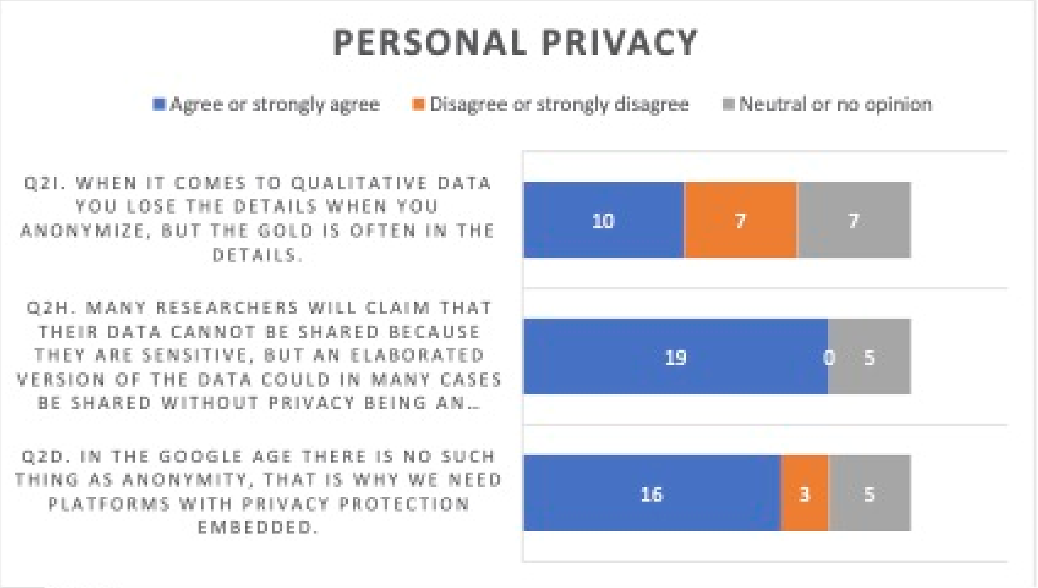


Figure 4

4. Have you since we last spoke (primo 2018) have experienced any changes or developments in the knowledge infrastructure for research data in Norway?

   4.1 What are your views on these changes?
   4.2 What impact do you think these changes have on the road ahead?

   4.3 Have you experienced or noticed any cultural changes the last year regarding open science in general?
   4.4 What about data sharing specifically?

Thank you for your participation. Once I have completed the transcriptions you will get the text for review.

The initial consent form given for this study needs to be revised. I have asked for permission to publish data after "anonymization and final consent from you", anonymization of the interviews have shown to be difficult, even if directly identifiable information such as name and workplace is removed, triangulation and/or familiarity with the subject will in many cases make identification possible.

I will therefore send you an updated consent form, along with the data from both of the interviews and the questionnaire, as soon as transcription of this interview is completed. There I will ask you to decide again whether the information you have given can be published in a relevant open archive when I publish the results from this study. You will have the possibility to share only parts of the information or all of it. Your decision on what to share does not affect the results from my study in any way.

Live Kvale

Appendix 1. Responses to question 6 regarding research data support person. (mixed English Norwegian according to respondents)

If you do not see the need for such a position, please give a short explanation	Jeg ser ikke helt hensikten med å ha en egne rolle for dette på institusjonene (som er slik jeg tolker spørsmålet). Den nødvendige (totale) kompetansen bør finnes hos forskeren, forskerstøttepersonell og hos arkivene/infrastrukturene i fellesskap; Jeg tror ikke man trenger en person, man trenger mange som har ulik kompetanse. En type person som typisk vil være en datakurator/data scientist/data røkter vil kunne jobbe med teknisk håndtering av data eks. på relevant infrastruktur, ev. også være involvert i fasilitering av og gjennomføring av analyse i samarbeid med forsker. Andre vil ha fokus på mer administrativt relaterte ting, lovverk, organisasjonsreglement, krav som stilles til forsker mm. Og kan bidra til å hjelpe at forsker er i stand til å følge opp disse. En trenger kanskje en annen kompetanse igjen for å sørge for at dataene gjøres FAIR og deles i tråd med regelverk.
Workplace - Where does this person work and who are they employed by?	En ansvarlig per institutt; in the research group, reaearch department university; University; Biblioteket, alternativt lokalt på forskningsgruppe; På biblioteket el. Fagenhet; Forskningsinstitusjonene, som en del av støttetjenestene der. Det er nok en fordel om datapersonen har fagspesifikk kompetanse, Vedkommende er ansatt på forskningsinstitusjonen og jobber på instituttet/avdelingen; At the University of Oslo, in the Research administration department. His new office is very close to the office of the recently employed data protection officer.; Universitet/Høyskole/forskningsinstitutt; Institutt på de store fakultetene, fakultetet ellers. Når IT er i faglig linje bør de være hos sentral IT med rolle ute hos forskerene; Institutt; Sentralt i forskningsavdelingen. Ansatt av Udir;Kan arbeide i forskningsadministrasjon sentralt eller ved fakultet/institutt, eller ved bibliotek. Ansettes av ledelse ved denne enhet; Jobber i de nasjonale datainfrastrukturene. Vertsinstitusjon ansetter; Helst institutt/avdeling, men ha god kontakt med fakultet; På universitetet, enten biblioteket eller på et fakultet. Det må være i ei stilling hvor kontakten med fagmiljøet kommer naturlig; Biblioteket, institusjonen; Ved et universitet/høgskole, fakultet eller institutt
Background – brief description of work experience and educational background	Bør ha egen forskningserfaring, IT og arkiv kunnskap. God til å lytte, kommunisere og finne praktiske løsninger på vanskelige dilemmaer/problemstillinger (dvs ikke en mekanisk regelfølger!); "IT bakgrunn; Forskningsassistent"; Må ha kompetanse på både (1) teknologiske sider ved datahåndtering, (2) relevant forskningsetikk og (3) grunnleggende jus på området; researcher linked to the field of the research group, plus background in data science; master degree; good knowledge of statistics and methodology; I fremtiden ser jeg for meg at man har egne datarøkter-utdanninger, på samme måte som man har bibliotekarer i dag; Utdannet med IT/datakompetanse, el. innen naturfag eller samfunnsfag, ev. humanist innen f. eks. lingvistikk, filosofi. Kan være spesialbibliotekar el. universitetsbibliotekar, IT-ingeniør el. faglig ansatt. Arbeidserfaring med - å håndtere data i ulike format - forholde seg til retningslinjer og dokumentasjon - domenekompetanse el. generalist avhengig av arbeidsplass; Gjerne innenfor det fag/område som personen skal serve. Personen bør også ha den nødvendige tekniske kompetansen knyttet til datahåndtering; Utdannet i IKT + disiplin. Bibliotekar med god kjennskap til IT (gjerne programeringskunnskaper), med utdanning/erfaring innen <relevant fagfelt> søkes; "Programmeringskunnskap / scripting; Kunnskap om LIMS og databaser; Kunnskap om metadata og metadatastandarder; Sans for orden; Utdanning innenfor Informasjonsvitenskap; Kompetanse innen dataanalyse og søk; Kan være stor bredde slik det er innenfor det forskningsadministrative feltet. Høyere utdanning, minimum mastergrad. God forståelse for forskning og spesialisert kompetanse på kuratering for arkivering/deling etter gjeldende standarder; Høyere utdanning; He has a PhD in computational linguistic and many years of data intensive research. Recently he has gotten an accredited course in data stewardship offered by GoFAIR office; Litt usikker på om det bør være en person, men et team pga. det er så mye som man må ha erfaring medregelverk, finansiørers og institusjonens krav, kjennskap til de data som brukes ved det instituttet eller fakultet, kjennskap til metadata og hvilket arkiv som er best å bruke for de dataene mm; Mye kommer gjennom erfaring, men man må skjønne forskning, og man må skjønne det store bildet ndg håndtering av forskningsdata anno 2018; Spesialbibliotekar med spesialisering innen data; Universitetsbibliotekar med fag der datahåndtering er del av faget med forståelse for forskningsprosessen; Høyere utdannelse, gjerne PhD innen fagfeltet ved fakultet/institutt hvor han/hun er ansatt

Bio - Please provide a short description of who this person is.	Har en bachelorgrad i et disiplinfag, med påfølgende mastergrad i datarøkting. Alternativt doktorgrad; Interessert av data, nøyaktig og fleksibel; Har antagelig en ph.d., men kan også være en med masterutdanning.; John has a scientific oriented, analytical mindset. He was engaged several years in data driven research but he got more interested in the challenges related to ontologies metadata definitions and less interested in the scientific topic and the final publication. Therefore he decided to go into an administrative carrier. He has also strengthen his knowledge in data management by attending several conference, forum and lately by taking a course in data curation. With the experience and the age he is getting more keen to do end-user support. Thanks to his research background he is able to read through the lines and understand the researchers, and he is also able to understand that the first priority for a researcher is to get a position/new publication/new grant and that is why they do not like to invest much time in doing data management best practices. But John is now also good at convincing them that a by-product of a proper data management planning will be more citing and therefore more accreditations! With his background in computational linguistic John has nice experience in coding/scripting and he is not scared in giving some technical advice in how to do data management in practice. On the contrary, he likes this!; Ein service instillt person som evner å arbeide tett på fleire ulike forskerteam - i ein støtterolle med fokus på å beskrive og persistere forskningsdata - i sanntid. Må gjerne bidra inn i analyse av forskningsdata saman med forskerteamet, mtp å trekke inn andre relevante data som er kuratert av andre forskningsteam. Bedre innsikt i bruken av data vil gi bedre grunnlag for gode metadata og dermed øke gjenbruk av data; Teknisk ryddoman; Uvesentlig utover at det er en person som er erfaren/moden nok til å forstå omfanget av feltet for datahåndtering og -kuratering og grensene for hva som kan deles. Forståelse for hele kretsløp for forskningsdata i prosjekter; En person med gode It-kunnskaper og interesse for feltet; Utdannet bibliotekar med erfaring fra lisensarbeid, opphavsrett og interesse for data.
Skills - please add minimum three words that describes what this person is particularly good at.	Koding, Kommunikasjon, Punktlighet; Må ha kompetanse på både (1) teknologiske sider ved datahåndtering, (2) relevant forskningsetikk og (3) grunnleggende jus på området; Able to working with large databases, creative, punctual; Data management and storage for further use???; Koding, systematisering, jus; "Utarbeide rutiner og fokus på gode brukergrensesnitt; God kjennskap til håndtering av ulike dataformater; Strukturert og nøyaktig; Kan sette seg inn i ulike fagmiljøers behov; Kjenner disiplinens dhp-prosedyrer, metadatakrav, dataarkiver osv.; Beskrivende metadata og publisering; Transformasjon av data (få det på rett format); Forskerområdets terminologi; Script, Management, Definere policies; Dialog med sluttbruker; Systematikk, forskningsforståelse, standarder for arkivering slik at kuratering sikrer langvarig tilgjengeliggjøring i arkiver; Digitaliseringsløsninger, personvern, rettighetsjus; data, dhp, metadata, regelverk, krav, FAIR, søk osv. jeg tenker alt som har med forskningsdata å gjøre; struktur, etikk, forskning; Tolke policy og retningslinjer,Valg av lisenser, Metadatastandarder, DOI, vokabular for fag, Strukturert og nøyaktig; Dokumentasjon, organisering, planlegging, kjenner til ulike typer forskningsdata
Motivations – please describe what makes this person enjoy their work	Jobbe tett på mange forskningsprosjekter; (for stor variasjon; men bør ha engasjement overfor både god forskning og etisk forsvarlig dataforvaltning); serving the research group; the salary?; Nærkontakt med forskning; "Trives i miljø med varierende behov; Servicerettet; Får bidra til å medvirke til åpen vitenskap; Vedkommende er høyt verdsatt av alle i forskergruppen og på instituttet; Bidrar til å gjerne forskning transparent og verifiserbar; Øker publisering på investerte resursers (til å samle inn data); Gleden av orden; Liker systematikk og orden, organisering av materiale (data). Trives med å hjelpe forskere til å ta vare på sine data - og dele data på en trygg måte; Bygger unik kunnskap i organisasjonen; jeg mener at dette ikke kan være en person, og motivasjonen må være at de har ett team som kan bygge opp dette felles; Liker å sette seg inn i ulike fagområders behov; Liker å gi servise til andre; Liker å samhandle og ""oversette"" mellom fag og IT"
Other things	Gode evner til samarbeid kombinert med personlig integritet og faglig relevant kompetanse; Må ta noen spesialkurs og være med i data management nettverk; Har respekt og forståelse for ulikhet i de ulike fagdisipliner og jobber for å finne løsninger som er gode for alle.

Appendix F: Questionnaire

Questionnaire - Study on knowledge infrastructure for research data in Norway

Thank you for participating in the study for my PhD project on knowledge infrastructure for research data. Based on the interviews preformed this winter I would like you to answer some questions. Answering the questionnaire takes between ??? and ??? minutes.

The knowledge infrastructure is fragmented and the respondents to this questionnaire have different roles, please answer based on what you are familiar with and express your personal opinions. Please use the open questions to elaborate if there is something you would like to add or comment.

Please look through the questions (attached) and get in touch if you would like explanations or clarifications on my questions.

Please answer in nettskjema (link)

Regards,
Live Håndlykken Kvale

Preferred language:
English
Norwegian

1. **To what extend do you agree with the following statements**
   1=Disagree, 3= Agree and 5= Strongly Agree 1-5 No opinion

   a. The point of data sharing is to make research verifiable and/or reproducible.

   b. Transparency and the possibility to trace results in the data material is one of the things that defines research.

   c. Sharing research data is necessary to enable critique and evaluation of research.

   d. Data is not analyzed enough and given the large investments in data collection, it is only reasonable to share the material.

   e. When research data is shared openly it is possible for other researchers to use the data claiming that they have collected the data themselves.

   f. Data sharing increases the value of the research data if other researchers makes new data connections.

g. Research data should be freely available (as in no costs). There are great positive externalities (additional effect) with sharing of research data and charging for access to data that is already collected is unfortunate.

h. Innovation: by increasing access to research data people can use it for something positive.

i. Science develops by building on previous research and creating something for others to build on, without sharing science does not evolve.

j. Letting other researchers analyse your data can provide a different perspective, this is innovative and can drive the innovation.

2. In the interviews, the greatest concerns regarding data sharing relate to privacy and other ethical challenges. I would like you to share your opinions on the issue by considering the following statements.

The term "privacy protection parties" refers to NSDpersonvern, REC and NESH, and Datatilsynet.

1 Strongly disagree Strongly agree 5

a) Privacy protection is important for public trust in research

b) There should be more room for putting privacy aside in order to conduct important research.

c) Research ethics is a difficult subject; I try to avoid it.

d) In the Google age there is no such thing as anonymity, that is why we need platforms with privacy protection embedded.

e) Privacy protection parties often lack an understanding of research.

f) It is confusing and conflicting for researchers that NSD is both collecting data and has a responsibility in regard to privacy protection.

g) Privacy protection parties and ethical committees contribute to improve the quality of research.

h) Many researchers will claim that their data cannot be shared because they are sensitive, but an elaborated version of the data could in many cases be shared without privacy being an issue.

i) When it comes to qualitative data you lose the details when you anonymize, but the value is often in the details.

j) Research ethics is not that relevant for my work.

2. Your ideal data person

In several interviews, the need for a data person of some kind (data steward, data curator, data scientist, data librarian, "datarøkter") was mentioned. In order to get a better understanding of who this is or could be, I would like you to spend some minutes creating an image of an ideal person.

I am asking you to create an imaginary character here, so please use your imagination.

a. Position/job title *
If you do not see the need for such a position please give a short explanation on why there is no need for this.

b. Name

c. Workplace - Where does this person work and who are they employed by?

d. Background – brief description of work experience and educational background.

e. Bio - Please provide a short description of who this person is.

f. Skills - please add a minimum of three words that describes what this person is particularly good at.

g. Motivations – please describe what makes this person enjoy their work.

h. Other things - Feel free to add additional information about this person.

3. The Data management plan (DMP) is mentioned by several of you as a tool in good data practice, at the same time, it is not clear whom the DMP is created for. Please point out who will be the three primary users of the DMP, and range them 1-3 by importance, please use the extra space to explain your view (if other please specify).

The researcher, other researchers, the data steward/data person, The institution, The funder, the archive, the journals, Other?

5. Please select what you find to be the five most important reasons to make a DMP (please select no more than 5 of the statements below and range them 1-5 by importance)

1 2 3 4 5

a) To create awareness in the research community for the need for data stewardship.

b) By being in control of their own data management, the research gets better and more efficient.

c) DMP gives the universities an overview of the ongoing research projects

d) When the researchers think about what they are doing at an early stage, they can make intelligent choices about their data.

e) The DMP provide the archives with information (metadata) they need when data is deposited.

f) DMP gives the archive the possibility to plan for data that are going to be deposited there.

g) It makes the researcher think about how to make their data and metadata FAIR (Findable, Accessible, Interoperable, and Reusable).

h) A DMP creates awareness and agreements on data procedures within a research group, particularly important when several partners with different data needs are involved.

i) A DMP shows what data is produced in a project when the project is finished.

j) A DMP is primarily a tool for those who take care of the data after it has left the project.

k) A DMP is a planning with a commitment to making the data as open as possible.

l) A DMP makes researchers aware of the value of what they are collecting.

6 a) What do you think would be a reasonable amount of a research budget to spend on data management? *
Please give an approximate in percentage and an explanation.

6 b) Who should cover the costs of data management? *
The university (at some level)
It should be budgeted in the project.

other

If other please specify

7. In order to describe the knowledge infrastructure for research data please use a minimum of three adjectives:

that describe what the knowledge infrastructure is.

that describe what the knowledge infrastructure should be.

that describe what the knowledge infrastructure is not and should not be.

(use cloud with adjectives as ill?)

Please comment on your answer.

8. Below some of the key-functions mentioned in the interviews are listed, please rate their importance as of level of requirement in an infrastructure for research data. (1=Not at all relevant and 5= Very relevant) 1-5 No opinion

a) Data publishing, also implying metadata.

b) Enough storage capacity

c) Complete workflow: recording, pre-processing, storage and one click to archiving with verification.

d) Possibility to use the data without downloading

e) Access from anywhere, without requiring for example an institutional affiliation.

f) Provenance: easy to see who created the data, and in what context.

g) Tailor-made option to the various user needs

h) Brings the computing and analysis tool to the data.

i) Interoperability between different solutions provided by different providers

j) Heritable metadata from project to files.

k) Shares metadata with discovery tools and search systems.

l) User friendliness.

m) Automated connection with related articles.

n) User (researcher) involvement in the infrastructure development.

o) High capacity for uploading, ability to steam quantity of data to remote storage.

p) Alignment with journal policies.

q) Possibility to share data with reviewers only during article review.

r) Quality assurance, not just a dump for everything.

s) "Reviewed by" and "approved by" functions to assure data quality.

t) Advanced consent and privacy options for research participants.

u) Possibility to share parts of the data openly and that an agreement on collaboration is signed before full access.

Please comment on your response

9. Do you have any thoughts on how the infrastructure can better facilitate transparency and replication in research?

10. Is there anything else you wish to add?

Do you consent for the information given in this form to be made openly available as part of my research data? *
The responses to the questionnaire will not be linked to the interviews, but grouped by primary role (policymaker, researcher, research support, infrastructure provider).

Yes
No

Thank you for answering my questions.
I will contact you again for a final short interview in the beginning of 2019.

Appendix G: Preliminary and final codebook

Theme	Code	Sub-code
Knowledge infrastructure	Todays infrastructure	
	The role of journals	
	DMP	*DMP experience*
		DMP goals
	Dream scenario	Priorities
		Technical solutions
		Support
	Ethics	*Privacy*
	How much and for how long	Costs and benefits
	Practical experience with data	
	Infrastructural needs	Gaps
	Infrastructural plans	
	International aspects	
	Demands	Need for guidelines
		FAIR
		National requirements
		Experience of demands related to own work
		As open as possible
	Accreditation	Rewards
		Citation
	Cultural change	
	Knowledge and training	Own knowledge
		Researchers knowledge of data
		Knowledge resources
	Metadata and data review	
	Open Access	

	Roles, responsibilities, and collaboration	Other stakeholders
		Distribution of responsibility
		Library
		Data steward
		Own role
		Research support
		Collaboration
		Unit
	Socio-tech aspects	
	View on the future	
	challenges	
Disadvantages	Unethical use of data	
	data fitness for use	
	Data integrity – actitation and rights issues	
	Drowning in data	
	Confidenciality and personal privacy	
	Competition and when to share	
	costs	
	Can demands of sharing result in lower data quality	
	Security in data storage	
	Time consuming and labour intensive for researchers	
	Unclear ownership	
	Losing control when sharing openly	
	Difficulties of commercialising what is openly available	
advantages	Uncover fraud	
	better research	other researchers can understand your research better

		higher quality data and data documentation
		reproducibility
		Standing on the shoulders
		transparency
	historical value	
	visibility	
	Added value in the society	
	more back on investments	Avoid duplication
	Further usage	More interdisciplinary research
		Combining existing data for new findings
		new and more collaborations
		innovation
	Increase trust in research in the general public	
Data Journals		
data quality		
Own attitudes		
The research process		
New ideas and concepts		

Table 8 Preliminary codebook used for exploration of topics in phase one

Theme	Code	Description
Persona/data steward	Career	Career tracks and career possibilities for data stewards
	Centre	Centralising of research data support services across research support departments
	Trust	Researchers trust in external data stewards to advice and do data management
Mata management plan	User	Addressing users and usage of data management plans
	Experience with tools	Discussing practical experience with using the different, existing data management plan tools
Ethics	Personal privacy	Privacy protection in research
	Consent	Usage of consent as legal basis when processing personal data
	Public trust	The general public's trust in research
	Cost profit	Cost and profit aspect of data archiving
	Integrity	Research integrity
	Research essere	ethos of research, what research are or strive to be
	Research ethics	Thought or experiences regarding research ethics
	Internationalisation	Internationalisation in research and data sharing
	Embedded privacy	The use of embedded privacy in privacy protection
	Privacy vs. research	Balancing of the respect for privacy with conducting high quality research
Cultural change	Plan S	Plan S and Open Access
	FAIR	The FAIR principles for data sharing
	Data sharing	Cultural change regarding attitudes or experiences with data sharing
	DM in education	Changes regarding the embedding of data management in education of researchers
	Incentives	The usage of incentives for a cultural change towards open science
Infrastructural development	New infrastructure	Development of new infrastructure or infrastructural changes
	Organisation	Organisational changes or organisations impact on the infrastructure development
Method	My data sharing	Opinions/reflections/experience with the data sharing in this project
	Consent	Thoughts on the use of consent in this project
	Experience as participant	Own experience as research participant in this project
Qualifiers	Reflections	The qualifiers were used on the coding to sort the different ways the subject (noted by adding theme and codes) was addressed.
	Practical experience	
	Suggestions	
	Concepts	
	Critique	
	Motivation	
	Expectations	

Table 9 Final codebook

Articles

Article 1

Kvale (2021). Using Personas to Visualize the Need for Data Stewardship. *College and research libraries* 82 (3).

DOI: **https://doi.org/10.5860/crl.82.3.332**

Using Personas to Visualize the Need for Data Stewardship

Live Håndlykken Kvale

There is a current discussion in universities regarding the need for dedicated research data stewards. This article presents a set of fictional personas for research data support based on experience and requests by experts in different areas of data management. Using a modified Delphi study, 24 participants from different stakeholder groups have contributed to the skills and backgrounds necessary to fulfill the needs for data stewardship. Inspired by user experience (UX) methodology, different data personas are developed to illustrate the range of skills required to support data management within universities. Further, as a competency hub for data stewards, the development of a research data support center is proposed.

Introduction

Data are the entities researchers draw conclusions from and are essential for fellow researchers to examine and criticize results. Transparency and access to data, the analysis applied, and the conclusions drawn are part of what defines research.[1] Data sharing and data archiving is expected to resolve the reproducibility crisis in research and provide new insight.[2] Consequently, academic journals and research funders are increasingly requiring research data to be made available.[3] Along with requirements for sharing data in academic research, there has been a growing need for new skills for data managers, data stewards, data librarians, and data scientists.[4] These new roles are professionals who assist researchers in managing research data, avoiding data loss during the research process, and preparing the data for archiving and public access. Digital research data are easily lost, and steps to preserve data must be taken in all stages of the research process.[5] Consequently, skills to maintain and curate data are required, but which skills are needed? And where in the universities should curation services be offered? These questions are currently being explored[6] and debated in libraries and among infrastructure providers.[7] This paper draws on a study of stakeholders involved in research data management in Norway involving policymakers (this group included representatives from the Norwegian Ministry of Education and Research, the research council of Norway and the rectorate of one of the included universities), national infrastructure providers,[8] and researchers and research support staff (in university IT, library, and research office) from the four oldest universities in Norway. By using persona templates adapted from user experience (UX) methodology,[9] this paper explores how the data stewards are described by different stakeholders. The aim with

---

**Live Håndlykken Kvale is a PhD candidate at Oslo Metropolitan University; email: live.kvale@oslomet.no.*

the making of the personas has been to visualize how a data steward team could respond to the various necessary competencies and skills needed for data management support.

Internationally, "data steward" is one of several terms used in the literature and among practitioners to describe a person working with research data management (RDM). "Data librarian," "data manager," and "data curator" are examples of other titles with somewhat overlapping responsibilities.[10] The term data steward is used in this article, as it is less domain-specific than "librarian," "curator," or "scientist." The usage of "data steward" is intended to include all the different requirements for data management.

The research question investigated is:

Who are the data stewards in the universities?

a. What roles should data stewards play?
b. What services should data stewards provide as part of these roles?
c. What skills do data stewards need to carry out these services?

By developing a set of data personas, it becomes possible to illustrate and exemplify one possible response to each research question; it is not to be interpreted as a universal solution, but rather as an example of how roles, skills, communication, and services for data management may be organized. The findings also focus on potential obstacles and what to be aware of when developing data steward services.

Literature Review

A broad range of literature on RDM skills were identified through searches for "data steward," "data librarian," "data manager," and "data curator" in Web of Science and Scopus. These articles were supplemented by searching relevant journals that are not indexed in these databases, such as *JesLib* and the *International Journal of Digital Curation* and adding other relevant documents. The different articles highlight the skills required in data management and the different roles of data professionals. The articles were grouped into three categories according to how the data steward was described: 1) new responsibilities of the librarian; 2) the embedded data steward in the research environment; and 3) other approaches to data management services. In addition, the literature review contains a section on the usage of personas related to data management services.

In the library and information science literature a majority of articles on data stewardship aim at clarifying which skills are needed for the data professional librarian offering data management support to researchers at the university.[11] Both Brown and Federer emphasize that "support for researchers' data needs is a moving target"[12] that needs to be supported by a skills development program in libraries.[13] The most important skills identified by Federer relate to communication, presentation, relationship with researchers, teamwork, and one-to-one training.[14] This argument is supported by Kennan, who finds that communication skills in many forms were the most in demand for RDM positions; she further emphasizes the need for "boundless curiosity," including both the willingness and ability to learn new things.[15] Kennan identifies four different roles that ensure data management in the different stages of the data life cycle: "the data librarian/data manager," "the data IT and systems experts," "the data scientist," and "the data creator."[16] Cox and Corrall illustrate the role of the "research data manager"[17] in the breach between the faculty and the academic library, connecting the institutional repository manager role in the library with the research produced by the faculty. The data librarians described can either be skilled generalists in data management or be specialized in a particular discipline. Disciplinary specialization can be achieved through engagement with subject specialists and researchers.[18]

A data steward working in a research group or similar research environment with data management is here referred to as embedded data steward. These domain-specific data stewards are primarily used in data-intensive research within health sciences[19] and natural sciences[20] and specialize in data management in a single discipline. An editorial from *Nature Genetics*[21] starts with a clear statement regarding data stewardship, asserting that, "professional data stewards be trained and employed in all data-rich research projects, [which] raises the exciting prospect they will conduct research on data-intensive research itself."[22] Some articles describe solutions for data management within national research institutes[23] or data centers.[24] The articles on embedded data stewards are discipline-specific and involve a high degree of specialization with a focus on the development of best practices and domain-specific standards.[25] This illustrates how an embedded data steward needs to understand the methods and data they are working with in addition to preservation and metadata. None of the articles describing data stewards in research environments are from the humanities or the social sciences. These disciplines have traditionally been less data intensive, and research is often conducted without data sharing among collaborating researchers. Also, the humanities and social sciences cater to needs differently, such as the trend of digital scholarship centers run by university libraries that explicitly serve the field of the digital humanities.[26] These are some possible explanations to why the experiences with embedded data stewards in the humanities and social sciences are fewer and newer, which again could explain why examples of embedded data scientists in humanities and social sciences have not yet reached the literature.

While embedded and library-centric were the two large categories to be found in the literature, there are other approaches to data management services. One example is the one-stop research support described by Clements[27] where someone can find answers to all questions regarding research data in one place, possibly a web portal. Another approach by Delft University in the Netherlands places domain-specialized data stewards within the faculty departments.[28] The service is coordinated by the library but aims to integrate the services of the data steward in each faculty. Still, their goal is to provide "more granular disciplinary experts."[29] The report from Research Libraries UK and Matt Greenhall exploring digital scholarship in UK libraries argues for a "mixed economy of digital scholarship support"[30] whereby the library partner supplies other research support facilities at the universities with complementary expertise in data management. In literature on data scientists the term "Data unicorns"[31] is used to mean an unrealistic skillset for one person. Kennan transfers this to the idea of the data steward.[32]

What the literature on data stewards has in common is the exploration of professional domains and services new to librarianship. The primary challenges described include the targeting of the right level of specialization versus the general knowledge of data management and communication and collaboration between the different levels within the organization.[33]

In the context of RDM, there are three examples of the usage of personas[34] within the literature. Lage builds on the usage of personas to improve institutional repositories for publication.[35] Crowston includes "Abby the Science data librarian"[36] in the group of users of research data repository. Both focus on the researcher, presenting five[37] and eight[38] researcher personas with different needs in regard to data management and different interests in using institutional archives for research data. A recent report on education for data stewards from

Denmark[39] presents the use of personas to illustrate the different needs and skill sets requested for data stewards in both the corporate and research sectors.

Methodology

RDM is a rapidly developing domain. To grasp some of the changes and developments, a Delphi study with an expert group and multiple rounds of data collection was found to be suitable.[40]

The expert group of participants in a Delphi study provided the possibility of bringing preliminary findings back for discussion, contributing to the understanding of the perception of roles through negotiation, testing, and learning. A group of 24 stakeholders participated in the study (see table 1). The group contained representatives from policymakers and national infrastructure providers in addition to researchers and research support staff from four universities in Norway. Recruitment of participants from different stakeholder groups was to include different aspects of the development of the sociotechnical infrastructure for research data and potentially uncover gaps or disagreements. The data steward was one element highlighted from multiple stakeholders as a gap or missing link. The four universities are the oldest in Norway, are all multidisciplinary, and have well-established collaborations on administrative and technical infrastructure. From the policymakers, rectors of research at the four universities were invited to participate (unfortunately, only one of the four invited rectors agreed to participate) in addition to representatives from the Norwegian ministry of knowledge and research and the research council of Norway. The infrastructure providers represent different organizations that offer data archiving services to universities in Norway (these three are all publicly funded and offer different archive services). The researchers were invited based on their receipt of European Union (EU) funding. The EU requires data management plans from the projects they fund. Researchers were identified through the cordis webpage (https://cordis.europa.eu/en); of 25 invited researchers, 8 participated in the study. This way of identifying and recruiting researchers was done to avoid potential biases related to engagement with data management as a topic. It also gave a pool of researchers with different disciplinary backgrounds (biology, musicology, science studies, economics, neuroscience, psychology, philosophy, gender studies). The grouping of researchers as working either individually (RI) or collaboratively (RG) was done during the analysis of data from the first round of collection, as the needs described corresponded with how the researchers collaborated with other researchers on data, rather than with disciplinary backgrounds. The research support staff were recruited with the focus of including representatives from three types of

TABLE 1
The Participants Organized According to Role

Role/Stakeholder Category		**Individual Participant Codes**			
Researchers working individually	RI	RIZ	RIJ	RIL	RIB
Researchers working in groups	RG	RGV	RGD	RGA	RGW
Policymakers	PO	POU	POS	POK	
Infrastructure service providers	IN	INH	INO	INR	
Research support IT	IT	ITE	ITY	ITI	
Research support, research office	RO	ROC	ROX	ROT	
Research support, library	L	LM	LP	LG	LN

research support services (library, IT, and the research office). Five of the research support participants also had previous experience as researchers, and two provided IT services that were offered both locally and nationally. In some cases, the participants did not want their statements to be identified with them; these have been marked with stakeholder category.

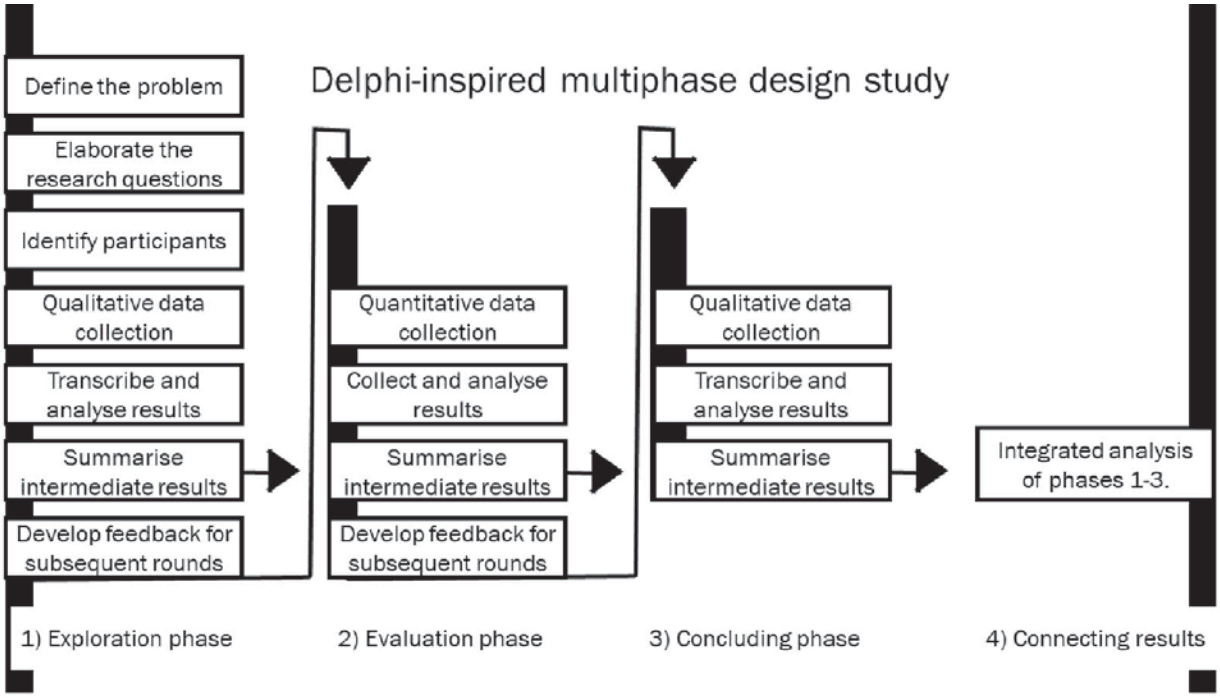
Within UX methodology, personas are commonly used to describe users of computer systems.[41] The development of personas builds on data collected through interviews or surveys, with the aim of creating fictional characters, either based on the participants or to fill the roles they describe. By creating personas, system developers flip the focus from the system to the user, aiming to create a product that fits well for some users, rather than merely adequately for everyone.[42] With the ongoing changes in the data management landscape and current infrastructure development, the data stewards will be central users. Still, who will fill these roles is not clear.

By using the expert group of the Delhi study to develop data stewardship personas, this study is not claiming to offer a universal solution but provides an illustration of how the different roles could be distributed. Data steward personas can be useful both to system developers and to the universities that employ data stewards and develop data management services.

Figure 1 illustrates the different phases of the study. In the exploration phase (January 2018), open interviews approximately one hour long were conducted with the participants. Data stewardship and skills for data management were but two of the several themes brought up in the interviews.

To further explore the expectations of the different stakeholders interviewed regarding data stewards, the second round of data collection (September 2018) had a section dedicated to the data steward (see appendix) inspired by UX-persona design. All answers were given in free text and were optional.

FIGURE 1
Delphi-inspired Multiphase Method[43]



In the concluding phase (March 2019), a first draft for three personas was developed based on the preliminary findings. This draft was presented and discussed with the participants in open interviews lasting about 30 minutes. The persona drafts were shared with participants prior to interviews as part of the interview guide. The findings presented in this article are from all three rounds of data collection and include an integrated analysis[44] of the results. Quotes presented in this paper are marked with a participant code and 1 or 2, referring to first or second interview (for instance, "RGA2"). Data from the survey are not linked to participant.

The collected data were, for the most part, qualitatively coded and analyzed thematically,[45] initially using the software NVivo and later using XML for thematic coding and Python script for extraction. Some of the results from the surveys, such as background, education, and skills, were counted and treated quantitatively.

Most participants granted permission to share the whole or parts of the data with directly identifiable information such as names removed. They all had the opportunity to review data they contributed ahead of publication and to indicate if there were parts they did not want published. The data, including the XML codebook, Python script, interview guides, transcripts, survey, and consent forms can be accessed through Zenodo.[46]

Findings

The findings first present the need for data stewardship before exploring in greater detail the skills and background requested for data stewards, which are used in the development of the personas.

The Need for Data Stewards

Several participants pointed towards a need for data stewardship. The vocabulary used to describe this need varied along with expectations as to how this role should be filled. The practical challenges of data planning, data management, and data curation were explored, along with collaborative skills between existing research support services, data stewards, and researchers.

Data management does require knowledge of research. Respondent ITE (research support, university IT cf Table 1) emphasized that, "the researchers know their data so only they will be able to describe their data, but they need help from the data curators" (ITE1). The data curator role described by ITE is defined as working with the departments to create continuity and preserve valuable digital data. ITE also believed in employing data curators to avoid data loss when temporary staff leave and further describes the need for data stewardship and data management as a consequence of data-intensive research. Among the researchers, RIZ feared that the general data steward would not be able to understand the context: "To do this type of job you must know the context, and to do this on an industrial scale might work in some cases but probably not in all" (RIZ1). As RIZ pointed out, some data types might be easy to structure and organize with a lower degree of specialized knowledge, whereas other types require a higher degree of specialization and domain-specific knowledge.

Understanding when different types of knowledge are required is also an issue, as well as understanding what one can expect researchers to do themselves versus what they need additional expertise to do, such as making data interoperable and creating a data management plan (DMP). Participant ROC addressed the long-term perspective: "I don't believe any researcher can have the responsibility to follow the data from collection […] until they are ready to be stored for maybe 1000 years." Making decisions on what should be selected for long-term storage itself requires expertise in addition to performing the actual data preservation.

Two of the researchers working in larger collaborations have hired, or are in the process of hiring, data stewards. RGA works on a multidisciplinary project that generates a large volumes of data from a variety of sources, while RGD collects social science data from previous projects for reuse in a new context. Both agreed on the need for data management: "The largest need is for human resources to manage data" (RGA1) and, "For us, research data means how to integrate data from all these sites, how to harmonize, standardize, and integrate them, and then how to analyze them in a way that something new comes out of that" (RGD1). RGD also described how several people are working with different aspects of data management, from data cleaning and access control to the re-collection of consent from participants.

Collaboration is also suggested as a challenge: "I believe it is important with such a holy trinity that IT, library, and administration could become if they would work together" (RGA1). She pointed to the need to combine different people with different skills and different backgrounds to solve complex issues and to create robust data management services. Also, one of the research office staff noted that collaboration is key: "There is not one such person, one that knows everything, it is a Kinder egg, three, four or five things you need to have thorough knowledge of" (ROT1). Metaphors such as "Kinder egg" or "Holy Trinity" have similarities with the "data unicorn,"[47] indicating that expectations for research support services for research data ought to be collaborative to deliver the complexity of skills required. In the survey, one participant from the library explained that, rather than a person, she saw as the best solution a team of people with different competencies complementing each other.

Researcher RGW explained that data management was the responsibility of the professor in charge of the lab: "In our group we don't actually have a data manager, but it is mostly the job of the professor. The data type has been fixed a couple of years ago that the data should be analyzed in such and such a way, so it has the same data structure. The professor acts like a data manager also. But because we are temporary researchers, and we have our own style, [the] professor should decide the data structure" (RGW2). As the majority of the researchers in the lab are there temporarily, it is the responsibility of the lab, and the professor to decide the structure and formats of the data, to "act like a data manager." Still, since RGW's description pointed to a high level of awareness, there is likely a formal or informal protocol for data management in the lab, and the responsibility belongs to the principal investigator. Also, among research support staff, it is agreed that the researchers themselves should be responsible for knowing basic data management: "I think that as a researcher, I would not say you are obliged to, but you should know basic data management" (LM2). This does not, however, exclude the need for dedicated data managers and further highlights the need for available training.

The participants described the following needs for data stewardship:

- As research is becoming increasingly data intensive, larger research groups may need to hire data managers.
- Data loss from PhDs, postdocs, and other temporary staff when leaving the university is a challenge.
- To find the right balance between the generalist and the specialist is important in terms of playing the right data management support role.
- A closer collaboration between IT, library, and research offices is needed.

- All researchers cannot be expected to do data management on their own, yet it is the responsibility of the researcher to ensure good data management in his or her research.

Collaboration and Communication between the Different Support Levels

In the survey, there were 20 responses to the question regarding the workplace of the data steward, which showed a general agreement that closeness to the research environment is essential. In particular, the researchers emphasized that the employee should work in the research groups. The research group or/and departments (10) were mentioned most frequently, but the research administration (4) and the library (5) were also suggested as appropriate work environments for the data steward. One suggested the national research data infrastructures as the appropriate place to employ the data steward. In the interviews, several participants elaborated on this by emphasizing how collaboration and communication between different levels of support within the universities are crucial:

> *You need to create a system where these people actually work together and are able to interact in a good way. [……] There is a pulverization of responsibilities absolutely everywhere, and with such a research data center it might be possible to avoid this, given the entry points and the information flow and such. (RGA2)*

Both responses show how the workplace of the data steward is one issue to consider, but many challenges are related to organizational culture, organization, and information flow. One of the researchers suggested coordination between the universities to ensure standardized and high-quality services: "You need a way to assure that you even out the pressure from place to place, so you don't end up in one bubble, each with the development of strange subcultures; this is important to avoid, difficult to avoid but very important" (RIJ2). The workplaces of the data stewards need to be interconnected in networks of information and skills exchange locally and, perhaps, nationally and internationally. As RIJ notes, hiring data stewards without facilitating knowledge exchange can easily create dysfunctional subcultures rather than interoperable data.

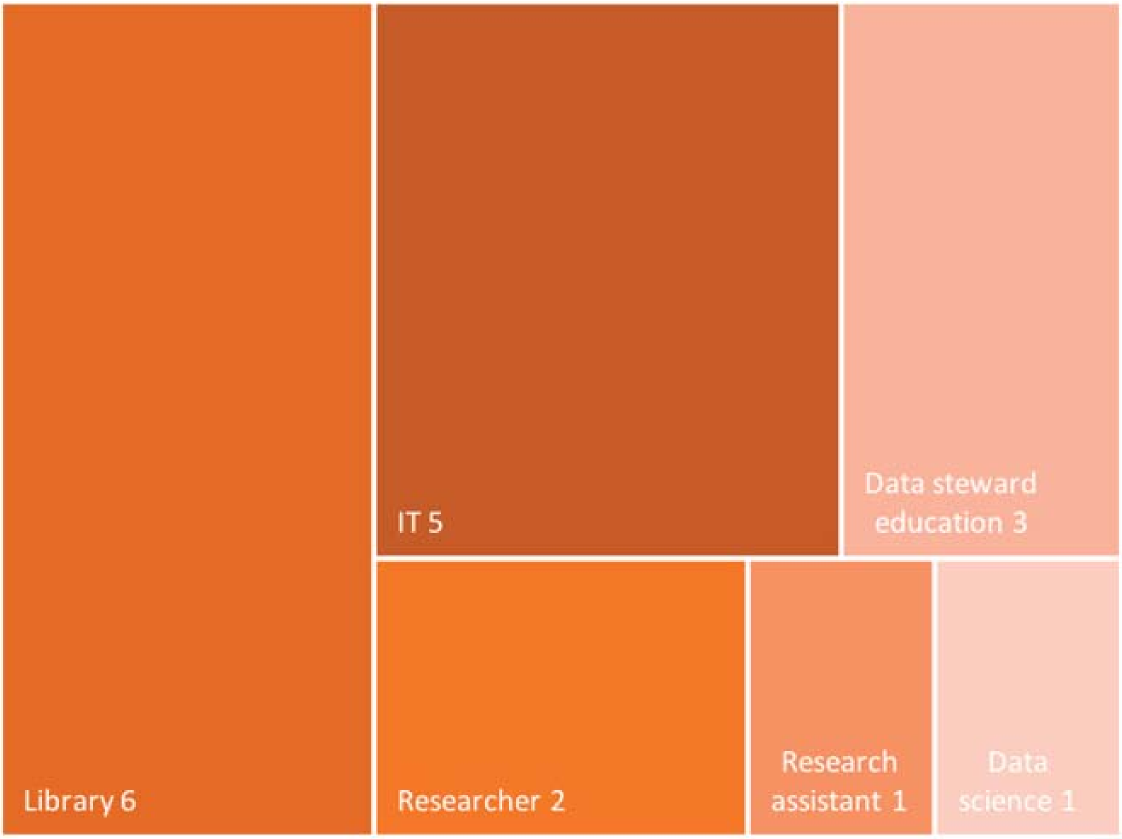
Speaking the Same Language

The respondents (21) mentioned several educational backgrounds in different combinations; the results have been split and grouped in figure 2. Further, four mentioned the master's level, and five mentioned the PhD level as the appropriate educational level. Others suggested higher education without specifying the degree level. The respondents all suggested that the ideal candidate would be a highly educated person preferably with research experience, often in combination with a background in data stewardship, IT, or library and information science (LIS).

As one of the staff members at the university explained, a PhD degree can be a gateway to communication with the researchers to "speak their language" and create trust:

> *It helps if they all speak the same language. That is part of the success in my department, where half of the staff have a PhD, so we can communicate with the researchers.… You first need some positive and some negative experiences in order to make the transition. Someone has done this internally … others who have not committed huge mistakes yet, they just continue to build, data on data on data and more data, without any control.* (ITI2)

FIGURE 2
Preferred Background for Data Stewards



The notion that experience of research can be one way of creating trust and knowledge of the data types and methods used in the field is another point of entry. However, ITI believed that the need for data management must often be experienced by the researchers before it can be taken seriously. Similar backgrounds help in creating relational bonds and trust between researchers and data stewards:

> *I think this is also a kind of confidentiality. There is something like a role you trust, like that person is to be really trusted, so I think it would be, I don't know, but if I was a researcher I would be a bit, I don't know maybe awkward to contact somebody who is just a data manager and is not related closely to my field.* (RGD2)

One of the researchers described a fear that the data stewards operate using their own agendas:

> *If you enter on the side in that mean of adding an additional agenda beyond solidity and such, sometimes that might be an advantage for some types of projects, for others it might be alarming, both economic and in terms of work environment.* (RIJ2)

Research experience among data stewards or similar disciplinary backgrounds are possible strategies to create common ground between data stewards and researcher. These strategies might also help to avoid additional agendas on the part of data stewards. Interest in research or the research topic might help to assure the researchers that solidity and reproducibility are the stewards' primary motivations for data management. There are, however, already several agendas present in the field of data management, such as economic interests,[48] and an interest to explore existing data in new ways through data science.[49] For the researchers, on the other hand, the purpose of data management is primarily to document and archive research data for their own reuse and for reproducibility. One of the policymakers pointed to this conflict of interests and motivations: "A risk in this area, and what we have seen until now that the area suffers from, is that library and archive people, bureaucrats, and non-researchers have taken a strong role of leadership" (PO). When policymakers, archivists, IT developers, data scientists, and librarians all see different potential in research data, these interests might come to overshadow the core: the quality of research and the challenge in overcoming the reproducibility crisis.

The question of what motivates the data steward in doing their job becomes important for building relations between the researchers. Twelve participants answered this question. Ethical motivations and genuine engagement in research were seen as the most important motivations: "Engagement both with good research and ethical data management," "the enjoyment of assisting researchers in taking care for their data and sharing data in a safe way," and "[Contributing] to making research transparent and verifiable." Other responses described a methodical person with a genuine interest in research who can provide a valuable contribution by organizing, providing services, building something together as a team, and contributing to science.

Dividing Tasks but Maintaining Responsibility

When asked to write a short biography, nine participants responded. One of the descriptions given was that of a "technical and tidy person," and other characteristics included a good overall understanding of research and of the research data life cycle: "The person must be mature or experienced enough to understand the range of the field of data management and curation, and the limitations for what should be shared and [to] understand the whole lifecycle of research data in projects." Another participant described "a service-minded person able to work closely with several research teams." Thus, both emphasize that technical and social skills are necessary, along with experience, knowledge, and the ability to provide professional guidance. One participant gave a longer description of a researcher who wanted to work in-depth with data and who enjoys both the service and the problem-solving aspects of data stewardship.

Balancing the interest in the research with motivations to keep the data structured and documented to enhance the quality of the research results without adding additional agendas is important. Still, the involvement of the data stewards must be balanced in such a way that the responsibility of the research data is not completely transferred away from the researchers:

> *You hope that data management should become embedded in normal research practice, for much of this can, with fairly simple means, become part of existing routines.… Because the problem, if you get a data manager in the group, is that the others might not take as much responsibility for the data management.* (LM2)

The interviewed researchers shared the concern of LM. A data steward must provide support without creating an excuse to transfer the responsibility from the researchers; when data are deposited in an archive, a transfer of responsibility can take place:

> *The researchers themselves must be responsible … I realize this myself in part of these discussions, that one thinks ´yes, we create this role, and then everything is solved,´ but it is not at all in that way. Because the researcher sits with the data set and needs to make sure this is in order, and then you need a curation function, but that again depends on the data set and where you are in your research process.… But from the moment we have a publication, with a corresponding data set, made available, then the data set will still need curation, but then you are more on the library side. First the researcher needs to sign off the responsibility, and then others take it on.* (RGA2)

Another option proposed by RIZ is not to create data stewardship positions, but to distribute responsibility among existing researchers in a group:

> *I would say that the competency should be in the group, and not in an extra position; I believe there are other positions more important to prioritize, so I guess I am against all these, but the nearer the better.* (RIZ2)

RGA and RIZ work in extremely different research environments: while RGA works in a collaborative and data-intensive environment, which employs its own data managers, RIZ is a theorist and collaborates with other researchers on publications. RIZ's point of assigning responsibility for ensuring data quality to the researcher is representative of the view of many researchers, in particular those working independently. She argued that the quality of your data is the quality of your research, and your responsibility as a researcher.

Fifteen participants listed different skills as being necessary for data stewardship; some skills were mentioned several times. The skills mentioned are analyzed and grouped in table 2. Different labels, such as personal skills, general skills, research skills, knowledge of law and policy, technical skills, and archiving skills, differentiate the variety of skills listed. The label "general skills" is used for skills that are found to apply to more than one of the other categories. Knowledge of metadata are most commonly mentioned. However, none of the researchers mentioned metadata explicitly. There is one mention of "data management and storage for further use," while another writes about "coding, systematization and law"; this is the response that best reflects the feedback from the researchers, along with responses that emphasize personal skills, such as creativity, punctuality, and good communication skills.

The Personas

Based on the analyses, the placement of a support service in the right context, with appropriate channels of communication and collaboration, appears to be one of the major challenges of delivering appropriate services. As a workplace for two of the data steward personas, the Research Data Service Center (RDSC) has been developed. The RDSC draws on inspiration from the development of digital scholarship centers, but with a multidisciplinary approach and with the emphasis on strengthening collaboration between the different research support services within a university. Several participants requested better collaboration to provide better data management support; the suggested RDSC is one response to this. In the RDSC, the library, IT, and the research administration are aligned in a partnership for coordinated research data support. Further, three different data steward personas filling different roles and levels of support are presented: the RDM service coordinator, the data curator, and the

TABLE 2
Data Stewardship Skills (times mentioned in parentheses)

Personal Skills:	General Skills:	Research Skills:	Law and Policy:	Technical Skills:	Archiving Skills:
Structured and organized (4)	Knowledge of research (4)	Knowledge of discipline-specific terminology (2)	Understanding and interpretation of policies (3)	Programming, coding, scripting (4)	Metadata related (6) (hereunder: metadata demands, standards, documentation, descriptive metadata)
Accurate (5)	Research ethics (3)	Ability to understand discipline-specific needs (1)	Knowledge of law and juridical aspects (2)	Technical aspect of data management (1)	Familiarity with organizing and planning for different types of research data (2)
Dialog with end user/ communication (2)	Knowledge of the FAIR* principles (1)	Statistics and methodology (1)	Define policies (2)	Ability to work with large databases and LIMS (2)	Systematization (2)
Creative (1)	Data management and storage for further use (1)		Personal privacy (1)	Digitization (1)	
Flexible (1)	Ability to work with guidelines and documentation (1)		IP-law (1)	User interface (1)	Search (1)
A problem solver able to think outside of the box (1)			Familiar with DMP procedures (2)	Data transformation (1)	Data archives (1)
Good listener (1)					Archival standard for curation and secure long-term archival storage (1)

*Findable, Accessible, Interoperable, and Reusable: FAIR Guiding Principles for scientific data management and stewardship.

data manager. Again, it is necessary to emphasize that personas are fictive entities, and real people could be filling these roles. The number of data stewards will vary depending on institution size. The survey responses gave a mix of male, female, and gender-neutral names, and the personas have been carefully constructed to reflect this. The author selected illustration photos to give the personas more of an identity by providing them with a face; care has been taken to avoid stereotyping. The names and photos were presented to the participants in the final interview; none of the participants presented any opinions on either, but focused on the roles and skills embedded in each persona while referring to each with the names.

The Research Data Service Center

The RDSC is run collaboratively by IT, the library, and the research office at the university. The RDSC has been established to solve issues of RDM support and training but also espouses other related research skills, such as data visualization, data analysis software, and support on statistics. The services they offer are divided into core services provided by RDSC staff and coordinated services where the RDSC is the host for related networks and courses. RDSC is designed to be user-centered and responsive to current needs among researchers who are testing and offering the latest in technologies for research data. By having an approval function for data management plans and, by coordination, network meetings of data managers, they map and respond to the knowledge level and needs of their local environment. The RDSC are up to date on challenges and needs in their community. Further, they collaborate closely with different departments at the university to ensure that data management training is offered to researchers and graduate students.

Core Services

- DMP review and consultancy
- One-to-one data management support for PhDs and researchers
- Courses in data management
- Coordination of the "peer-support network" of data managers

Coordinated Services

- Hosting courses focusing on skills for research (Python, poster design, R and other courses provided by the Carpentry community)
- Hosting other peer support networks (such as Carpentry study group, R-ladies, and the like)
- FAIR training courses

There are three groups of staff at the center: permanent staff, student staff, and associated staff. In addition, they collaborate closely with the data protection officer and with a network of data managers hired by a research group.

The permanent staff includes one RDM service coordinator (Kim) and data curators of whom David is one.

Based on requests, student staff are hired from a pool of data science students and PhD candidates. This offers students interested in data management an opportunity to practice and brings new expertise into the center. Some of these students end up being hired as data managers in data-intensive research groups upon graduation.

Associated staff work at the research office, library, and in IT but have some tasks at the RDSC. Typically, expertise on data analysis software and statistics are offered by IT staff along with support on writing. DMPs and grant fulfillment are offered by the research office, and metadata and data archiving are offered by library staff. In addition, each individual brings their own skills—some with graphic design, others with ontology building, artificial intelligence, interaction design, or semantic web technologies. This renders the center an interdisciplinary environment that focuses on collaboration and RDM, as well as the proliferation of skills for data-centered research.

The RDM Service Coordinator – Kim Smith

Photo 1: Kim Smith, Ill. from Colourbox

Kim Smith is the coordinator and communicator with the RDM service. She has a master's degree in LIS and several years of experience at the university library. Kim works as RDM Service Coordinator at the RDSC and is responsible for the data management services at the university. She oversees and coordinates everyone involved at the RDSC. Kim enjoys teaching and presides over several of the RDM training courses offered at the university. Through a series of workshops held at the center, she has given several researchers and master's students their first RDM course. She also advises on privacy and copyright issues, and while she does not have a background in law, experience has made her able to advise on many of the issues that occur. When in doubt, she consults the data protection officer. Kim is also responsible for the review and approval of DMPs. The workload is, however, shared, and the plans are reviewed collaboratively at the center. Through DMP reviews, Kim, David, and other staff at the RDSC are able to identify potential challenges at an early stage and offer support. In addition, Kim is active in the international coordination work done with the Research Data Alliance. Core skills:

- Communication and interpretation
- Policy expertise
- Research ethics and personal privacy
- Intellectual property law
- Data management plans
- Metadata

Motivation: Kim believes in contributing to making research transparent and verifiable and building new knowledge in the organization.

Kim believes that proper data management can solve the reproduction crisis and help rebuild trust in research in society in general. With a background as a librarian, she is focused on data quality and longtime curation. Kim is also concerned about maintaining the legacy of prominent researchers at her university. Her colleagues describe her as structured and strategic.

The Data Curator – David Carpenter

Photo 2: David Carpenter, Ill. from Colourbox

David holds a PhD in computational linguistics and many years of experience with data-intensive research. Recently, he has taken a course in data stewardship. David has a scientifically oriented, analytical mindset. He had been engaged for several years in data-driven research, but he became more interested in the challenges related to ontologies and metadata definitions and less interested in scientific topics and final publications over time. David is good at convincing researchers that a by-product of proper data management is an increased number of collaborations, citations, and accreditations. Core skills:

- Systematization
- Making data FAIR
- Metadata, documentation, and provenance
- Data archives and archiving
- Coding
- Data mining
- Formatting and data transformation

Motivation: David enjoys translating among disciplines, understanding researchers' needs, and solving problems.

David loves research and the university as a work environment, but he prefers working with the data rather than publishing. He is described as accurate and systematic.

The Data Manager – Kari Anderson

Photo 3: Kari Anderson, Ill. from Colourbox

Data manager Kari Anderson is the disciplinary specialist, while the staff at the research support center are the generalists. She is one of the data managers working in the data-intensive research groups at the university. The data managers meet monthly at the peer support network at the RDSC to exchange experiences and solve concrete problems. Kari makes sure there is an agreement on standards and protocol for data management within the research group. When new staff is hired or if students are participating, she makes sure they are briefed in data management before touching anything. Kari identifies with the other researchers in the group. She is good at picking up on potential issues at an early stage, and if someone has problems with conversions, transfer, or the merging of data, she loves the challenge. She is also focusing on deleting what is obsolete, rather than keeping every version of everything.

Kari has a PhD in neuroscience and is fascinated by classification. Through statistical classification, she has developed an interest in AI. She was working closely with a research group during her master's and was later hired as a PhD. During her PhD period, her role gradually became more of a data manager, and when a new center for brain research was established, she was hired as a data steward. She is also taking some extra courses within data science to work with still more methods and disciplines as a data manager/data scientist.

Through the RDM network at the university, she learned of the Research Data Alliance and is now engaged in the health data interest group, where she keeps up to date. Still, her heart is most at home in the R-ladies network. Core skills:

- Documentation
- Working with large databases
- Coding
- Systematization
- Data transformation
- Metadata standards
- Interoperability

Motivation: Kari loves working in the creative environment of research while still clocking office hours.

At the lab, she is described as the right hand of the professor, the go-to person for the people working there, and a creative and hard-working part of the team.

Persona Summary

By creating the personas Kim Smith, David Carpenter, and Kari Anderson, the aim has been to visualize and concretize one example of how both a team providing general support and a data steward working within a research group can function. What is crucial is that the data stewards have a genuine interest in contributing to research and a combination of the right soft skills and knowledge of research along with technical, law and policy, or archival skills. The personas can be applied both in the development of software solutions and as inspiration when creating better research data support at the institutions.

Conclusion

The findings from this study show that outreach, education, and problem solving are only some of the keys to the creation of a functional service for data management. There are several concerns that must be taken into account as a service is developed.

Four primary challenges for providing data stewardship at universities are identified:

- Placement of responsibility: Researchers must retain their responsibility for data throughout the research cycle. When depositing to a data archive, responsibility can be transferred if the selected archive offers curation services.
- Communication: Lines of communication between support levels must be established to avoid closed subcultures and to exchange best practices between domains.
- Knowledge of data and methods: There is a need for local and specialized expertise within an increasing number of domains. It is necessary to find the appropriate degree of disciplinary knowledge to provide support. Knowledge of research is essential; however, the researchers are responsible for data management in their projects.
- Joint research support effort: Research data management requires several different types of expertise that traditionally are spread among different research support departments at universities. The creation of a general research data support team or center with connection to the research office, IT, and the library is crucial to cover all aspects of data management.

One solution can never fit all, and, while a general team will be able to solve and support a wide range of issues, many larger research communities need dedicated staff with specific knowledge of the issues and concerns that are relevant for their research data. While data management is gradually becoming current practice within several data-intensive communities, it is also needed among researchers producing and collecting small heterogeneous datasets, referred to as the long tail of research data;[50] a research data support center is an attempt to resolve this. A general team will function as a professional network for discipline-specific research data staff and could potentially assist research groups in recruitment and transfer of skills and knowledge across disciplinary boundaries. Motivated by contributing to research, data stewards can be recruited among both graduate students and researchers; however, understanding of research and research methods is important.

APPENDIX

Questions describing the data steward in the survey.

Your ideal data person	
In several interviews the need for a data person of some kind (Data Steward, Data Curator, Data Scientist, Data Librarian, "Datarøkter") was mentioned. To get a better understanding of who this is or could be, I would like you to spend some minutes creating an image of an ideal person. I am here asking you to create an imaginary character so please use your imagination.	
a.	Position/job title If you do not see the need for such a position, please give a short explanation on why there is no need for this.
b.	Name
c.	Workplace: Where does this person work and who are they employed by?
d.	Background: Please give a brief description of work experience and educational background.
e.	Bio: Please provide a short description of who this person is.
f.	Skills: Please add a minimum of three words describing what this person is particularly good at.
g.	Motivations: Please describe what makes this person enjoy their work.
h.	Other things: Feel free to add additional information about this person.

Notes

1. Robert King Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago, IL: University of Chicago Press, 1973).

2. Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, MA: MIT Press, 2015); Peter T. Darch, "Limits to the Pursuit of Reproducibility: Emergent Data-Scarce Domains of Science," in *Transforming Digital Worlds*, eds. Gobinda Chowdhury et al., vol. 10766 (Cham, Switzerland: Springer International Publishing, 2018), 164–74, https://doi.org/10.1007/978-3-319-78105-1_21.

3. Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (Los Angeles, CA: SAGE Publishing, 2014).

4. Alma Swan and Sheridan Brown, "The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs," *Report to the JISC* (Truro, UK: Key Perspectives Ltd, 2008); Robin Rice, *The Data Librarian's Handbook* (London, UK: Facet, 2016).

5. Kitchin, *The Data Revolution*.

6. Michael J. Scroggins et al., "Thorny Problems in Data (-Intensive) Science," in *Publications* (Los Angeles, CA: UCLA Center for Knowledge Infrastructures, 2019), https://escholarship.org/uc/item/31b1z69c; Marta Teperek et al., "Data Stewardship Addressing Disciplinary Data Management Needs," *International Journal of Digital Curation* 13, no. 1 (December 27, 2018): 141–49, https://doi.org/10.2218/ijdc.v13i1.604; German Council for Scientific Information Infrastructures (RfII), "Digital Competencies: Urgently Needed! Recommendations on Career and Training Prospects for the Scientific Labour Market" (Göttingen, Germany: German Council for Scientific Information Infrastructures [RfII], 2019), http://www.rfii.de/?p=4015.

7. German Council for Scientific Information Infrastructures (RfII), "Digital Competencies: Urgently Needed!"; Philipp Conzett and Lene Østvand, "Støttetenester for forskingsdatahandtering på UiT Noregs arktiske universitet – erfaringar og forslag til beste praksis," *Nordic Journal of Information Literacy in Higher Education* 10, no. 1 (May 31, 2018): 65–80, https://doi.org/10.15845/noril.v10i1.283.

8. In Europe and Norway there is a strong tradition for national custodians of research data; Kristin R. Eschenfelder and Kalpana Shankar, "Of Seamlessness and Frictions: Transborder Data Flows of European and US Social Science Data," in *Sustainable Digital Communities: 15th International Conference, IConference 2020, Boras, Sweden, March 23–26, 2020, Proceedings*, eds. Anneli Sundqvist et al., vol. 12051, Lecture Notes in Computer Science (Cham, Switzerland: Springer International Publishing, 2020), 695–702, https://doi.org/10.1007/978-3-030-43687-2.

9. C. Lewis and J. Contrino, “Making the Invisible Visible: Personas and Mental Models of Distance Education Library Users,” *Journal of Library and Information Services in Distance Learning* 10, no. 1/2 (2016): 15–29, https://doi.org/10.1080/1533290X.2016.1218813.

10. Mark D. Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3 (March 15, 2016): 160018, https://doi.org/10.1038/sdata.2016.18; Sara Rosenbaum, “Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access: Data Governance and Stewardship,” *Health Services Research* 45, no. 5p2 (2010): 1442–55, https://doi.org/10.1111/j.1475-6773.2010.01140.x; Swan and Brown, “The Skills, Role and Career Structure of Data Scientists and Curators”; Jingfeng Xia and Minglu Wang, “Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions,” *College & Research Libraries* 75, no. 3 (May 1, 2014): 362–88, https://doi.org/10.5860/crl13-435; Anna Clements, “Research Information Meets Research Data Management … in the Library?” *Insights: The UKSG Journal* 26, no. 3 (November 1, 2013): 298–304, https://doi.org/10.1629/2048-7754.99.

11. Xia and Wang, “Competencies and Responsibilities of Social Science Data Librarians”; Rebecca A. Brown, Malcolm Wolski, and Joanna Richardson, “Developing New Skills for Research Support Librarians,” *Australian Library Journal* 64, no. 3 (July 3, 2015): 224–34, https://doi.org/10.1080/00049670.2015.1041215; Lisa Federer, “Defining Data Librarianship: A Survey of Competencies, Skills, and Training,” *Journal of the Medical Library Association* 106, no. 3 (July 2018): 294–303, https://doi.org/10.5195/jmla.2018.306; Lyn Robinson and David Bawden, “‘The Story of Data’: A Socio-Technical Approach to Education for the Data Librarian Role in the CityLIS Library School at City, University of London,” *Library Management* 38, no. 6/7 (August 15, 2017): 312–22, https://doi.org/10.1108/LM-01-2017-0009; Mary Anne Kennan, “‘In the Eye of the Beholder’: Knowledge and Skills Requirements for Data Professionals,” *Information Research: An International Electronic Journal* 22, no. 4 (December 2017): 1601; Fabian Cremer, Claudia Engelhardt, and Heike Neuroth, “Embedded Data Manager, Embedded Research Data Management: Experiences, Perspectives and Potentials,” *Bibliothek Forschung Und Praxis* 39, no. 1 (April 2015): 13–31, https://doi.org/10.1515/bfp-2015-0006; Andrew M. Cox and Sheila Corrall, “Evolving Academic Library Specialties,” *Journal of the American Society for Information Science and Technology* 64, no. 8 (August 2013): 1526–42, https://doi.org/10.1002/asi.22847.

12. Federer, “Defining Data Librarianship.”

13. Brown, Wolski, and Richardson, “Developing New Skills for Research Support Librarians.”

14. Federer, “Defining Data Librarianship.”

15. Kennan, “In the Eye of the Beholder.”

16. Kennan, “In the Eye of the Beholder.”

17. Cox and Corrall, “Evolving Academic Library Specialties.”

18. Minglu Wang, “Supporting the Research Process through Expanded Library Data Services,” *Program-Electronic Library and Information Systems* 47, no. 3 (2013): 282–303, https://doi.org/10.1108/PROG-04-2012-0010; Ricardo L. Punzalan and Adam Kriesberg, “Library-Mediated Collaborations: Data Curation at the National Agricultural Library,” *Library Trends* 65, no. 3 (WIN 2017): 429–47, https://doi.org/10.1353/lib.2017.0010; T.P. Bardyn, T. Resnick, and S.K. Camina, “Translational Researchers’ Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library,” *Journal of Web Librarianship* 6, no. 4 (2012): 274–87, https://doi.org/10.1080/19322909.2012.730375.

19. Meredith N. Zozus et al., “Analysis of Professional Competencies for the Clinical Research Data Management Profession: Implications for Training and Professional Certification,” *Journal of the American Medical Informatics Association* 24, no. 4 (July 2017): 737–45, https://doi.org/10.1093/jamia/ocw179; Gabriele Schnapper et al., “Data Managers: A Survey of the European Society of Breast Cancer Specialists in Certified Multi-Disciplinary Breast Centers,” *Breast Journal* 24, no. 5 (October 2018): 811–15, https://doi.org/10.1111/tbj.13043; Martin Dugas and Susanne Dugas-Breit, “Integrated Data Management for Clinical Studies: Automatic Transformation of Data Models with Semantic Annotations for Principal Investigators, Data Managers and Statisticians,” *Plos One* 9, no. 2 (February 28, 2014): e90492, https://doi.org/10.1371/journal.pone.0090492; R. Esser, “Biostatistics and Data Management in Global Drug Development,” *Drug Information Journal* 35, no. 3 (September 2001): 643–53, https://doi.org/10.1177/009286150103500302; George Hripcsak et al., “Health Data Use, Stewardship, and Governance: Ongoing Gaps and Challenges: A Report from AMIA’s 2012 Health Policy Meeting,” *Journal of the American Medical Informatics Association* 21, no. 2 (March 2014): 204–11, https://doi.org/10.1136/amiajnl-2013-002117; Hong Huang et al., “Prioritization of Data Quality Dimensions and Skills Requirements in Genome Annotation Work,” *Journal of the American Society for Information Science and Technology* 63, no. 1 (January 2012): 195–207, https://doi.org/10.1002/asi.21652; Crystal Kallem, “Data Stewardship,” *Journal of the American Health Information Management Association* 79, no. 9 (2008): 58–59, 63.

20. John Cartwright, Jesse Varner, and Susan McLean, “Data Stewardship: How NOAA Delivers Environmental Information for Today and Tomorrow,” *Marine Technology Society Journal* 49, no. 2 (April 2015): 107–11, https://

doi.org/10.4031/MTSJ.49.2.11; T.A. Boden, M. Krassovski, and B. Yang, "The AmeriFlux Data Activity and Data System: An Evolving Collection of Data Management Techniques, Tools, Products and Services," *Geoscientific Instrumentation Methods and Data Systems* 2, no. 1 (2013): 165–76, https://doi.org/10.5194/gi-2-165-2013; A.J. Barrett, "Socioeconomic Aspects of Materials Data: Serving the User," *Journal of Chemical Information and Computer Sciences* 33, no. 1 (February 1993): 22–26, https://doi.org/10.1021/ci00011a004; Kenneth R. Knapp, "Scientific Data Stewardship of International Satellite Cloud Climatology Project B1 Global Geostationary Observations," *Journal of Applied Remote Sensing* 2 (2008): 023548, https://doi.org/10.1117/1.3043461; Kenneth R. Knapp, John J. Bates, and Bruce Barkstrom, "Scientific Data Stewardship: Lessons Learned from a Satellite-Data Rescue Effort," *Bulletin of the American Meteorological Society* 88, no. 9 (September 2007): 1359–61, https://doi.org/10.1175/BAMS-88-9-1359; Xin Li et al., "Toward an Improved Data Stewardship and Service for Environmental and Ecological Science Data in West China," *International Journal of Digital Earth* 4, no. 4 (2011): 347–59, https://doi.org/10.1080/17538947.2011.558123; R.R. Downs and R.S. Chen, "Designing Submission and Workflow Services for Preserving Interdisciplinary Scientific Data," *Earth Science Informatics* 3, no. 1 (2010): 101–10, https://doi.org/10.1007/s12145-010-0051-6; R.R. Downs et al., "Data Stewardship in the Earth Sciences," *D-Lib Magazine* 21, no. 7/8 (2015), https://doi.org/10.1045/july2015-downs; Helena Karasti et al., "Knowledge Infrastructures: Part I (Guest Editorial)," *Science & Technology Studies* 29, no. 1 (2016), http://ojs.tsv.fi/index.php/sts/article/download/55406/pdf_1; E.A. Kihn and C.G. Fox, "Geophysical Data Stewardship in the 21st Century at the National Geophysical Data Center (NGDC)," *Data Science Journal* 12 (2013): WDS193–96, https://doi.org/10.2481/dsj.WDS-033; T.P. Lauriault, P.L. Pulsifer, and D.R.F. Taylor, "The Preservation and Archiving of Geospatial Digital Data: Challenges and Opportunities for Cartographers," *Lecture Notes in Geoinformation and Cartography*, no. 9783642127328 (2010): 25–55, https://doi.org/10.1007/978-3-642-12733-5_2; D.J. Lowe, "The Geological Data Manager: An Expanding Role to Fill a Rapidly Growing Need," *Geological Society Special Publication* 97 (1995): 81–90, https://doi.org/10.1144/GSL.SP.1995.097.01.10; R.D. McDowall, "Understanding Data Governance, Part I," *Spectroscopy (Santa Monica)* 32, no. 2 (2017): 32–38; C.J. Moore and R.E. Habermann, "Core Data Stewardship: A Long-Term Perspective," *Geological Society Special Publication* 267 (2006): 241–51, https://doi.org/10.1144/GSL.SP.2006.267.01.18; T. Nadim, "Data Labours: How the Sequence Databases GenBank and EMBL-Bank Make Data," *Science as Culture* 25, no. 4 (2016): 496–519, https://doi.org/10.1080/09505431.2016.1189894.

21. "European Open Science Cloud," *Nature Genetics* 48, no. 8 (2016): 821, https://doi.org/10.1038/ng.3642.

22. Ibid.

23. Cartwright, Varner, and McLean, "Data Stewardship."

24. Boden, Krassovski, and Yang, "The AmeriFlux Data Activity and Data System"; Li et al., "Toward an Improved Data Stewardship and Service for Environmental and Ecological Science Data in West China"; Kihn and Fox, "Geophysical Data Stewardship in the 21st Century at the National Geophysical Data Center (NGDC).

25. Cartwright, Varner, and McLean, "Data Stewardship"; Knapp, Bates, and Barkstrom, "Scientific Data Stewardship"; Huang et al., "Prioritization of Data Quality Dimensions and Skills Requirements in Genome Annotation Work.

26. Rikk Mulligan, *SPEC Kit 350: Supporting Digital Scholarship (May 2016),* SPEC Kit (Chicago, IL: Association of Research Libraries, 2016), https://doi.org/10.29242/spec.350.

27. Clements, "Research Information Meets Research Data Management … in the Library?"

28. Teperek et al., "Data Stewardship Addressing Disciplinary Data Management Needs."

29. Teperek et al., "Data Stewardship Addressing Disciplinary Data Management Needs."

30. Matt Greenhall, "Digital Scholarship and the Role of the Research Library," The result of the RLUK digital scholarship survey (London: RLUK, 2019), https://www.rluk.ac.uk/wp-content/uploads/2019/07/RLUK-Digital-Scholarship-report-July-2019.pdf.

31. Saša Baškarada and Andy Koronios, "Unicorn Data Scientist: The Rarest of Breeds," *Program* 51, no. 1 (January 1, 2017): 65–74, https://doi.org/10.1108/PROG-07-2016-0053; Kennan, "In the Eye of the Beholder."

32. Kennan, "In the Eye of the Beholder."

33. Teperek et al., "Data Stewardship Addressing Disciplinary Data Management Needs."

34. Kathryn Lage, Barbara Losoff, and Jack Maness, "Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder," *portal: Libraries and the Academy* 11, no. 4 (2011): 915–937; Kevin Crowston, "User Personas" (DataOne - Data Observation Network for Earth, 2015), https://www.dataone.org/personas/abby-science-data-librarian; Lorna Wildgaard et al., "National Coordination of Data Steward Education in Denmark: Final Report to the National Forum for Research Data Management (DM Forum)" (National Forum for Research Data Management (DM Forum), 2020), https://doi.org/10.5281/zenodo.3609516.

35. Jack M. Maness, Tomasz Miaskiewicz, and Tamara Sumner, "Using Personas to Understand the Needs and Goals of Institutional Repository Users," *D-Lib Magazine* 14, no. 9/10 (2008): 15.

36. Crowston, "User Personas."

37. Crowston, "User Personas."

38. Lage, Losoff, and Maness, "Receptivity to Library Involvement in Scientific Data Curation."

39. Wildgaard et al., "National Coordination of Data Steward Education in Denmark."

40. Erio Ziglio, "The Delphi Method and Its Contribution to Decision-Making," in *Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*, eds. Michael Adler and Erio Ziglio (London, UK: Jessica Kingsley, 1996), 3–33.

41. H. Rex Hartson and Pardha S. Pyla, *The UX Book: Process and Guidelines for Ensuring a Quality User Experience* (Amsterdam, Netherlands; Boston, MA: Elsevier, 2012).

42. Hartson and Pyla, *The UX Book*, 266.

43. Live Kvale and Nils Pharo, "Understanding the Data Management Plan as a Boundary Object through a Multi-Stakeholder Perspective" (forthcoming), *International Journal of Digital Curation* 15, no. 1 (2021).

44. John W. Creswell and Vicki L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. (SAGE Publishing, 2018), 80.

45. Johnny Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd ed. (London, UK: SAGE Publishing, 2016).

46. Data from a three-phase Delphi study used to investigate Knowledge Infrastructure for Research Data in Norway, KIRDN_Data, 2020, http://doi.org/10.5281/zenodo.3673053.

47. Kennan, "In the Eye of the Beholder"; Baškarada and Koronios, "Unicorn Data Scientist."

48. OECD, ed., *Data-Driven Innovation: Big Data for Growth and Well-Being* (Paris, France: OECD Publishing, 2015), http://dx.doi.org/10.1787/9789264229358-en.

49. Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired* (June 2008), https://www.wired.com/2008/06/pb-theory/.

50. Bryan P. Heidorn, "Shedding Light on the Dark Data in the Long Tail of Science," *Library Trends* 57, no. 2 (2008): 280–99, https://doi.org/10.1353/lib.0.0036.

Article 2

Kvale & Pharo (2021). Understanding the Data Management Plan as a Boundary Object from a Multi-Stakeholder Perspective. *International Journal of Digital Curation* 16 (1)

DOI: **https://doi.org/10.2218/ijdc.v16i1.746**

IJDC | *Research Paper*

Understanding the Data Management Plan as a Boundary Object through a Multi-Stakeholder Perspective

Live Kvale Nils Pharo

Abstract

A three-phase Delphi study was used to investigate an emerging community for research data management in Norway and their understanding and application of data management plans (DMPs). The findings reveal visions of what the DMP should be as well as different practice approaches, yet the stakeholders present common goals. This paper discusses the different perspectives on the DMP by applying Star and Griesemer's theory of boundary objects (Star & Griesemer, 1989). The debate on what the DMP is and the findings presented are relevant to all research communities currently implementing DMP procedures and requirements. The current discussions about DMPs tend to be distant from the active researchers and limited to the needs of funders and institutions rather than to the usefulness for researchers. By analysing the DMP as a boundary object, plastic and adaptable yet with a robust identity, translating between worlds (Star & Griesemer, 1989) where collaboration on data sharing can take place, we expand the perspectives and include all stakeholders. An understanding of the DMP as a boundary object can shift the focus from shaping a DMP which fulfils funders' requirements to enabling collaboration on data management and sharing across domains using standardised forms.

Submitted 18 December 2019 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Live Kvale, Pilestredet 35, 0166 Oslo. Email: live.kvale@oslomet.no

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The data management plan (DMP) is promoted as a tool for ensuring good data management and a first step for making data as open as possible, enhance reproducibility and reusability of collected data and avoid data loss (Michener 2015). Funders are increasingly requiring DMPs to be submitted along with research proposals and updated during the research projects (European Research Council 2017). Following funders requirements, a growing number of either discipline or funder specific templates and tools for DMPs have been developed. Within the research data management community the current focus is on making machine actionable, readable and interoperable DMPs exploiting the "thematic, machine-actionable richness with added value for all stakeholders" (Miksa et al. 2019). Several studies on DMPs take a quantitative approach to measuring effects, either in actual shared data or as successful funding (Diekema, Wesolek, and Walters 2014; Johnson and Knuth 2016; Mischo, Schlembach, and O'Donnell 2014; Van Loon et al. 2017; Westra 2017). Other articles focus on the importance of writing a DMP (Nature 2018) or how to write one (Burnette, Williams, and Imker 2016; Michener 2015; Wright 2016). All are useful and applied approaches with a focus on meeting requirements and receiving funding. This paper takes a different approach by aiming to understand the DMP as an object and document in the research process by investigating how the DMP is perceived by different stakeholders that all claim an interest in the plan and the planning.

The research questions investigated in this paper are:

1) What perspectives on DMPs are held by different stakeholder groups?
2) How do these perspectives help or hinder DMPs as tools to support data management?

Background

In 2017, the European Union's[1] (EU) Horizon 2020[2] (H2020) programme updated the EU's research data policies to require that new projects funded by the programme had to create a DMP (European Commission, 2016; European Research Council, 2017). A policy document from the Norwegian Ministry of Education and Research (Kunnskapsdepartementet, 2017) made recommendations for universities and university colleges to require DMPs. The policy document described the DMP as a document containing plans for how research data will be managed through the research lifecycle to make data sharing an embedded part of the workflow. Further it should be a guiding document to help researchers in the project planning. Also the DMP should aid institutions, the research council and others in ensuring that the requirements are met, and, it should serve as inspiration for other researchers to learn best practice (Kunnskapsdepartementet, 2017). The DMP is also expected to increase awareness and improve the way researchers document data and to enhance reproducibility (Kunnskapsdepartementet, 2017). The current template for H2020 DMPs from the European Research Council (ERC) focuses on how data can be made findable, accessible, interoperable and reusable (FAIR) and describes the costs associated with data management (European Research Council, 2017). Unlike the EU (European Research Council, 2017) guidelines, the policy document from the Norwegian Ministry of Education and Research emphasises that each institution is responsible for approving the DMPs (Kunnskapsdepartementet, 2017), and calls for

---

[1] Also funding associated countries, such as Norway, under the same conditions.
[2] https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020

the design of a tool to support development of DMPs. This has led to two national research data storage providers developing and publishing generic DMP tools[3].

In the EU and in Norway as addressed in this study, DMPs are relatively new to all stakeholders, including the researchers. The European Union ran a pilot requesting DMPs from selected thematic areas funded by Horizon 2020 between 2014 and 2016. In 2017, this pilot was extended to cover all areas of Horizon 2020 (European Commission, 2013, n.d.). In Norway, a pilot on DMPs for climate research was done from 2014 to 2015, the aim being to learn whether DMPs would encourage more data sharing (The Research Council of Norway, 2014). This attempt was regarded as unsuccessful due to a lack of experience and knowledge for evaluating the DMPs amongst reviewers (Schjølberg, 2015). In its 2017 policy, the Norwegian Ministry of Education and Research points to research institutions as responsible for assessing DMPs. In doing so, they shift the focus from the evaluation of DMPs as part of funding applications to the creation of DMPs as part of research workflows. Consequently, universities are now establishing workflows for DMPs (NTNU, n.d.; The Artic University of Norway, 2019; University of Oslo, 2019). There are no national guidelines or criteria for evaluation of DMPs.

In the United States, DMPs have been a standard requirement in grant application for a decade (Mischo et al. 2014), and analysis of DMP guidelines and DMPs is an established part of the literature on data management (Berman, 2017; Burnette et al., 2016; Diekema et al., 2014; Dressel, 2017; Hardy, Hughes, Hulen, & Schwartz, 2016; Johnson & Knuth, 2016; Thoegersen, 2015; Van Loon et al., 2017; Williams et al., 2017; Wright, 2016). In reading and analysing the literature two tendencies become evident. Studies on DMPs tend to present the perspectives of one or two stakeholders and thus cover different aspects of and approaches to the DMP. Steinhart, Chen, Arguillas, Dietrich and Kramer (2012) use a survey to investigate researchers' experiences with DMPs. Researchers' perspectives are found in case studies describing how the DMP was applied in a research group (Burnette et al., 2016; Dressel, 2017) or discipline (Dressel, 2017). Other studies use quantitative approaches to measure the effect of DMPs either by grant success rates (Mischo et al., 2014) or by evaluating the effectiveness of research support by assessing the quality of DMPs (Johnson & Knuth, 2016; Van Loon et al., 2017). These studies present a research support perspective. Two studies have used content analysis to assess the requirements from the funders (Thoegersen, 2015; Williams et al., 2017). The results from these studies are useful for assistance in the writing of DMPs. Diekema and colleagues (2014) investigate researchers, research offices, and academic libraries in the role of infrastructure providers. They find that although researchers often are positive towards sharing data, they lack the necessary skills to do so. Researchers were unfamiliar with data repositories and existing data management services from the library. Further they noted that data management mandates had little impact on the workflow of researchers and research office respondents. Diekema and colleagues propose that the library needs to make researchers aware of existing research data management services and infrastructure to bridge the data management skills gap (Diekema et al. 2014).

Current literature on DMPs presents an applied approach to the document as a tool, and on how making the DMPs machine-actionable can be beneficial for multiple stakeholders (Cardoso, Proença, and Borbinha 2020; Miksa et al. 2019; Simms et al. 2017). Less emphasis is placed on the content of the plan, why the plan is written and for whom. The review of DMP literature by Smale and colleagues (2020) does, however, suggest that there is no evidence to support a claim that researchers benefit from filling out a DMP. This suggests that it might be a good idea to take one step back and problematise the influence of the varying interests held by different stakeholders when creating a DMP and the tension between these interests.

Leading theorists in the area of data management emphasise that different stakeholders sometimes hold conflicting interests (Bowker, 2005, p. 123) and that including multiple stakeholders when examining the functions of data in scholarship is important (Borgman, 2015, p. 14). We believe this multi-stakeholder approach is the strength of the study presented in this

---

[3] NSD: https://nsd.no/arkivering/en/data_management_plan.html
Sigma2: https://www.sigma2.no/content/easydmp

paper, providing a broader understanding of the DMP, which is helpful in the practical approach to writing a DMP.

Theoretical Framework

The main characteristic of boundary objects is that they mean different things to groups of people working in different contexts and facilitate coordination and collaboration between these different groups. According to Star and Griesemer:

> 'Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use and become strongly structured in individual-site use. These objects may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognisable, a means of translation' (Star & Griesemer, 1989).

In this paper, we will focus on standardised forms, described as methods of communication between different groups with different interests. One example of a standardised form which Star and Griesemer use in their original work on boundary objects is a document of procedures for data collection and curation, 'a precise set of procedures for collecting and curating specimens' (Star & Griesemer, 1989). This description of a precise set of procedures for collection has strong similarities with some of the descriptions we found of the DMP. Still, Star acknowledges the challenges in collecting, disciplining and coordinating distributed knowledge (Star, 2010). Her example of the complexity of creating forms is from a research study on epileptic patients from late nineteenth century England. She found that much information was scribbled down on the edge of the form because it did 'not fit the actual form' (Star, 2010). The information was later discarded as unimportant because it was not part of the information family members of the epileptic patients were asked to collect. Star describes these documents as "revealing the relations of class and medicine" in England at the time. Thus, Star asks, 'how do forms shape and squeeze out what can be known and collected?' (Star, 2010). This is a problem which emerges frequently in the era of automation and digital forms: there is often little space for scribbling on the side. The different approaches to the standardised form, either as a set of procedures or boxes to fill in supplying the requested information, call for different levels of involvement from the contributor.

Method

We have used a modified Delphi study (Ziglio, 1996) to explore the understanding and application of DMPs among different Norwegian stakeholders involved in research data sharing. A Delphi study is characterised by the use of an expert panel to elicit opinions on a shared reality from different perspectives. Data collection is performed in several rounds with the intention of reaching consensus or solving an issue.

A group of 24 experts took part in the study. Table 1 contains an overview of the participants. The group consisted of policy-makers, representatives of national service providers, and researchers and research support staff from four Norwegian universities. The participants were invited based on their involvement in the development of policies, infrastructure or data-related research support. The research support staff were recruited to include representatives from different research support services at the universities, including libraries, research offices and IT departments.

We invited researchers who were appointed as project owners of H2020-funded projects to participate in the panel. Of the 25 researchers contacted, eight participated. These eight researchers hailed from different disciplinary backgrounds (biology, musicology, science studies, economics, neuroscience, psychology, philosophy, gender studies), and they differ in levels of prior knowledge of research data management.

The participants were not promised anonymity, only that their names would not be used. Identification might be possible with triangulation and local knowledge. As a result, quotes in cases in which the informant does not wish to be identified in connection with the statement do not include the full participant code.

Table 1. The participants organised according to role

Role/stakeholder category	Participant code			
Researchers working individually	RIZ	RIJ	RIL	RIB
Researchers working in groups	RGV	RGD	RGA	RGW
Policy-makers	POU	POS	POK	
Service providers	INH	INO	INR	
Research support IT	ITE	ITY	ITI	
Research support, research office	ROC	ROX	ROT	
Research support, library	LM	LP	LG	LN

Data were collected in three phases, as shown in Figure 1. The first phase, the 'exploration phase', was conducted using open interviews lasting approximately one hour in January/February 2018. The purpose of this phase was to obtain an initial overview of the panel members' opinions on the DMP, or 'defining the problem'. Interviewees were asked a set of questions concerning research data management, including their needs for data management, their experiences with DMPs and their perceptions of the aim of a DMP.

In the second phase, the 'evaluation phase', conducted in August/September 2018, participants answered a survey containing nine questions on topics such as data stewardship, DMPs, ethical aspects of data sharing and core functions in a research data infrastructure. The survey was designed to further explore issues and tensions uncovered in the first interviews. Several of the questions were formulated as statements that the participants were asked to agree or disagree upon.

The third, 'concluding phase', was conducted using interviews in March/April 2019. These interviews lasted approximately 30 minutes and were based on results derived from the two former phases. Among the questions asked in the final interview were how does the DMP best reflect the different needs of the different stakeholders, and participants were asked whether they had thoughts on the preliminary findings of this study, such as the differences reported by researchers working individually and in groups.

Based on requests from some of the participants, the questions were sent to all participants prior to the data collection, in all three phases. The participants were also sent the transcripts from the interviews and were asked for permission to share the complete material or parts of the data material to which they contributed. The data is available in Zenodo (Kvale, 2020). In this paper, data regarding DMPs from all three phases are reported and integrated in the analyses (Creswell & Plano Clark, 2018, p. 80). The interviews were qualitatively coded and analysed thematically (Saldaña, 2016).

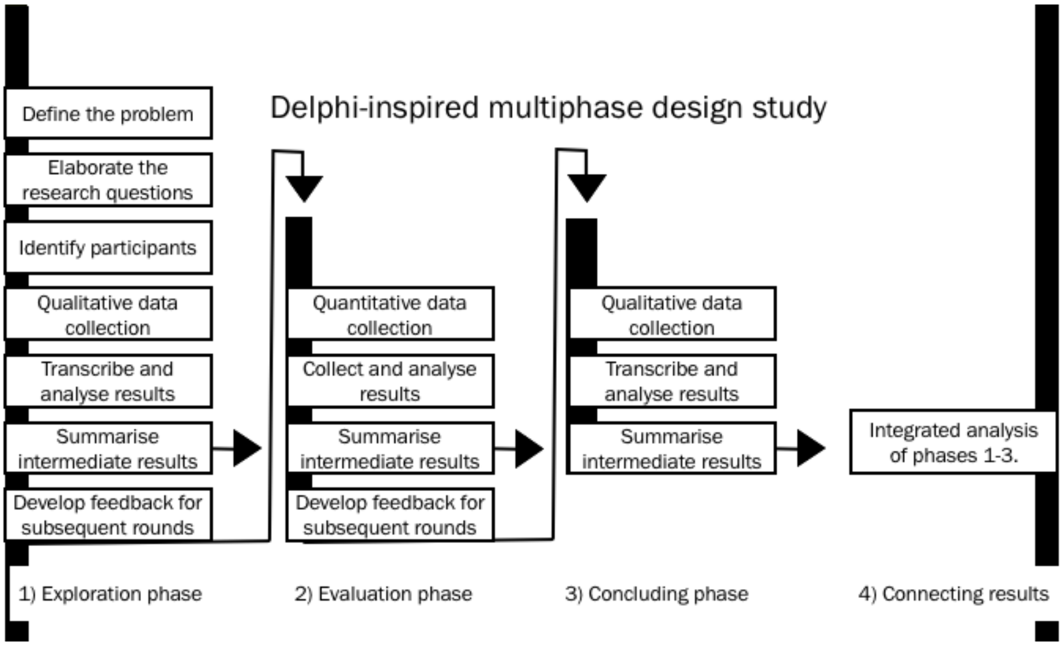


Figure 1. A Delphi-inspired multiphase design study.

Findings

The findings reported here are based on the integrated analysis of the material from all three phases of the study. Findings were subsequently grouped according to three main themes to highlight different issues regarding the DMP: Sharing a common goal, Different perspectives on the DMP, and Different practice approaches to the DMP.

Sharing a common goal

Analysis of the first interviews revealed 12 different perceived purposes for using a DMP amongst our participants (Table 2; rows A-L). These purposes were used in the survey to understand the extent to which there was agreement among stakeholders about the purpose of a DMP. To cover other views expressed in the exploration phase, three additional options were added (rows M-O). In the survey, the participants were presented with a list of purposes for making a DMP and were asked to select the five most important reasons to make a DMP. The third column in Table 2 shows the number of times each of the aims was selected.

Table 2. Aims of the DMP. (n = 24 participants)

	Reasons to make a DMP	Frequency
G	It makes the researcher think about how to make their data and metadata FAIR (findable, accessible, interoperable, reusable).	21
H	A DMP creates awareness and agreement on data procedures within a research group, which is particularly important when several partners with different data needs are involved.	19
B	When researchers are in control of their own data management, the research gets better and more efficient.	15

D	When researchers think about what they are doing at an early stage, they can make intelligent choices for their data.	15
K	A DMP is a plan with a commitment to making the data as open as possible.	6
C	A DMP gives universities an overview of ongoing research projects.	5
A	To create awareness in the research community of the need for data stewardship.	5
L	A DMP makes researchers aware of the value of what they are collecting.	4
F	A DMP gives the archive the possibility to plan for data which are going to be deposited there.	3
O	I am not familiar with data management plans.	3
E	A DMP provides the archives with information (metadata) they need when data is deposited.	3
N	I don't see why DMPs are important.	2
M	Other reasons: 'Power and competency to avoid ethical brakes in terms of personal privacy'.	1
J	A DMP is primarily a tool for those who take care of the data after it has left the project.	1
I	A DMP shows what data will be produced in a project when the project is finished.	1

Four aims (G, H, D and B) were selected significantly more often than the others (by 15 to 21 of the participants).

The most important reasons to make a data management plan (as selected by participants) are:

G. The DMP makes the researcher think about how to make their data and metadata FAIR (findable, accessible, interoperable, reusable).

H. A DMP creates awareness and agreement on data procedures within a research group, which is particularly important when several partners with different data needs are involved.

D. When researchers think about what they are doing at an early stage, they can make intelligent choices for their data.

B. When researchers are in control of their own data management, the research gets better and more efficient.

All of these reasons emphasise the researcher both as the creator of the plan and the primary beneficiary of thorough planning. The different stakeholders agreed on a common goal of a data management plan. Aims G, H, D and B have in common that the goal of a DMP is to improve data management by making researchers plan for sharing their data internally within research groups and externally (FAIR) by creating procedures for documentation and collection at an early stage. The survey brings the areas of agreement to the surface, and it therefore appears to be a broad agreement among different groups of stakeholders about the purpose and role of DMPs. The interviews, however, tell a different story, with perspectives and approaches varying according to the different contexts in which each group of stakeholders work.

Different perspectives on the DMP

The first interviews reveal five different perspectives on the DMP reflected by the vison of the stakeholders and have been analysed, grouped and labelled accordingly.

The participants representing policy-makers and research support services agreed largely on the DMP being a reflection of the extent to which data can be shared and on how data sharing is an aspect of open research. One of the librarians stated ‘it is about the researchers already in the design phase reflecting on how to work as openly as possible’ (LG). Meanwhile, one of the policy-makers focussed on the management, publication and associated costs for which the DMP should be used to prepare: ‘What type of data to collect and how to take care of them, how to make them available and possibly how to fund data management’ (POU). Another policy-maker focused on similar aspects by putting forward the need for data stewardship: ‘It is for the whole research environment to become aware of their need for data stewardship’ (POK). Both policy-makers hold a funder perspective on the DMP, emphasising that it is used to manage how data can be made available and enable the calculation of data sharing costs. The librarian, on the other hand, focussed more on the structured planning for the research process with data sharing as the ultimate goal. We have labelled these *the sharing and open science perspective* and *the stewardship perspective*, respectively.

Researchers have divergent views on what the DMP is, based on whether they work in collaborative environments or in more individual-based research environments in which the sharing of data among collaborators is less common. RGV, RGW, and RGD are all researchers who work in groups in which data is shared both within the group and with external partners. They described what is categorised as an internal *protocol and procedures perspective* on the DMP. In contrast, he individual-oriented researchers (RIZ, RIB, and RIL) had no experience with DMPs. Both RGW and RGV described the DMP as a document used for agreeing on standard procedures. As such, the DMP becomes more of an internal document for the respective research group. Researcher RGD described the entire research project as a DMP: ‘Actually, the whole project is like a big data management plan’ (RGD). The research project RGD is referring to combines data from different locations and previous research in a new databank for the researchers to collaborate on. As it is described, the research project itself is to a large extent about managing data, and the description of the project becomes the DMP. In the second interview, one year later, RGD described another DMP document used in the same project. In it, the data manager had created a detailed protocol for how to work with the data in the project to ensure that all researchers involved in the project followed the same procedures when working on the existing data or adding new data.

Researchers working independently or in collaborations in which there is little or no sharing of data among collaborators, express less knowledge of the DMP. Researchers RIZ, RIB, and RIL were, as stated above, unfamiliar with DMPs. However, RIB was familiar with aspects of data management and shared comprehensive descriptions of how she analysed data and how the data could be accessed as supplements to journal articles. Documenting data was a clear part of RIB’s research, even if there were restrictions on sharing the data. Researcher RIJ is a researcher within philosophy with experience of ethical committees and a strong interest in privacy protection and research ethics. RIJ’s understanding of a DMP was similar to that of the policy-makers and research support staff, with more emphasis on aspects regarding personal privacy.

The service providers presented a more differentiated view. INR focussed on how the DMP is useful for several stakeholders, stating it is ‘a tool for planning with archiving and sharing in mind’ (INR) and, ‘for the researcher and the institutions to make sure their researchers fulfil the demands’ (INR). This aspect of control for the institution was not promoted by other stakeholders in this study. INO focussed on decisions that should be made prior to data collection: ‘for the researcher to think about what he is doing at an early stage, so he can make intelligent choices’ (INO). INH emphasised that the DMP is a document the researchers do not create themselves: ‘When I got this task [to create a DMP tool], I thought those that are using my tool are going to be researchers, but although we put all the effort to facilitate the creation of a DMP, it still requires some competence in data management that is not likely to be present in the end user, in the researcher’ (INH). These quotes from the service providers present a *curating perspective* and *fulfilling requirements perspective*.

Stakeholders view of users

The survey asked participants to name who the DMP primarily is written for and rank the users of the DMP according to their importance on a scale of 1-3. The results show that the participants to a large degree agree on the DMP being for the researcher (a score of 52 out of 72). Other central users of the DMP are the research institutions (24/72) and other researchers (17/72). In addition, funders (9/72), archives (5/72) and data stewards (2/72) were mentioned. To illustrate the different opinions, a relational visualization is used to show relationships and connections between the data (Figure 3). This visualization shows that the different stakeholder groups point at different users of the DMP, and that there is no clear coherence in the responses apart from the common agreement about the researcher. All participants point at several users of the DMP which again illustrates the different perspectives of the plan listed above. Accordingly, DMPs should be developed to be used by different stakeholders for different purposes, with primary focus on the researcher. To follow up on this aspect we invited the participants to give concrete suggestions on practice approaches to how the plan could be perceived as useful for the different stakeholders.

Different practice approaches to the DMP

In the final interviews, the participants were asked to make suggestions about how the DMP could be developed to respond to the needs of its different users, and specifically about how the DMP should be developed to become a useful document. The different stakeholder groups suggested different approaches to developing the DMP as a practical tool.

Among the research support services, participants emphasised the DMP as a document in which to display and encourage best practices on data management by embedding checklists and good examples. The document should, according to them, be developed to reflect the researchers' perspective on the research process. One of the participants working in IT support put it like this:

> 'You should give some kind of best practice, both tips and strategies, for how you as a researcher should do best practice on data. If you ask questions that not just irritate the researcher but rather enlighten them on something they didn't know. The questions should make them think 'wow, I do have to think about this', I believe that this could be a useful approach' (ITI).

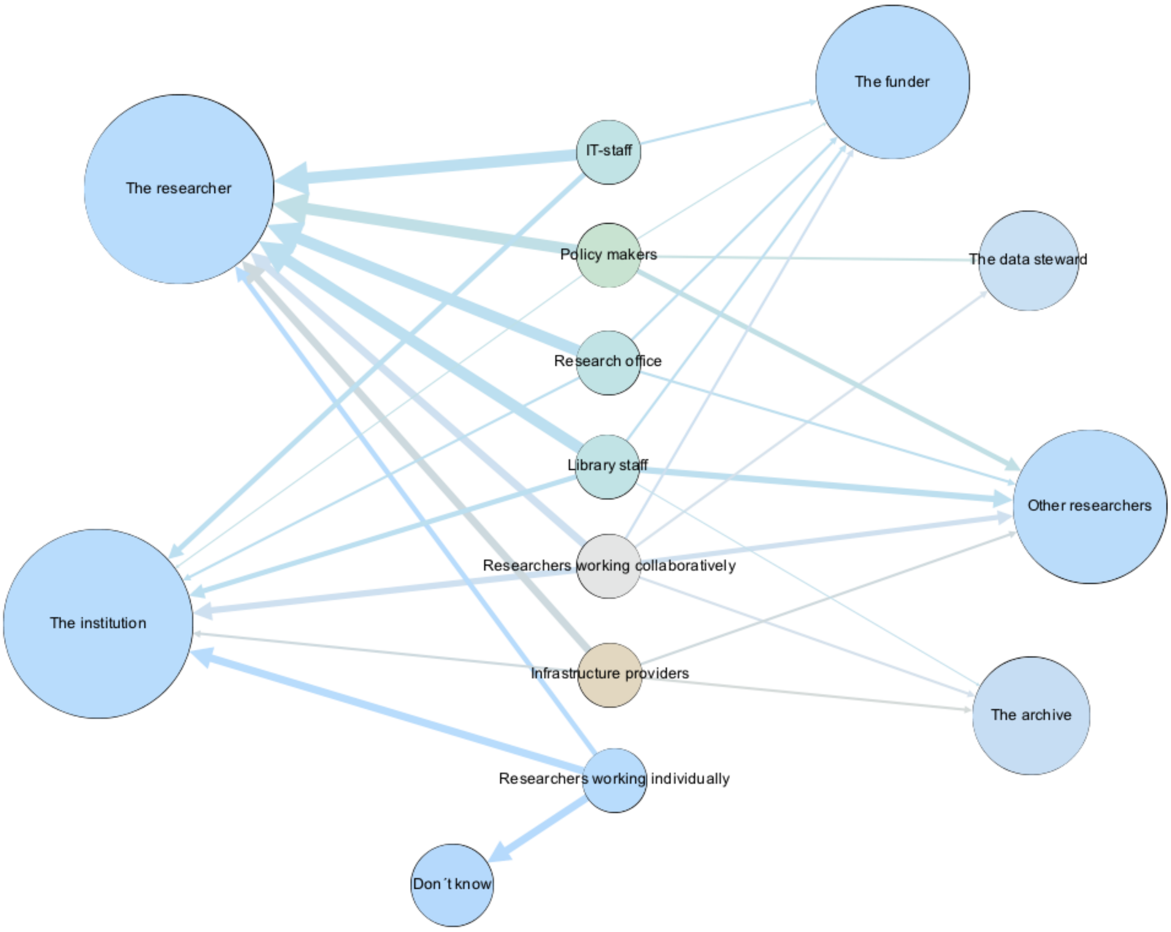


Figure 2. Relational visualization of responses to the question: Who is the data management plan primarily for?

One of the library staff participants described it in a similar way:

> ‘What I think and believe will be important is informative help texts for the different sections, in a way translating the computer syntax, that you might have a bullet point checklist or question that the researchers should ask themselves when answering that section’ (LM).

Aiming at creating guidance that encourages researchers to reflect on data management practices was typical among the research support staff. One participant was concerned about how the interests of the other stakeholders could reduce the plans’ relevance and usefulness for the researchers:

> ‘We see in Norway that there are some entities who think they should use the DMP for all other types of purposes, to their own advantage. And it is possible that there will be types of secondary use, for the institutions to monitor research, and for archives it might make the archiving process easier, but that should not be the primary function, and one should not create templates focussing on this instead of the researchers’ needs’ (LG).

Their concern was that the interests of other stakeholders will make the DMP less useful for researchers. The research support staff therefore suggested the creation of a list of questions formulated in such a way that researchers find it useful to reflect upon the questions, supplemented with tips and best practices on how to improve data management.

The researchers were looking to their international research communities for best practice and evolving methods, standardisation and expectations. This was expressed both by RIB and RGW:

> 'It is evolving. Some groups have high data management standards and you try to adapt to it. But that is also expensive, depending on what you do, so then if that is the standard, we have to invest in that kind of resources. So, specifically, in the data we are influencing each other internationally' (RGW).

> 'Within many areas there has become an increased focus on reproducibility and then it is important to actually have the data, the codes and what else might be needed for replication' (RIB).

The increased awareness described by RIB was found among several stakeholders in the final interviews and will be investigated separately in a forthcoming paper. RGV described the need for simple language and relevance.

> 'I realize that some believe that the current forms [referring to experience with existing templates] are fairly simple and clear; at the same time it is in practice very difficult' (RGV).

RGV further presents the idea of using a decision tree to visualise and decide what is relevant for different researchers to consider in their plans and to supply the research office with extensive knowledge of data management to guide the researchers in writing the plan.

> 'I think what really matters, when these in reality often are complex issues, is another person between us and the [service provider] ….and the person managing this as an advisor must have extensive knowledge of the whole field, not just disciplinary glasses on, seeing only what is relevant in medicine or sociology, but one that understands the background for certain questions and understands what is important to maintain and legitimate interests for those who are part of the research' (RGV).

This request points at the data steward role, i.e. someone somewhere in the university with competence in data management.

One of the researchers described a lack of coherence between policy and practice:

> 'Basically, I had a chat with EU and they say that once your DMP is accepted, unless it is absolutely necessary, please do not spend too much time on it because the main goal is the scientific research, and we have only two years of funding, so then, yeah, so we keep it as such but there are not big changes in the way I manage my data, so I did not really think about it again' (Researcher).

This reflects notions of a funder with no clear interest in data management, revealing a conflict between the requirements for updating a DMP and the standard model (Bowker, 2005, p. 121) for scientific publication.

Among the policy-makers, there were different opinions on the DMP. However, they do agree that it should be a useful document for researchers. POK was clear in her opinion that, 'it should not be up to the authorities to specify what type of DMPs are good for the researchers to make; that is none of our business' (POK). Her point was that the researchers should write the DMP in a manner they find it useful rather than being forced to use a certain template or tool to create it. POU expressed concern about the extent to which the DMP is relevant for researchers in its current form:

> 'Today the DMP is formulated a bit like a questionnaire, generating a PDF and that's the way it gets, maybe not that useful, it becomes more like an exercise, a bureaucratic exercise' (POU).

She further explains how it is difficult to complete, 'Because concepts like metadata are not something most researchers relate to' (POU). She also proposed automating parts of the DMP creation process,

> 'So that some information can be automatically added, and others be automatically proposed. Ideally, the calculation of data management costs can increasingly be automated' (POU).

What POU proposed is automating as much of the DMP as possible so that researchers do not have to spend time on trivial questions. Her approach stands in contrast to that of both POK, who wished to minimise the formal requirements of a DMP, and POS, who described the function of the DMP in a way similar to that of the research support services.

> 'What I communicate to my researchers is that you will always generate data. Describing it in a DMP, even if you work on an individual research project, and being explicit about how you are going to structure your data, so that you actually can reproduce your results at a later stage, improves the research process [….] I don't know if this is currently reflected in the DMP, but I believe it is a way to strengthen the research process' (POS).

Among the service providers, we found a change in their views of the DMP compared with what they had expressed in the earlier phases of this study. One service provider, a technician and service provider delivering tools to the universities, referred to the DMP as a complex document serving several purposes and stakeholders:

> 'I have only become more uncertain about what a DMP is; the more I try to understand it, I am certain that I had a simpler perception of the problems last time we spoke, and I am a bit frustrated over my own lacking capability to get anywhere. Because it is important to very many, but for different reasons something that results in it being perceived as pretty useless for everyone because it tries to solve too many things at once' (INO).

This statement by INO reflects some of her difficulty in understanding research. In a prior stage of the study INO focussed on the institutions and their wish to have an overview of research data collected and control over where and how it was archived, but in the year which had passed she had become more uncertain about the DMP.

Another service provider, INH, emphasised meeting the needs of the researchers. In the first interview, INH did not believe that the researchers would be capable of filling out the DMP. In the later interview, however, she stated that:

> 'Researchers have to upgrade along the way in the research process, and this as the DMP should be drafted or ready since the very beginning because it is part of, I mean it is part of the research process itself. So, making the plan is not for the sake of making a plan; it is part of the research' (INH).

INH also suggested that the DMP should be developed to serve the needs of the researchers:

> ‘First of all, the researchers have to interact actively with the DMP, so it has to be in electronic form, it has to be modified and customised as much as possible, so this means that actually the guidelines should be really high-level guidelines, but the action form of the DMP should be dedicated to the community specific. Only in this way can you make sure that it is not an [exercise of] checking boxes’ (INH).

INH further emphasized that data management should be ‘customized to the scientific topic’ and a digital tool.

The understanding presented by INR overlaps with that of INH, who also struggled to develop a relevant DMP tool:

> ‘It should not just become a questionnaire ending in a document you send to the funder because they require it’ (INR).

Still, her approach to guiding researchers in the right direction is somewhat different from that of INH, who focused on a community specific DMP, whereas INR focussed on embedding a detailed level of institutional guidance:

> ‘There should be a guidance in the DMP so that when you answer questions you are guided in the right direction. So that with naming conventions, really what to name the files, and how to structure data, there might be similarities, and then you can get help and suggestions as to how you should name your files’ (INR).

Another suggestion by INR was to use a guide for the classification of data according to sensitivity, so that, while writing the DMP, the level of sensitivity is defined for the data to be collected.

The different practical approaches to DMPs point in different directions. The researchers requested on-the-ground support, and development of a peer-network to share best practices. The research support staff focussed on well-formulated questions encouraging researchers to reflect and make decisions for data management. Both approaches imply a lower level of automation and a higher level of flexibility or a more manual plan. Among the policy-makers, the opinions differed: one was clear that such decisions should be left to research communities, another focussed on the importance of the DMP as a useful tool for researchers, whereas the third suggested that more information should be automatically added, a notion shared by the infrastructure developers.

Discussion

The findings presented perceptions of the DMP held by the different stakeholders. The different stakeholders understand and apply the function of the DMP differently: The curating and fulfilling requirements perspective, the sharing and open science perspective, the stewardship perspective, and the protocol and procedures perspective. The perspectives illustrate the different backgrounds of the various stakeholders. The researchers reflect on how DMPs could be useful in a research group, while the service providers reflect on how they can be used to assist in planning for the archiving process and fulfilling formal requirements. In Star and Griesemer’s terms, this would constitute worlds of the different actors. Research, in general, and the sharing of research data, in particular, requires cooperation between different stakeholders. By writing a DMP, researchers plan for their data to move from collection through analysis and to sharing as was agreed upon in the goal of the DMP. In this sense, DMP creation facilitates translation between the different worlds and the different stakeholders as a standardised form by creating context for research data so that these can be understood and interpreted in the different worlds independently of disciplines, institutions or national boundaries.

The ideal DMP, as described by the support services participants, is a guiding tool which poses questions researchers have not thought about. The service providers presented a different understanding, focussing on meeting formal requirements and possibilities of automation. The researchers look to their international research communities for best practice and request support functions in their research institutions. However, in encounters with policy-makers, researchers are confronted with the blunt reality of little time and money to think about data quality and the continuous pressure to publish. The tension between different perspectives, the research reality and the higher goals, can be resolved by a common understanding of the DMP as a document which is not a product of consensus, but an everyday translation between worlds and communities. Star (2010) points at a common misunderstanding regarding the need for consensus as a basis for cooperation, claiming that the use of boundary objects can explain how cooperation can continue unproblematically without consensus. The DMP may perform the role of a boundary object for different data management stakeholders.

We find that there are two issues which need to be clarified in order for the DMP to function optimally as a standardised form translating between worlds, formalising procedures and standardising methods: the *degree of standardisation* and the *degree of automation*.

When it comes to degree of standardisation, the policy-makers problematised how the DMP today becomes more of a bureaucratic exercise than an actual plan and emphasised that they do not want to interfere with what should be in the plan. At the same time research communities are continuously developing best practices for data management and there is no static standard for how data management should be done. This suggests that a lower degree of standardization would give the DMP flexibility over time and across methods and disciplines. The DMPs should therefore be developed more as open documents to fill the needs of the researchers, in their planning for sharing of the data. The DMP is never a goal in itself, rather it is the reflection it triggers regarding data sharing that is the desired output. The different stakeholders agree on the goal of a DMP: To improve data management by making researchers plan for sharing their data internally within research groups and externally (FAIR) by creating procedures for documentation and collection at an early stage.

To achieve this, research support and infrastructure developers need to take one step to the side and leave the researchers with autonomy to shape the content of DMPs according to the design of their research projects. We suggest formulating open-ended questions concerning data management issues for researchers to reflect upon how data best can be structured and documented for reuse and sharing. In addition, researchers should be supported with best practices to ensure high-quality data management.

The degree of automation refers to tools for DMPs and how they should be developed. Information which is on a general level or project information could preferably be imported or connected to other sources. Harvesting data from DMPs to repositories and research administrations tools does, however, come at the cost of the autonomy of a plan. Automated decisions do require a preselection of options, which again would be limiting the possible choices for the researcher. We therefore argue that the level of detail in the DMP itself will and should vary significantly between research projects. Automated input of general information could be useful, this type of information should however be kept to a minimum. Further we do not find that automated decision-making and harvesting standardized output is beneficial for the researchers in their planning of data management.

Our suggestion is therefore to focus on balancing the guidance and decision-making, leaving flexibility for the researcher in the creating the DMP.

Conclusion

Creating consensus between data management stakeholders might not be necessary for cooperation or successful conduct of data management planning. With respect to research question 1, we identified four different perspectives amongst the participants. The different stakeholders have different perspectives each of which reflects to some extent their backgrounds and roles. The perspectives need to be considered if the DMP is to work as a document translating between different stakeholders and supporting the longevity of research data. Despite conflicting approaches to how the DMP should be developed, the stakeholders agreed on a common goal of creating the DMP and that the DMP has a purpose for several stakeholders, including themselves. Considering research question 2, our findings suggest that conflicting perspectives currently result in researchers becoming more distanced from the DMP, and that DMPs risk becoming merely a bureaucratic exercise. If leaving the shaping of the plan more open to the researchers to adapt to their needs, it can become useful in helping researchers plan for data sharing. The DMP should allow researchers to scribble down what is most relevant in each unique research project. The lack of coherence and the complexity of DMPs could be turned into a strength. If the DMP is to function as a standardised form facilitating co-ordination and collaboration between different groups of people, the degrees of standardisation and automation must be balanced, leaving the researchers with flexibility in the development and implementation of the plan. Only then can the DMP function as a boundary object translating between worlds. By formalising procedures and standardising methods, the DMP can become a boundary object, enhancing reproducibility and enabling data sharing.

References

[journal article] Berman, E. (2017). An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Integrated Findings to Develop Research Data Services. Journal of eScience Librarianship, 6(1), e1104. https://doi.org/10.7191/jeslib.2017.1104

[book] Borgman, C. L. (2015). Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, MA: MIT Press.

[book] Bowker, G. C. (2005). Memory Practices in the Sciences. Cambridge, MA: MIT Press.

[journal article] Burnette, M., Williams, S., & Imker, H. (2016). From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project. Journal of eScience Librarianship, 5(1), e1101. https://doi.org/10.7191/jeslib.2016.1101

[book] Creswell, J. W., & Plano Clark, V. L. (2018). Designing and Conducting Mixed Methods Research (3rd ed.). SAGE Publishing.

[journal article] Diekema, A. R., Wesolek, A., & Walters, C. D. (2014). The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories. Journal of Academic Librarianship, 40(3–4), 322–331. https://doi.org/10.1016/j.acalib.2014.04.010

[journal article] Dressel, W. (2017). Research Data Management Instruction for Digital Humanities. Journal of eScience Librarianship, 6(2), e1115. https://doi.org/10.7191/jeslib.2017.1115 [guidelines] European Commission. (2013). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[guidelines] European Commission. (2016). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. European Commission.

[guidelines] European Commission. (n.d.). Data management—H2020 Online Manual. Retrieved November 1, 2019, from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[guidelines] European Research Council. (2017, April 12). European Research Council Data Management Plan template. Retrieved November 19, 2019, from ERC: European Research Council website: https://erc.europa.eu/content/erc-data-management-plan-template

[guidelines] European Research Council. (2017, April 21). Guidelines on Implementation of Open Access to Scientific Publications and Research Data—In projects supported by the European Research Council under Horizon 2020. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

[journal article] Hardy, L. J., Hughes, A., Hulen, E., & Schwartz, A. L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. Journal of Empirical Research on Human Research Ethics, 11(2), 191–198. https://doi.org/10.1177/1556264616636233

[journal article] Johnson, A., & Knuth, S. (2016). Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach. Journal of EScience Librarianship, 5(1), e1089. https://doi.org/10.7191/jeslib.2016.1089

[report] Kunnskapsdepartementet. (2017). Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata (p. 48). Oslo: Kunnskapsdepartementet.

[research data] Kvale, L. (2020). Data from a three-phase Delphi study was used to investigate an emerging community for research data management in Norway: KIRDN_Data. https://doi.org/10.5281/zenodo.3673053

[journal article] Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. PLOS Computational Biology, 11(10). https://doi.org/10.1371/journal.pcbi.1004525

[journal article] Mischo, W., Schlembach, M., & O'Donnell, M. (2014). An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals. Journal of EScience Librarianship, 3(1). https://doi.org/10.7191/jeslib.2014.1060

[editorial] Nature. (2018). Everyone needs a data-management plan. Nature, 555, 286. https://doi.org/doi: 10.1038/d41586-018-03065-z [guidelines] NTNU. (n.d.). Open Data. Retrieved September 24, 2019, from https://innsida.ntnu.no/wiki/-/wiki/Norsk/NTNU+Open+Data

[book] Saldaña, J. (2016). The coding manual for qualitative researchers (3rd ed.). London: SAGE Publishing.

[presentation] Schjølberg, R. (2015, May). Research Council of Norway's Policy for Open Access to Research Data, Data management plans. Presented at the NorStore Research Data Management training seminar, Trondheim, Norway. Retrieved from https://www.sigma2.no/sites/default/files/media/norstore_rune_nfr_-_data_retningslinjer260515.pdf

[journal article] Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. Science, Technology, & Human Values, 35(5), 601–617.

[journal article] Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science, 19(3), 387–420. https://doi.org/10.1177/030631289019003001

[journal article] Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements. Journal of EScience Librarianship, 1(2). https://doi.org/10.7191/jeslib.2012.1008

[guidelines] The Artic University of Norway. (2019). Retningslinjer for oppfølging av datahåndteringsplaner ved UiTs enheter. Retrieved from https://uit.no/Content/633034/cache=20191706140204/Retningslinjer%20oppfolging%20DMP_publisert.PDF

[guidelines] The Research Council of Norway. (2014). Klimaforsk NOK 40 million for climate research projects within an open thematic framework. Retrieved July 8, 2018 from https://www.forskningsradet.no/en/Funding/KLIMAFORSK/1254017839874/p1184150364108?visAktive=true&progId=1253982678608

[journal article] Thoegersen, J. (2015). Examination of Federal Data Management Plan Guidelines. Journal of EScience Librarianship, 4(1). https://doi.org/10.7191/jeslib.2015.1072

[report] University of Oslo. (2019). Datahåndteringsplan ved UiO—Kompetanse, maler og organisering (p. 11). Retrieved from University of Oslo website: https://www.uio.no/for-ansatte/arbeidsstotte/fa/forskningsdata/291119_dhp_rapport_endelig-nettversjon.pdf

[journal article] Van Loon, J. E., Akers, K. G., Hudson, C., & Sarkozy, A. (2017). Quality evaluation of data management plans at a research university. IFLA Journal, 43(1), 98–104. https://doi.org/10.1177/0340035216682041

[journal article] Westra, B. (2017). Assessing data management practices and needs through data management plans. Abstracts of Papers of the American Chemical Society, 253.

[journal article] Williams, M., Bagwell, J., & Zozus, M. N. (2017). Data management plans: The missing perspective. Journal of Biomedical Informatics, 71, 130–142. https://doi.org/10.1016/j.jbi.2017.05.004

[journal article] Wright, A. (2016). Electronic Resources for Developing Data Management Skills and Data Management Plans. Journal of Electronic Resources in Medical Libraries, 13(1), 43–48. https://doi.org/10.1080/15424065.2016.1146640

[book] Ziglio, E. (1996). The Delphi Method and its Contribution to Decision-Making. In M. Adler & E. Ziglio (Eds.), Gazing into the Oracle—The Delphi Method and its Application to Social Policy and Public Health (pp. 3–33). London: Jessica Kingsley

Article 3

Kvale, L.H., & Darch, P. (2022). Privacy protection throughout the research data life cycle. Information Research, 27(3), paper 938. Retrieved from http://InformationR.net/ir/27-3/paper938.html
DOI: **https://doi.org/10.47989/irpaper938**

Privacy protection throughout the research data life cycle

Live Håndlykken Kvale and Peter Darch

Abstract

Introduction*. The sharing and reuse of research data is gradually becoming best practice in research. However, multiple frictions exist between realising stakeholders' ambitions for research and research data sharing and addressing legal, social and cultural imperatives for protecting data subjects' privacy. Through identifying and addressing conflicts between personal privacy and research, our paper offers advice to research data management services on how to approach personal privacy in research data sharing using the research data life cycle as the context.*
Method*. A three-phase Delphi study on a population comprising 24 stakeholders involved in research data curation in Norway. Data were collected during 3 consecutive rounds over 14 months.*
Analysis. *The data were analysed qualitatively. Following the third round of data collection, the entire corpus of data was analysed using exploratory sequential design methods.*
Conclusion. *The findings show multiple tensions between maintaining research subjects' right to privacy and advancing research through data sharing. This paper identifies and analyses three particular sources of tension: 1) maintaining trust with the research participants, 2) managing divergent views of privacy in international and intercultural research collaborations and 3) interpreting and applying policy. The divergent motivations and perspectives on privacy held by different stakeholders complicate these tensions. Researchers, research data management support staff and data organisations must reconcile these motivations and resolve tensions throughout the data life cycle, from collection to archiving and eventual sharing. Through dialogue and negotiation, all stakeholders involved in data sharing should aim to respect the research subjects' own understandings of privacy.*

Introduction

Policymakers and funding agencies increasingly require researchers to share research data openly (European Research Council, 2017; cOAlition S, 2019; National Science Foundation, 2011). Sharing *human subjects' data* (identifiable data from living persons) across national boundaries promises enormous benefits—for instance, in addressing global health emergencies, such as COVID-19, or in facilitating new research in social science (Havemann and Bezuidenhout, in press; Research Data Alliance, 2020; Kim, 2015; Lee and Jeng, 2019).

The open sharing of such data may pose considerable privacy risks to human subjects (GDPR, 2016; Nissenbaum, 2010, p. 4). Nevertheless, funding agencies often leave it to researchers and research support services to make difficult decisions about whether human subjects' data can be shared (European Commission, 2016; Research Data Alliance, 2020). Researchers struggle to access guidance in making these decisions (Jorge and Albagli, 2020; Modjarrad, et al., 2016; Research Data Alliance, 2020). University libraries' research data services (RDSs), which support researchers in planning, collecting and storing data, are potentially suitable entities that can provide such guidance (Pinfield, et al., 2014; Tenopir, et al., 2017).

However, to date, library and information science research in scholarly data sharing has largely focused on non-human subjects' data (Borgman, 2015; Darch, et al., 2015; Palmer and Cragin, 2008; Scroggins, et al., 2019; Tenopir, et al., 2017; Yoon and Schultz, 2017), leaving open the question of how to better configure RDSs in supporting researchers in balancing privacy concerns with the requirements and benefits of sharing human subjects' data.

Because multiple stakeholders with divergent perspectives are involved in RDSs, we investigate how perspectives on privacy influence research data sharing in practice. By identifying the conditions under which friction between privacy and research becomes visible, we provide advice for research data management services on how these can play a role in translating the needs of research versus privacy throughout the research data life cycle in a specific context.

Research questions:

1) What perspectives on privacy are held by stakeholders in the curation of research data on human subjects?
    a. How do these perspectives differ by role?
    b. What factors shape these perspectives?
2) How do stakeholders' perspectives on privacy shape their data curation actions?
    a. How do differences in perspectives between stakeholders cause friction during data curation?
    b. How are differences in perspective between stakeholders contested, negotiated and resolved?

Background

Versions of the research data life cycle are widely used within research data management to emphasise how a single dataset can pass through multiple contexts and be handled by different people and institutions. Challenges regarding sharing of human subjects data, including interview data or images, complicates this picture further, they represent a pressing issue. The cultural, legal and social contexts in understanding personal privacy are briefly described to illustrate how privacy should not be simplified to the current national privacy legislation implemented at the university level. Human subjects and the context in which they find themselves must be included when researchers are asked to share research data 'as open as possible and as closed as necessary'. Raising awareness regarding personal privacy amongst RDSs is necessary to ensure that the protection of privacy is maintained throughout the research data life cycle.

Current state of research data management

Research data life cycle models include various stages of processing datasets. One such model, derived from a synthesis of multiple models representing a range of disciplines, is presented in Figure 1 (Corti, 2014). A single dataset can pass through multiple institutional, organisational and cultural contexts during the life cycle. For instance, a researcher may collect a dataset in a remote field site in one country, take this dataset back to their home university in another country for analysis and then hand off the dataset to a data repository hosted by another university for long-term curation. In each context, the dataset may be subject to different regulations, policies, cultural perspectives and practices relating to privacy.

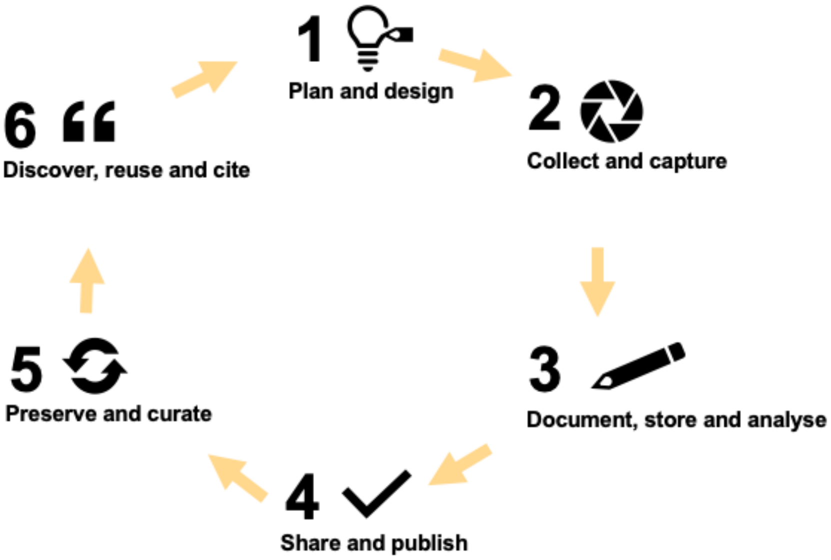


Figure 1: The research data life cycle (Corti, 2014, p. 17)

The *findable, accessible, interoperable* and *reusable* (*FAIR*) principles enshrine how research data should be made available for further research (Wilkinson, et al., 2016). The *collective benefit, authority control, responsibility* and *ethics* (*CARE*) principles are a supplement to FAIR and address human subjects' data (The Global Indigenous Data Alliance, 2019). Focused on data collected from Indigenous populations, the CARE principles emphasise protecting the privacy and dignity of research subjects (Caroll, et al., 2020).

University libraries increasingly offer RDSs that support planning, collecting and storing data (Kvale, 2021a; Tenopir, et al., 2013, 2019). Such services can include training for researchers in research data management, consultative RDSs and policy development—frequently in collaboration with the IT Centre and Office of Research (Tenopir, et al., 2017). The task of RDSs in planning the sharing of human subjects' data for further research requires that library staff acquire a deeper understanding not only of the law but also of research subjects' perspectives on what personal privacy means and the challenges researchers face when conducting human subjects research (Hardy, et al., 2016; Jackson, 2018). Institutions failing to protect personal privacy risk losing public trust (Guillemin, et al., 2018; McDonald, et al., 2008), and while privacy protection adds a layer of complexity to research data management, it can also be viewed as an opportunity to increase awareness regarding privacy and information security (Borgman, 2018).

Privacy and the challenges of human subjects' data

Research and research data sharing have become increasingly global, whereas understandings of privacy in Library and information science scholarship and practice on data sharing often remain linked to specific cultures and contexts (Jackson, 2018). To our knowledge, the alignment of the requirements of different research partners in different contexts has not been addressed in the literature. This section addresses the concept of privacy, relationships between privacy, context and culture, and how these relationships relate to collecting and sharing research data.

The meaning of privacy changes over time and can vary according to culture and context (Elias, 2014; Solove, 2002). In this paper, we define *personal privacy* in research data management as the power and right of research subjects to control their personal information or data (Floridi, 2013; Solove, 2010). The *fair information practice principles* (*FIPPs*) are rules for protecting privacy in record-keeping systems. The FIPPs approach privacy as providing control of personal information to the information subject (Zureik, et al., 2006) by regulating who can access personal information and for what purposes (Floridi, 2013; Inness, 1992). The FIPPs emphasise that information subjects should be able to find out what information about themselves an organisation stores and how the organisation uses this information. The FIPPs also state that personal information collected for one purpose cannot be used for a different purpose without the consent of the information subject (HEW Advisory Committee on Automated Data Systems, 1973). These perspectives are enshrined in principles governing human subjects research, as described in the Belmont and Menlo principles, the General Data Protection Regulation (GDPR) and the OECD Guidelines on the Protection of Privacy and Transborder Flows of Human Subjects Data (GDPR, 2016; OECD, 1980, 2013; U.S. Department of Health, Education, and Welfare, 1979; U.S. Department of Homeland Security, 2012).

In practice, however, full compliance with FIPPs and associated regulations is virtually impossible given the vast quantities and types of human subjects' data. The task of managing all the data that exist about them is overwhelming for an individual. The administrative burden of compliance on data-holding organisations is also immense. Instead, Nissenbaum introduced context as an approach to understanding privacy, taking account of the 'roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)' where information sharing is taking place to establish appropriate privacy-protecting practices (2010, p. 132). The context in which data were collected includes the researcher's original purpose for data collection and the data subject's culturally shaped motivations for allowing their data to be collected, understandings of what the data will be used for and perspectives on privacy. Nissenbaum's focus is on whether transfers of data from one context to another preserve the original *contextual integrity* of the data or whether they violate the expectations or goals of the data subject about the purposes for which the data will be used or their understanding of how their privacy may be at risk and may be protected.

Maintaining contextual integrity can be particularly challenging when a dataset is transferred across cultural boundaries, especially to a cultural context in which very different understandings of privacy apply. Several cross-cultural studies of privacy use Hofstede's indices for evaluating cultures (Bellman, et al., 2004; Zureik and Stalker, 2010), particularly the Individualism index, which differentiates individualistic societies, such as the US, from collective societies, such as Bangladesh, while Japan, France and Norway are in the middle (Hofstede, et al., 2010). The Globalization of Personal Data project found that members of individualistic societies were more likely to prioritise the protection of personal privacy ahead of other values, such as promoting public health, than members of collective societies (Zureik and Stalker, 2010).

Privacy and data sharing in practice

Laws regulating privacy help direct whether and under what conditions research data from human subjects can be archived and reused. Conversely, cultural understandings of privacy are often embedded in privacy laws (Nissenbaum, 2010). Approaches to privacy vary between Europe, where the law places responsibility on the government to act, and other countries, such as the US, where businesses are responsible for privacy protection (Lane, et al., 2014; Zureik, et al., 2006).

European approaches were embedded in the GDPR, which harmonised privacy law across the European Single Market (GDPR, 2016). The GDPR allows the collection of human subjects' data for research 'insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' (GDPR, 2016). The GDPR does not allow the open sharing or publishing of data without either anonymisation or the informed consent of the data subject. However, these measures do not guarantee privacy, as anonymised data are liable to be re-identified (Barocas and Nissenbaum, 2014, p. 50), and processes for gathering human subjects' consent typically occur at the start of the data collection process, often long before their sharing is envisaged.

Frictions between privacy theory and data management practice

Multiple sources of friction between stakeholders complicate privacy management in data curation. For example, in interactions between individuals, conflict may occur when different stakeholders involved in various stages of the data life cycle hold divergent values that influence how they approach privacy and data management (Bowker, 2005). Library professionals are typically trained and socialised to value open research, including open data sharing (Carroll, et al., 2020; Melinder and Milde, 2016). However, open data sharing is often incompatible with privacy protection and anonymisation requirements, meaning that researchers—who must protect their research subjects' privacy—may find themselves at odds with the policy of funding bodies (de Koning, et al., 2019).

Other sources of friction arise when researchers operate across countries and cultures and are subject to divergent national legislation and/or cultural norms. Research is increasingly being conducted in online environments, in which the sharing of human subjects' data can readily occur across legal and cultural differences (Ess and Hård af Segerstad, 2020). Researchers working in international environments may also face the challenge of complying with multiple sets of potentially incompatible funding agency requirements. Research data sharing opens up new challenges for cross-cultural ethics (Rappert and Bezuidenhout, 2016). The attention given to international ethical guidelines, such as the CARE principles, illustrates the need to look closer at the practices of sharing human subjects' data.

Methods

To address the research questions, the first author of this paper conducted a Delphi study to observe how stakeholders involved in research data management approach research data sharing and associated privacy issues, the conflicts they encounter and the compromises they make to enable data sharing. Delphi approaches are characterised by using an expert group of research participants and collecting data in multiple rounds (Ziglio, 1996). This method offers a way of systematically collecting solution-oriented opinions on a subject or problem. A Delphi study typically contains three phases (Figure 2). In each phase, data are collected and analysed, and the intermediate results are used in the development of the next data collection phase. The data collection process focuses on gathering participants' perspectives, assessing the extent to which these perspectives agree and eliciting from participants potential solutions to the issues raised. The multi-phase nature of Delphi studies enables participants to reflect on and respond to the experiences and perspectives of other respondents, including those working in roles and institutional contexts different from their own (Tapio, et al., 2011). Unlike focus group interviews, Delphi studies afford confidentiality to individual research participants and provide them with equal possibilities to express themselves (Landeta, et al., 2011). The multiple phases of data collection also enabled the first author to observe the developments that occurred over time.

Research participants

The study participants comprised researchers and staff involved in developing policies, building and operating infrastructure and providing support for research data management.

The participants (n = 24) were recruited from the Universities of Bergen, Oslo, Trondheim and Tromsø, all major Norwegian research universities, and from national providers of policy or infrastructure in Norway (Table 1). The research support staff covered a wide range of university-based services involved in research data management. The researchers—representing the largest group in the study—were principal investigators on projects receiving grants from the European Union in 2017 (European Commission, 2020). The researchers came from different disciplinary backgrounds (humanities, sciences and social sciences), with five using data on human subjects in their research. Two had extensive experience with national research ethics review boards.

Stakeholder group	Number of participants	Participant code			
R Researchers	8	R1	R2	R3	R4
		R5	R6	R7	R8
PO Policymakers	3	PO1	PO2	PO3	
IN Infrastructure providers	3	IN1	IN2	IN3	
IT IT research support	3	IT1	IT2	IT3	
RO Research support, research office	3	RO1	RO2	RO3	
L Research support, library	4	L1	L2	L3	L4
Total	**24**				

Table 1: Research participants

Research phases

A Delphi study comprises three phases (Ziglio, 1996). In each phase, data were collected and analysed, and the intermediate results were used in the development of the next phase (Figure 2). Inspired by a multiphase-design mixed-methods study (Creswell and Plano Clark, 2018), the first and third phases involved interviews and the second phase comprised a questionnaire. This approach provided both quantitative data, which enabled comparisons between stakeholder groups, and rich qualitative data, in which the participants elaborated on issues relevant to their perspectives.

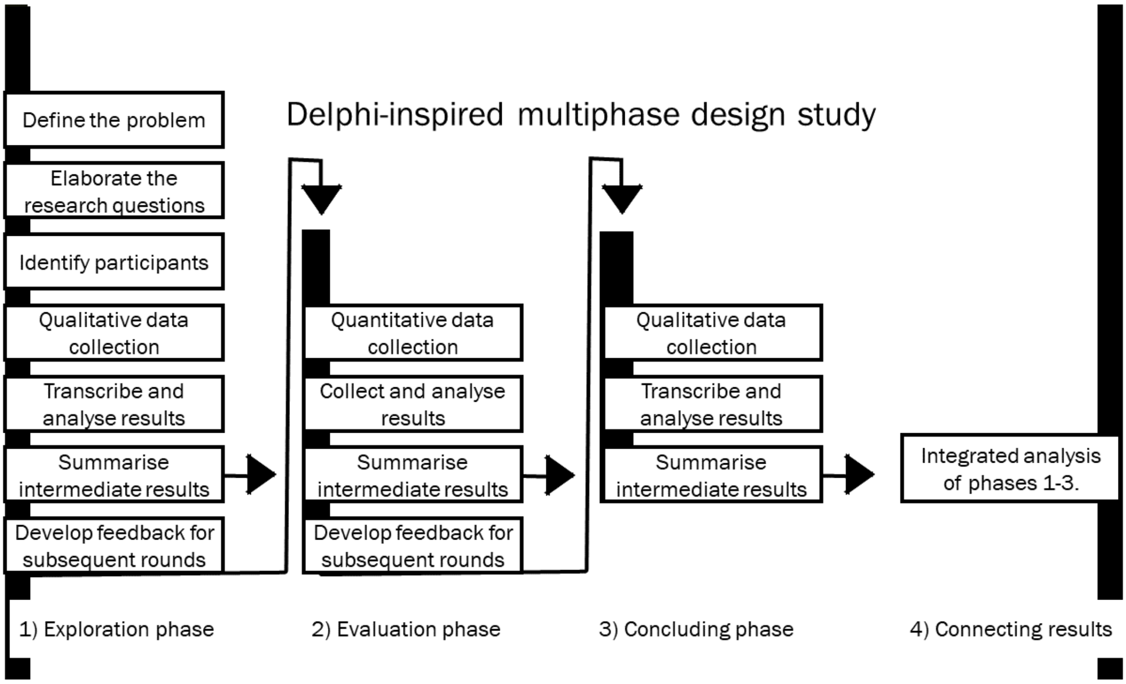


Figure 2: The research design (Kvale and Pharo, 2021)

In the first phase, conducted at the beginning of 2018, the interviews were approximately one hour long, yielding a total of 24 hours of recordings. The interviews used open-ended, exploratory questions. The participants were asked about how they worked with research data, what challenges they encountered and how they imagined an ideal solution to these challenges. Table 3 presents some quotes from the interviews that exemplify the issues and perspectives raised by respondents from each type of role. The first author transcribed the interviews, yielding 215 pages of transcripts, and developed a preliminary inductive codebook based on the topics and themes explored in the interviews. The codes and keywords were noted during transcription and then structured according to themes in a preliminary codebook (Saldaña, 2016). This codebook was then used to code the transcripts using nVivo. The results of this analysis informed the preparation of the questionnaire in the second phase and the integrated analysis of all data after the completion of all three phases.

In the second phase, conducted in September 2018, the participants answered a questionnaire in which they were asked to share opinions about 10 statements regarding privacy that originated from the first round of interviews (see Appendix). The participants were asked to state their level of agreement with each statement on a Likert scale. The results of this questionnaire were used to develop the interview protocol for the final round of interviews (Figure 2).

The third phase, conducted in April 2019, involved 30-minute interviews with each participant, yielding a total of 12 hours of recordings and 98 pages of transcripts, which included questions on personal privacy and public trust in research in the context of data sharing. The questions aimed to better understand how each participant encountered and dealt with conflicting demands regarding data sharing and privacy. The respondents were also asked about issues they had brought up in their previous interview. The preliminary codebook was developed into a final codebook, grouped according to the themes explored in the final interview, with qualifiers describing whether it was experiences or reflections that were shared and code terms related to the subject (Table 2). A Python script was used to extract the coded text, with 540 occurrences of the code ‘personal privacy’ and 245 occurrences of ‘practical experience’.

Finally, data from all three rounds of the study were analysed thematically using the themes and codes of the final codebook (Table 2). This article presents findings from themes relating to privacy and ethics, illustrated with quotes. Most of the quotes used were translated from Norwegian for the purpose of this article, while three participants were interviewed in English. Parts of the material presented in this article have previously been presented in poster format (Kvale and Darch, 2020).

Code	Description	Application
Reflections	Sharing of thoughts or reflections on the subject	These two codes were used as qualifiers to sort quotes in which the participants were referring to practical experiences or reflections on the issue.
Practical experience	Referring to own experience on the subject	
Consent	Thought or experiences regarding the use of consent	‘Much research is conducted on data collected by governmental agencies in one way or another; much of this is data in registries. Mainly, I believe that privacy protection is important, and that embedded privacy is crucial. I do not believe we manage to collect the benefits of the data if we don’t find a good solution for sensitive data’. (PO3)
Embedded privacy	Thought or experiences regarding the use of embedded privacy in privacy protection	
Personal privacy	Thought or experiences regarding aspects of privacy protection	
Public trust	Thought or experiences regarding public trust in research	‘I do not believe it is possible to conduct research without trust in research […] If research is to be publicly funded, it must be trusted. It is as simple
Cost profit	Thought or experiences regarding the	

	cost and profit aspect of data archiving	as that; it takes so little to destroy that trust, and by that, remove the will to fund'. (IN3)
Integrity	Thought or experiences regarding research integrity	'Personally, I would always argue for increasing quality assurance in research. Quality is what research is: to deliver knowledge which is relevant for those who potentially are interested in learning or applying. But it needs not only to be relevant but also to be solid. So, quality for me is above all else in research'. (R2)
Research essere	Thought or experiences regarding the ethos of research, what research are or strive to be	
Research ethics	Thought or experiences regarding research ethics	
Internationalisation	Thought or experiences regarding internationalisation in research and data sharing	'The idea of GDPR was to have free exchange of data and research collaborations across national boundaries—something which becomes extremely difficult when GDPR is practiced so inconsequently in the different countries'. (L3)
Privacy vs. research	Thought or experiences regarding the balancing of the respect for privacy with conducting high quality research	'Regarding privacy protection, I believe the commercial interests are much more dangerous than the researchers. I would say that it should be much freer for research and stricter for commercial use'. (R1)

Table 2: Qualitative codes within the ethics theme and examples of quotes coded with the different codes

How they work with data	Challenges they face	Ideal solution
'For us, research data means how to integrate data from all these sites, how to harmonise, standardise and integrate them and then how to analyse them in a way that something new comes out of that'. (R4)	'We had a data request and sent the data we used here, which are the translated transcripts. However, we explained that we did not consider it relevant to bring the original language audio here. But the question is if we should use the original audio? If these should be stored here? And there are hundreds of these. But it is not clear if it is us or our sub-department, the project on site, who are responsible. (R2)	'Access from anywhere, without requiring, for example, an institute affiliation […] To be able to use the data without downloading, to be able to read and understand the data from others, like properly defined metadata… What else? And find who created the data'. (R8)
'We receive and disseminate data for research purposes primarily, but also for educational purposes and, occasionally, for commercial purposes. But research is our primary focus. We receive data from researchers, but also from the National Bureau of Statistics; much of our data come from the Bureau of Statistics, where we accommodate and disseminate for research free of charge. We also have an agreement with the National Archives for the archiving of research data'. (IN2)	'It is more difficult to combine these requirements [of policymakers] technically. We have the natural attitude of the researcher of keeping safe their own discovery and their own data, so we need to provide a platform, a technical platform that once it is seen by the researcher as an advantage—not something which is just, "I must use this because I have been told to use this tool"… They must clearly see the advantage in using some tool'. (IN1)	'The technology is in place; this is not a technological challenge. The challenges are culture and organisations, and it is completely feasible to do something about this if you have vitality and time, because it will take time to change work routines, and when these are changed, you might be able to change the culture within the organisations, and this is something policy-makers clearly want'. (IN3)
'The plan is that I shall be one of the driving forces behind this from the side of the library, preparing the whole organisation for research data sharing'. (L4)	'We often talk about research data, and do things form a Norwegian perspective. While most researchers have an international perspective, this can sometimes conflict with the library perspective. The research disciplines operate in an international context, and the libraries are used to operating institutionally. The national dialogue again, tends to consider Norway as distinct from the rest of the world'. (L3)	'Collaborations between those providing retrieval services—those who build an archive and implement metadata standards—and research communities. Collaboration is key'. (L1)
'I prepare the institution for the	'This is fairly new at the university, and	'Quality assurance must be a

storage or archiving of research data so they can be made openly available, partly open or not open but can be retrieved and the research reproducible'. (RO1)	the challenges are big and small. Just opening the box of everything regarding research data, it surprises you: "Wow, did we really have this little overview?". Then how and in what order do we approach this? To build one service and infrastructure with the technical, the knowledge and the consciousness'. (RO2)	requirement, which can be discussed, but there should be a certain quality requirement. And then it is payment: open data implies free of cost, but should there be a cost for archiving?' (RO3)
'I have been working much on the national goals and guidelines for open access […] and now the national strategy for access and sharing of research data. So what I will be doing in the time to come is to ensure that the strategies and guidelines are implemented'. (PO3)	'Partially, it is to create a culture of data sharing, as this is not yet common practice in all disciplines, at least not in the open. People probably store data, but to what extend the storage is open varies. Also, I think we have a job to do in standardising to meet the FAIR principles'. (PO2)	'A bit like EOSC [European Open Science Cloud], one entry point, less choices and more streamlining, less work for the researchers—of course, they must describe their data and those things, but a service level that took care of the rest, including curation, access, long term archiving, retrieval and did this FAIR'. (PO1)
'How I work with data depends on which role I have, as I used to be a researcher, then I began as an IT-architect ten years ago and was looking into the lack of infrastructure for data management in research. So, I wrote a memo about the need to establish an infrastructure for open data'. (IT2)	'Sometimes, the demands for accessing data are challenging due to either size, speed, or it is sharing across nations or technologies. But the largest challenge is to keep the focus on open science and FAIR; the funders are saying that if you are not FAIR and open science in your data management, you will not receive funding. Still, the infrastructure is not in place because everyone likes making policy without paying for implementation'. (IT3)	'We need to think of a virtual data catalogue based on good disciplinary standards according to various attributes and ensuring that they are safe in terms of not being modified, being available and compatible over time with new standards'. (IT1)

Table 3: Descriptive results of the roles of the interviewees

Research ethics

Permission to collect non-sensitive personal data for the purpose of this study was granted by the Norwegian Centre for Research Data, Data Protection Services (study 56829 2017.11.22). To balance the privacy of the research participants with the authors' desire to make the research data open, the participants signed two consent forms: the first regarding participation in the study and the second regarding the publication of pseudonymised data in a repository (Kvale, 2021b). Full anonymisation of the data was not possible given that the context and details regarding the work of each participant allowed for identification by their colleagues. Prior to signing the second consent form, the respondents received a copy of their data to review. Six participants chose not to allow open sharing of all or parts of their data in a repository.

Findings

Realising the benefits of data sharing while protecting privacy is often a core ethical challenge of research data sharing, as reflected in the following quote from an interviewee: 'When it comes to the storage and management of data, I believe […] there is a fundamental conflict between different values: the need for high quality scholarship and personal privacy' (R6). Here, we present findings about how this conflict is negotiated by various stakeholders involved in data sharing. Three particular dimensions of this conflict emerged from our study and will be addressed here:

1) Maintaining trust with research participants;

2) Managing divergent views of privacy in international and intercultural research collaborations;
3) Interpreting and applying policy.

These themes highlight how privacy in data management is a complex subject, involving trust, cultural differences, personal relations and compliance with policies.

Maintaining trust with research participants

The various groups of stakeholders involved in our study largely agreed that privacy protection is important for maintaining public trust in research (Figure 3). However, this trust may be undermined when human subjects' data are transferred not only from research participants to researchers for the purpose of research but also to other stakeholders, including other researchers, data stewards or data organisations and back to the research participants. These transfers can lead to research participants losing a sense of control over their own data and may raise concerns about how these data may be used. This section identifies challenges researchers face as they resolve tensions between requirements to transfer data to others (e.g. for curation or for fulfilment of open data mandates) and the necessity to maintain their research participants' trust.

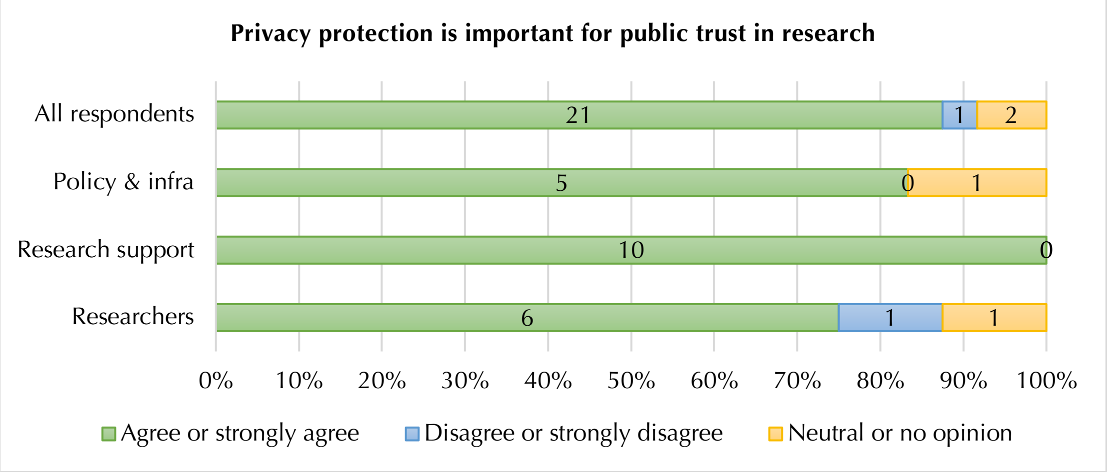


Figure 3: Privacy protection and public trust in research; responses from the questionnaire

Interpersonal trust between research participants and researchers

For researchers who worked on studies involving the long-term engagement of participants, even over decades, maintaining relations of trust between participants and researchers as data were shared with others beyond the initial study was critical for protecting this engagement: 'We always have to do everything to maintain the trust' (R4). For example, R4 was part of an international research team working at multiple sites across Europe and the US. In her project, data from previous research were added to a large database, which allowed partners to access these data and add new data from follow-up studies. The researchers collecting data perceived a limit to what they could ask of their participants. Exceeding this limit could have reduced a participant's trust: '[We] could do even more things, of course, but then you draw a line. I don't go further than this because it is not worth it. I might lose trust if I go further' (R4). R4 further described how participants trust researchers to protect their confidentiality and not be negligent with their data. The researcher explained that the participants with whom they were in contact displayed a high level of trust: 'The research participants here are really, really committed, so they really want to contribute, but I think they are not overly conscious about the privacy issues because they have a lot of trust in the research group' (R4). This trust was fostered by R4’s research team, who worked actively to update the participants on research progress and engage in dialogue with them.

Other researchers in our study echoed these sentiments. For instance, R2 expressed awareness of the fact that participating in a study and contributing data was a burden for research participants and that participant trust could only be maintained if these participants believed this burden was proportionate to the benefits of the study:

> *We are dependent on high-quality information from people. I believe that when you work with people and want them to contribute their data, you are also obliged to communicate that they benefit from the research being conducted and that the research somehow is relevant for them as well. For people to not be instrumentalised, we need a fundamental trust in research. (R2)*

Providing research participants with their own data

Once their data are collected, the human subjects often have no further involvement in the research process. However, according to the GDPR, they retain the right to access information about themselves unless it is deemed not to be in the best interests of their health (GDPR, 2016, art. 15; The Norwegian Personal Data Act, 2018, § 16. c). The divergence in the ways the participants’ interests are regarded by the different participants suggests a need for further knowledge regarding this aspect of privacy protection in research from a data-sharing perspective.

For instance, R4, who worked with health data, did not routinely share with participants their own data, even when the participants wanted to access them: ‘If they are interested in brain research, they are also naturally interested in their own brain data. Sometimes it is difficult to say, “Sorry, we cannot [provide the individual results]” (R4). R4 justified this reluctance by arguing that participants would not be able to interpret their data correctly, leading to potentially harmful outcomes. While sharing data with the participants could, by way of transparency, enhance participants’ trust in the researchers, R4 placed greater weight on protecting the participants’ health. The only exception R4 made was when the data revealed previously undiagnosed medical conditions, in which case, she has a moral duty to inform the research subject.

Multiple other stakeholders in our study also considered the dilemma of when to provide research participants with their own health data, reaching a range of divergent perspectives. For instance, policymaker PO2 took a more cautious approach than R4 about whether to inform a research participant about a potential medical condition:

> *When you know that someone has a mutation, making them exposed to diseases with large consequences, should one backtrack through the data and inform the participant? And I would say no, one should not do this unless permissions for such connections are explicitly granted. (PO2)*

Meanwhile, IN1 was far more sympathetic to the notion of sharing a research participant’s data with the participant: ‘Sharing with the owner, the data owner [data subject], is the key mechanism to gain trust’ (IN1).

The different conclusions reached by R4, PO2 and IN1 illustrate the lack of consistent perspectives across stakeholders, underlining the need for greater infrastructural support to minimise tensions between stakeholders as they navigate thorny ethical issues relating to human subject data sharing.

Managing privacy in international and intercultural collaborations

The interviewees handling personal data in international collaborations encountered conflicting cultural understandings of privacy within their collaborations. These differences created barriers to data sharing across collaborations.

Divergent understandings of what is considered sensitive data

Conflicting perspectives on privacy amongst different researchers can lead to tension and frustration within an international research team. For instance, L1, a librarian and data steward, worked as a researcher on a project involving multiple international partners. Differences emerged regarding which parts of the data were considered sensitive:

> *I was part of a data collection project in France, where we also had partners from Japan. And when the participants talked about what food they like [...] this was considered sensitive information by the Japanese researchers and could not be made available. (L1)*

While the Norwegian research team wanted access to data about research subjects' food preferences and did not see any ethical problems with accessing these data, the perspectives of the Japanese data collectors took precedence, frustrating researchers from other countries.

Other participants not only echoed how understandings of what is considered sensitive change over time and place but also explained how these understandings can vary within a single legal jurisdiction or local context. For instance, both IT3 and PO1 referenced how the implementations of the GDPR can vary considerably within Europe: 'The interpretation of the GDPR is very north/south; it is completely different in Spain than the Nordic countries' (IT3) and 'I have spoken with researchers [in other European countries] who can barely conduct their research if one is to follow the Norwegian implementation of the GDPR' (PO1).

Meanwhile. R3 found differences in understanding across multiple generations within the same family:

> *With the [grandmother], there is something strange regarding the father of her child, some vague formulation about a quick separation. Her child also does not say anything, apart from 'my father disappeared quickly'... However, when I interviewed the [grandchild] sometime later... then the story was revealed: the father was a German soldier. (R3)*

Privacy protection through local partners

When researchers collect human subjects' data in a context different from their own, they use strategies to understand and respect the participants within their own context. Partnerships and the empowerment of communities through citizen science or with researchers in local universities are strategies to ensure correct interpretations and translations of contextual differences.

By understanding privacy as a context-sensitive cultural phenomenon, researcher R2 and her group involved local partners and used applied ethics, defined as the interaction between ethical theory and moral practice, as an approach to protecting participants' privacy according to the participants' own preferences.

R2 discussed the ethical challenges she encountered when conducting interviews about how local communities adapt to climate change in rural Bangladesh. R2 described Bangladesh as a more collective society than many Western societies; in Bangladesh, the needs of the local community more frequently take precedence over those of individuals. Through dialogue with research partners from local universities and by using their local knowledge, R2 and her team conducted interviews on the street rather than in homes or other closed surroundings, which would have been the preferred context in Europe. This choice created some new challenges regarding who responded to the interview questions:

> *We realised that even if we had only one informant in a village, then […] at least 10–12 others around him added to his responses. He would pass the questions on, 'Oh God do I actually have some debt anywhere,' and the others would reply, 'Yes, you have, there and there,' which means sharing relatively sensitive information with others looks different in a Western context than in many other cultural contexts where you don't have the individual-based, but the group-based [society]. (R2)*

Although the economic situation of an individual is an example of information that, in some contexts, is regarded as sensitive information, in this case, it was not. Being a collective society also implies differences in what information is shared with whom; the private sphere includes the village rather than being limited to individuals or a nuclear family.

This example illustrates the need to understand privacy as a context-sensitive cultural phenomenon. R2's perspective on privacy as an individual right was challenged when conducting research in a different culture. R2 suggested that dialogue and interaction between different scholarly disciplines working with human subjects' data and different cultures are needed to properly reflect on how to protect privacy in research across cultures and contexts.

Another aspect of understanding the context in which one operates highlighted by R2 is the need for researchers to have an awareness of the power structures in which research participants are embedded.

These power structures can involve gender, education level and religion. Research participants' perspectives on privacy are also affected by how they experience themselves in relation to their surroundings, in the way that privacy is about subjects' control of personal information or data in a context. Without making an effort to understand this context, the researchers might fail to protect the participants' privacy. These examples illustrate the importance of understanding the contexts in which the research participants operate. For R2, the answer to how researchers should approach power structures is reflection, aimed at finding solutions and respecting and balancing the needs of the participants and of the research: 'There are different structures, and we need much more reflection' (R2).

R2 also described challenges regarding storing, depositing and deleting data from the project. The original interview recordings and transcripts were kept by collaborators in Bangladesh, while the Norwegian researchers used the translated transcripts. Requirements from the Data Protection Services at the Norwegian Centre for Research Data to delete the original recordings did not apply, as those recordings were kept in Bangladesh. However, the division of responsibility for data between the Norwegian and the Bangladeshi teams was not formalised. Retrospectively, the researcher questions whether it was right to split the material in this way, suggesting that better guidance for how to approach archiving in international research collaborations is needed.

The importance of an international approach is echoed by one of the librarians interviewed:

> *We often talk about research data, and do things form a Norwegian perspective. While most researchers have an international perspective, this can sometimes conflict*

> *with the library perspective. The research disciplines operate in an international context, and the libraries are used to operating institutionally. (L3)*

Interpreting and applying policy

Interpreting and applying personal privacy laws define the limits of the research project and the possibilities of sharing research subjects' data. Researchers perceived tensions between conducting research and protecting privacy (Figures 3, 4 and 5). While many researchers expressed a belief that they should have more discretion than they are currently allowed in determining the extent to which they trade protecting privacy for conducting important research, research support staff clearly disagreed. These disagreements contribute to tensions between different stakeholder groups in how privacy issues are handled in practice (Figure 4).

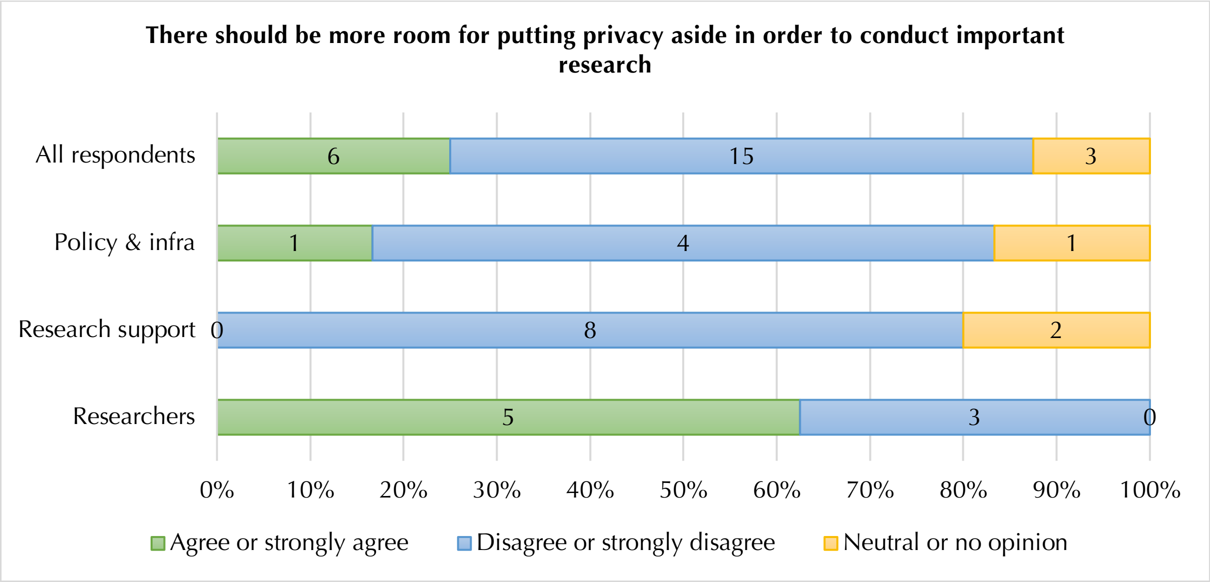


Figure 4: Putting privacy aside to conduct research

Researchers' dialogue with data protection services and ethical committees

The questionnaire showed that most researchers we studied thought that the providers of privacy protection services lacked an understanding of how research is conducted (Figure 5). Several of our interviewees expressed frustrations with the multitude of privacy protection offices with whom they must deal, including ethical committees, institutional privacy protection officers and the Data Protection Services from the Norwegian Centre for Research Data: 'I had a case where the regional ethical committees gave an o.k. for the research project, and then the local personal privacy officers at the hospital involved said, "No way"' (IT1) and 'The whole Norwegian Centre for Research Data system, to which I have had to relate […] they simply cannot understand qualitative data, they have no idea what qualitative data are' (R3).

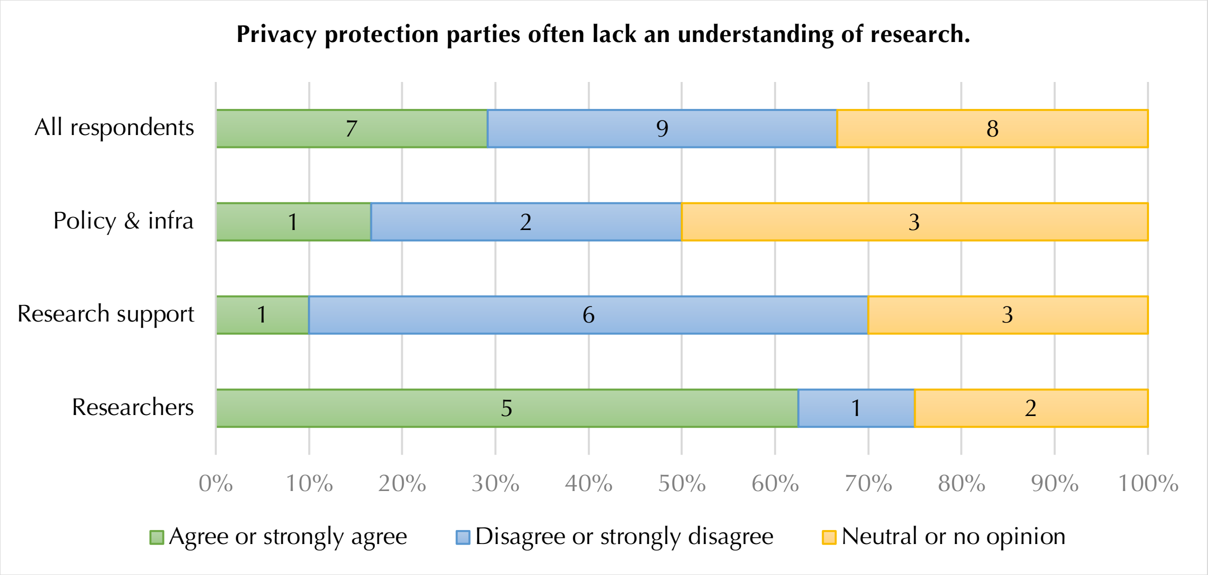


Figure 5: Stakeholders' views on privacy protection parties and their understanding of research

As a result, researchers often perceived that they had to make a choice between developing strategies to minimise disruption from their encounters with privacy protection services or suffering significant delays in their projects. For instance, R3, who had conducted a longitudinal study over more than a decade, explained how they received letters annually from the Data Protection Services from the Norwegian Centre for Research Data requesting the data to be deleted. To them, the frequency of these letters suggested that the privacy protection office lacked an understanding of longitudinal qualitative research: 'Every year, I received a letter asking me to delete the data […] and every year, I wrote back that this is a longitudinal study. I need to keep the data' (R3).

R3 discussed their dialogue with the National Data Protection Services regarding permissions for conducting interviews and collecting non-sensitive personal information: 'Suddenly, one person who understood qualitative research appeared, but otherwise, there were only zombies. Now, they have got other ones as well, thinking humans, not just sticklers for the rules' (R3). R3 explained that they had seen improvements over time regarding how the service dealt with qualitative data, but their many years of experience had left them with general mistrust in the service.

The Data Protection Services were familiar with this issue but claimed that the request to delete data did not come from them:

> *I have heard researchers multiple times claiming that The Norwegian Centre for Research Data told them to delete their data, and I have never said this to anyone. But still, this is the perception. We have a recurring communication challenge in making the individual [researcher] familiar with the legal system. (IN2)*

Although they have presumably changed their practice of requesting for data to be deleted, the mistrusts amongst researchers with such experiences remain.

Meanwhile, R7 highlighted the need for the help of data stewards or other research support staff when developing and submitting applications to the Data Protection Services, as a late response or rejection can result in substantial delays for a research project:

> *We have a project which is four months delayed only because the Data Protection Services doesn't manage to give us a go. If we only had some help with both*

> *designing the applications and sending reminders, when we would save so much time. (R7)*

Other stakeholder groups reported considerably more positive views of the Data Protection Services than the researchers. These divergent opinions suggest that research support staff should be careful when recommending these services, as they may encounter scepticism from researchers, leading to potential friction and mistrust between themselves and researchers.

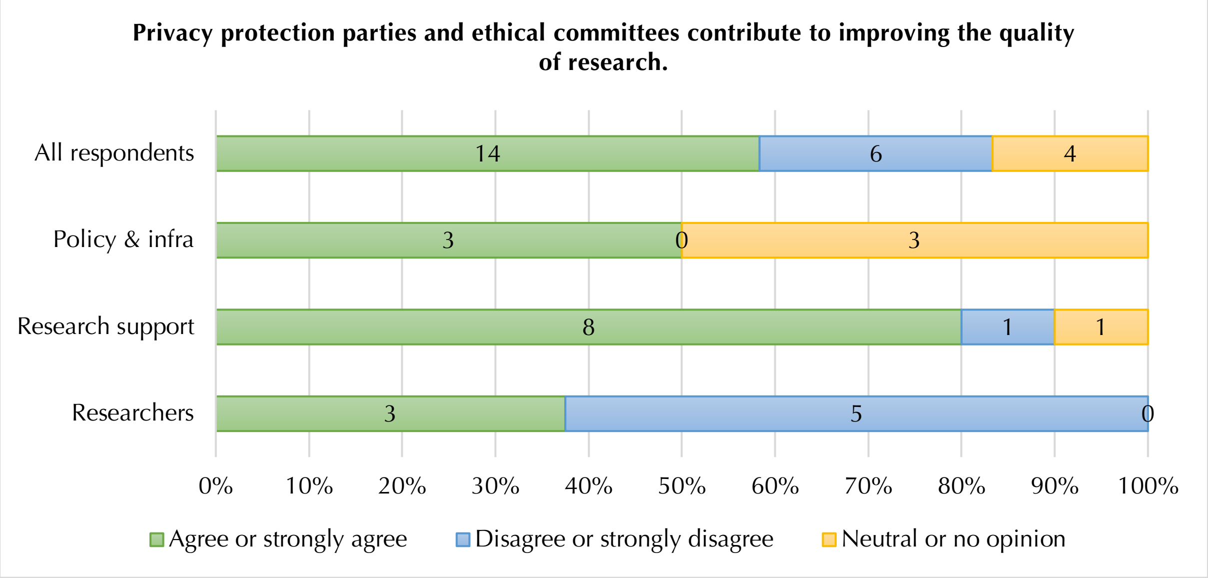


Figure 6: Stakeholders' views on privacy protection parties and their contribution to research quality

Similarly, researchers also disagreed with other stakeholders on the extent to which they perceived that the contributions from ethical committees and the Data Protection Services improved research quality (Figure 6), with research support service staff holding a significantly more positive opinion.

Awareness of how researchers perceive the Data Protection Services is important for library-based research support services to create trusting relationships with researchers. Appropriate guidance on designing research proposals that balance compliance with privacy regulations with the ability to conduct research using a range of methods and data sources should be made available to researchers early in the research life cycle. This knowledge and experience with what works are valuable to researchers in navigating tensions between complying with privacy law and conducting research.

Researchers' frustrations in complying with privacy law

Researchers' dialogue with legal advisors is central to developing projects that collect human subjects' data. The researchers we interviewed typically consulted legal advisors for advice on collecting, using and sharing data legally. However, challenges arose when researchers found this advice unreasonable.

For instance, R5 was frustrated by the limitations that informed consent requirements placed on her ability to share data openly. In her example, data were collected from filming musicians in her laboratory:

> *Then, when they enter our lab and we film them, they are happy about that, but still, we are not allowed to use that and give them visibility because the recordings are done within a research context. That, yes, is a bit strange. (R5)*

Legal restrictions meant that R5 was not allowed to share data collected in the laboratory (The Norwegian Personal Data Act, 2018, art. 6.4 and art. 5.1.b.), despite the data subjects' willingness for their data to be shared and publicly identified. To overcome this barrier, R5 now collects data by filming these musicians playing in concerts. The public nature of concerts allows for the data to be shared openly.

Another interviewee, R3, described how she chose not to comply with legal requirements. Upon completion of her research project, she was asked by the Data Protection Services from the Norwegian Centre for Research Data to either anonymise the project data, acquire new permissions from the research participants to retain their data or delete the data. R3 explained her perspective:

> *I would prefer not to delete this material because I am hoping to make a replica study and I was so busy at the time. So, I wrote back that the material had been deleted, which is not at all true. So sometimes the good intentions become its own enemy—when they demand something that is unrealistic, making us, as researchers, take shortcuts, hoping that no one will ever notice. (R3)*

R3 regarded complying with the Norwegian Centre for Research Data's requirements as infeasible. She could not contact the participants for informed consent, as she had already deleted their contact information and did not regard anonymisation as realistic. Meanwhile, deleting the data would have jeopardised her future research plans. Instead, R3 committed what she described as 'a small piece of civil disobedience' (i.e. breaching privacy law). When researchers falsely report having deleted data to their university, these data are instead hidden on a researcher's own computer or cloud storage account (e.g. Google Drive) rather than on secure media, such as university systems. This practice increases the risk of human subjects' data being accessed by hackers, potentially exposing data subjects to harm.

The burdensome and complex task of balancing research and privacy, as described by R5 and R3, was echoed by other researchers:

> *My experience is that most researchers experience this as burdensome tasks, "OMG, how do I go about this now?" and "What is the best thing to do here?" I think what we need are people providing guidance, assisting researchers in getting permissions and choosing responsible storage. (R2)*

Discussion

Providing research data management support is about facilitating the transition of data from one step of the research data life cycle to the next. Managing human subjects' data requires an additional layer of planning, including legal advice regarding personal privacy and applying for ethical approval. For researchers, our findings (see quotes from R2, R3, R5 and R6) demonstrate that personal privacy is often perceived as imposing burdensome, often complicated, requirements that may compromise researchers' ability to conduct innovative and high-quality research.

For university-library-based research data management services, delivering appropriate consultative support can include posing questions to researchers, being available for dialogue and initiating reflection on the part of the researchers rather than 'providing a choice between 'yes' or 'no' answers (Tenopir, et al., 2017). However, this approach requires that research data management support teams are familiar with the core principles of personal privacy ethics, privacy law and the researchers' own perspectives, knowledge and experience of handling human subjects' data. While the work of privacy protection officers involves ensuring that researchers follow the law, our findings suggest that the privacy evaluation needed in research is frequently more complex. Maintaining trusting relationships between stakeholders and working across national and cultural boundaries create ethical challenges regarding privacy that are not only about respecting the law but also about respecting the individuals who share their data with researchers (see quotes from R2 and L1) (Shankar, 1999). By applying a contextual approach to privacy protection (Nissenbaum, 2010), we argue that research data management services should encourage researchers to focus on context, transmission and actors when reflecting on how to protect the interest of data subjects.

Data subjects' trust in sharing their data

Managing human subjects' data requires awareness of the sender, recipient and subject (Nissenbaum, 2010). In research data curation, these placeholders are different actors located at different stages of the life cycle (Figure 7). In step 2, when data are collected from research participants, the sender and subject are the same. In steps 3 and 4, the role of the sender is held by the researcher. The repository or data organisation is the sender in steps 2 (when researchers are using data from archives or other data organisations), 5 and 6.

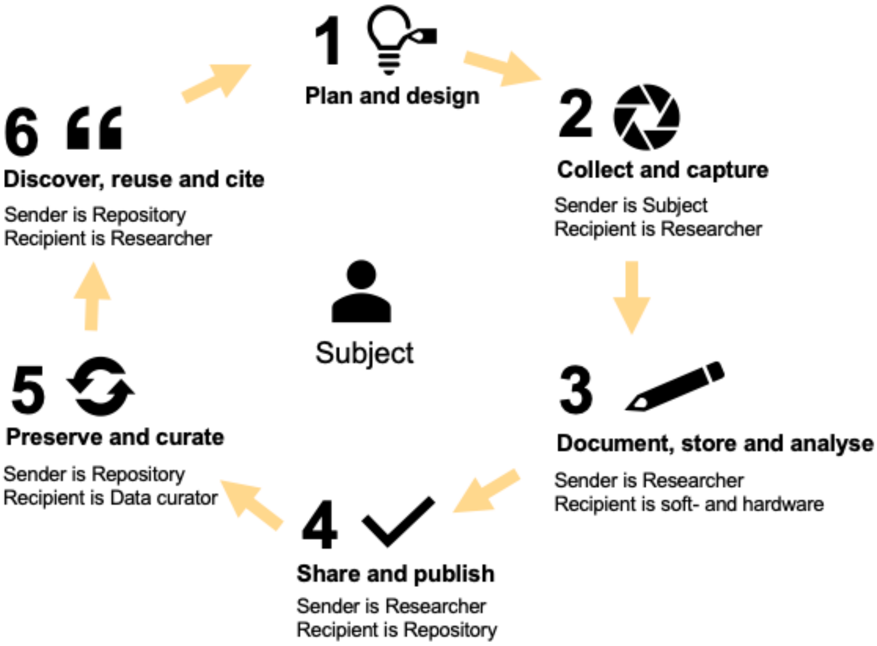


Figure 7: Personal privacy in the research data life cycle

Privacy protection is both a prerequisite and a condition for trust between the subject and other stakeholders involved as senders or receivers of human subjects' data (Floridi, 2005; Nissenbaum, 2010). Our findings show that successfully sharing human subjects' data between stakeholders requires researchers to build and maintain strong trusting relationships with research participants. These relationships, in turn, help researchers facilitate reuse and sharing for other research purposes. The participants' trust in the researcher as an individual and in the research and university as the context is crucial for data and research quality. Research participants often trust the university or research organisation rather than the individual researcher (Guillemin, et al., 2018). Research institutions represent the context in which participants trust their data to be processed according to explicit or implicit expectations.

The researcher maintains these trusting relationships by protecting the identity of the participants through not only anonymity but also what information is shared (Hardy, et al., 2016). Researchers should provide data subjects with knowledge of how privacy is protected throughout the life cycle and aim for shared stewardship and the empowerment of the data subject (Carroll, et al., 2020; First Archivist Circle, 2007; Shah, et al., 2021). This information should be given by the researchers both prior to data collection in stage 2 and during stage 4, when archiving or sharing data (Guillemin, et al., 2018). In stage 4, researchers should also provide the participants with information on where the data are archived and update participants on publications (Shah, et al., 2021).

When data are archived in a repository, responsibility for the data is transferred from the researcher to the data organisation, including responsibility for ensuring the compliance of access restrictions with privacy law (Eschenfelder and Shankar, 2020; Shankar, 1999). We suggest that the data subjects should be informed when data organisations take over this responsibility. In stage 4, the distance between the participants and their data increases, and an institutional trust relationship is required (Shah, et al., 2021). How digital solutions can minimise this distance and provide participants with increased control of the data regarding themselves should be further explored (Budin-Ljøsne, et al., 2017).

Privacy protection in international collaborations

Privacy protection in international research collaborations involving human subjects' data is complicated (Dilger, et al., 2019; Jorge and Albagli, 2020). Ethics oversight boards and their guidelines are often nation- or institution-specific, while researchers work globally. Initiatives to address personal privacy in a global research context would be valuable to highlight cultural differences in privacy and promote discussion of how to respect these differences (Carroll, et al., 2020; Melinder and Milde, 2016; Viberg Johansson, et al., 2021).

Transmission of data using fishing zones

Expectations regarding the transmission and sharing of human subjects' data are often tacit and commonly create misunderstandings between researchers and data subjects (Nissenbaum, 2010). Our examples demonstrate the complexity of transferring data between different contexts and how different understandings of privacy create obstacles. These misunderstandings could be mitigated if, before the start of the data collection process, researchers are explicit about how the data will circulate.

In international research, transferring data between jurisdictions might not always be necessary if appropriate storage and access are provided remotely. Options for researchers to work remotely in the jurisdiction of the data subjects could help in balancing conducting research with protecting privacy. Within Europe, the archiving of data where they are collected is referred to as 'the fishing zone agreement' (Eschenfelder and Shankar, 2020, p. 697).

However, the fishing zone approach is not always appropriate—for instance, in cases in which local laws do not provide data subjects with adequate privacy protection or for research on topics that are considered particularly sensitive in the local context. Researchers should always take care not to expose their participants to harm. When conducting research on exposed groups, dialogue with these groups and respect for their wishes might be the best protection. Ensuring that the research participants have the authority to control the data and the right to develop the cultural governance protocols highlighted in the CARE principles (Carroll, et al., 2020) is best achieved through local partnerships and dialogue between the researchers and the participants.

Creating common understandings of privacy in international research collaborations

When using human subjects' data, research should be grounded in an understanding of privacy that incorporates cultural sensitivity. Cultural understandings of privacy vary, particularly in relation to whether and how data can be shared. To respect the participants, researchers should reflect on any possible power structures and the cultural context of the participants and avoid enforcing their own understanding of privacy (Nissenbaum, 2010). Within archiving practice, the concept of shared stewardship is used to extend the notion of provenance for documents originating from Native Americans (First Archivist Circle, 2007). Shared stewardship requires the archivists to 'consult with the communities represented in order to understand how their cultural paradigms bear upon the materials in their custody' (Alcalá, et al., 2016, p. 332). Below, we suggest different strategies that researchers can use to reflect on power structures and the protection of privacy from the perspective of the subject:

- Actively drawing attention to tacit expectations regarding the collection and sharing of human subjects' data early on in an international collaboration to identify potentially conflicting views on privacy;
- Consulting surveys, such as the OECD Guidelines on Measuring Trust (OECD, 2017), or applying indices measuring cultural differences (Hofstede, 1984) to prepare for conducting cross-cultural research. As illustrated by R2's case of research in Bangladesh, the extent to which a society is collectivist or individualist may predict cultural attitudes towards privacy. The trust in government and public institutions numbers from the OECD complements the picture by indicating to what extent trust in universities as institutions can be expected from research subjects (OECD, 2017). Trust in the institution is central in participant recruitment and relevant for data quality (Guillemin, et al., 2018);
- Using a second translator to translate interview transcripts back to the original language for comparison against the original transcript. This can prove useful when working across cultures and languages, in which the same concepts could embody different meanings (Zureik, et al., 2006);
- Having local partners, either through a citizen science approach (Hardy, et al., 2016) or through formal collaboration with local universities, such as in the case of the researcher we studied working in Bangladesh, can ensure that participants' perspectives on personal privacy are respected. Dialogue with local partners regarding data collection helps ensure that participants are approached in settings where they feel safe, as in the case of R2, who conducted interviews in public. Local partners can also help in detecting whether the views of the researchers, *qua* cultural outsiders, affect the interpretation and analysis of the data (Hardy, et al., 2016; Jorge and Albagli, 2020). Local partners could also provide the participants with legal privacy protection that aligns with their own understanding of privacy and ensure shared stewardship (Alcalá, et al., 2016; First Archivist Circle, 2007). For instance, in R2's Bangladeshi example, involving local partners ensured that the original recordings were not moved outside Bangladesh.

Implications for research support services

As a result of the discussion, we provide the following advice to researchers and other stakeholders, as listed in Table 4, along with suggestions for how research data support services in universities should assist researchers in following this advice:

	Advice	Research support services should…
1	Researchers should always take care not to expose their participants to harm.	Include an ethics approach to privacy in research data management courses and training materials that target both the collectors and re-users of human subjects' data.

2	Assist researchers in finding solutions that do not compromise research quality by creating an understanding of different stakeholders' perspectives and motivations.	Focus on how to ensure research quality and transparency while protecting subjects' privacy by moving away from the open–closed dichotomy and their own ideals of open.
3	Use the entire legal space within the privacy legislation. Ensure that research participants have the authority to control data and the right to develop the cultural governance protocols highlighted in the CARE principles.	Create a dialogue on methods with legal experts and mediate between these experts and researchers to find solutions that allow innovative research.
4	Initiatives to address personal privacy in a global research context would be valuable to highlight cultural differences in privacy and promote discussion of how to respect these differences.	Facilitate seminars with the guidance of experts in applied research ethics to create a common platform for privacy protection in international research projects.
5	Encourage researchers to focus on context, transmission and actors when reflecting on how to protect the interest of data subjects.	Asist researchers in identifying the subject, sender and receiver at the different stages of the research data life cycle and use this as a basis for discussing strategies to empower data subjects and exercise cautions for privacy protection with a focus on transmission and context.
6	Use vignettes or double translations to ensure coherent understandings and translations of complex concepts.	Develop best-practice toolkits with examples of strategies that can be used to address power structures and protect privacy from the perspective of the subject.
7	Provide data subjects with knowledge of how privacy is protected throughout the life cycle and aim for shared stewardship and empowering the data subject.	Argue for and facilitate the subjects' right to be included and informed regarding decisions affecting their data.
8	Address differences regarding privacy and how data will circulate early in a project.	Assist researchers with designing informative and clear consent forms, in which the participants are provided with opt-out choices for data sharing.
9	Data subjects should be informed when data organisations take over responsibility for data.	
10	Explore how digital solutions can minimise distance between the data and the subject and provide participants with increased control of the data regarding themselves.	Be a driving force for investment in research data archives to balance privacy protection with access by having a dialogue with subjects.
11		Be a driving force for the inclusion of ethical training in data science curricula to ensure that the reuse of data follows the ethical standards expected from research on human subjects' data.

Table 4: Recommendation for research support services to follow up on advice given throughout the discussion

The interpretation of the law and possibilities to share and reuse data may conflict at two stages of the research data life cycle in particular: stage 2 in the design and collection of informed consent and stage 4, at the end of a project, when the data are either preserved or lost. Involving the research participants in stage 2 in decisions regarding the sharing or archiving of personal data is the best way to ensure participants' privacy. We recommend that researchers create a dialogue with the participants so that their opinions are heard.

In stage 4 of the research data life cycle, the participants are often unaware of the possibilities for preserving valuable research data, according to the GDPR art. 89 (1). Sharing human subjects' research data is often incompatible with the open publishing of data. Examples of strategies for the re-identification of data that are presumed to be anonymous illustrate how the sharing and reuse of research data containing personal information requires extra care and attention and how anonymisation is not always an option (Barocas and Nissenbaum, 2014). The importance of keeping records of current research and scholarship for future generations is currently not gaining enough attention (Thouvenin, et al., 2016). Long-term solutions for archiving human subjects' data with proper access control are necessary to protect current research histories from becoming lost. Privacy is far from dead, but it requires an infrastructure for data archiving with embedded and possibly also dynamic privacy solutions—preferably using globally distributed storage with access management, keeping the data local and the access global within the requirements of local norms and possibly also partnerships. The main challenge in designing such systems is the aggregation of the personal data necessary for facilitating dialogues with subjects.

Conclusion

Our findings demonstrate that researchers face the following challenges when sharing human subjects' data: 1) maintaining trust with research participants; 2) managing divergent views of privacy in international and intercultural research collaborations and 3) interpreting and applying policy. Successful data sharing requires aligning the work of multiple stakeholders, despite their often divergent perspectives and motivations.

Personal privacy protection in research involves respecting research participants, requiring awareness of roles, attributes and transmission principles. In digital research, multiple stakeholders are involved in data management, all of whom must demonstrate sensitivity towards data privacy and research participants. If and when data sharing is to take place, respecting the research participants and their perception of what information is sensitive and private must have priority.

The requirements of open research and international research collaborations make balancing personal privacy with data sharing a complex task for researchers. Providing expertise and guidance on how to best balance these requirements is part of research support and something that research data management support should offer. To facilitate the sharing of data 'as open as possible and as closed as necessary', we must acknowledge that different stakeholders in data sharing have different perspectives on how personal privacy and data sharing should be balanced. Increasing the quality and transparency of research must be the primary motivation for the sharing and reuse of data and must be carefully balanced with the privacy of the research participants when human subjects are involved.

Recommendations for further research and practical work

More knowledge and the sharing of best practices for balancing privacy with high-quality research without moving outside of the law are needed. We find that several researchers are interested in and motivated to share their data but struggle to find practical solutions to how privacy and open research can work together. Cases presenting knowledge on both solutions and potential hindrances would be helpful for RDSs in guiding researchers.

We also encourage the international research data community to involve privacy and research ethics experts in creating guidelines for protecting the privacy of research subjects in international research collaborations that involve data sharing. This could be achieved through the creation of an oversight board or a universal recommendation for how to protect privacy in dialogue with the subjects and, through this, empower the data subject and increase trust in research.

Acknowledgements

We would like to thank the participants of this study for sharing their time, knowledge, experience and thoughts on privacy and data sharing, which brought our attention to some of the difficulties in balancing privacy and research.

References

Alcalá, J. C., Star, S. L., & Bowker, G. C. (2016). Infrastructures for remembering. In S. Timmermans, A. E. Clarke, & E. Balka (Eds.), *Boundary objects and beyond: working with Leigh Star* (pp. 323–338). MIT Press.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. In J. Lane, V. Stodden, & H. Nissenbaum (Eds.), *Privacy, big data, and the public good: frameworks for engagement* (1st. ed., pp. 44–75). Cambridge University Press.

Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). International differences in information privacy concern: implications for the globalization of electronic commerce. In *Advances in Consumer Research* (Vol. 31, pp. 362–363). Association for Consumer Research.

Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. MIT Press.

Borgman, C. L. (2018). Open data, grey data, and stewardship: universities at the privacy frontier. *Berkeley Technology Law Journal*, *33*, 365.

Bowker, G. C. (2005). *Memory practices in the sciences*. MIT Press.

Budin-Ljøsne, I., Teare, H. J. A., Kaye, J., Beck, S., Bentzen, H. B., Caenazzo, L., Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M. K., Jones, E., Katić, V., Simpson, A., & Mascalzoni, D. (2017). Dynamic consent: a potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics*, *18*(1), 4. https://doi.org/10.1186/s12910-016-0162-9

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, *19*, 43. https://doi.org/10.5334/dsj-2020-043

cOAlition S. (2019). Accelerating the transition to full and immediate open access to scientific publications. https://www.coalition-s.org/wp-content/uploads/PlanS_Principles_and_Implementation_310519.pdf

Corti, L. (2014). *Managing and sharing research data: a guide to good practice*. SAGE Publishing.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd. ed.). SAGE Publishing.

Darch, P. T., Borgman, C. L., Traweek, S., Cummings, R. L., Wallis, J. C., & Sands, A. E. (2015). What lies beneath?: Knowledge infrastructures in the subseafloor biosphere and beyond. *International Journal on Digital Libraries*, *16*(1), 61–77. https://doi.org/10.1007/s00799-015-0137-3

de Koning, M., Meyer, B., Moors, A., & Pels, P. (2019). Guidelines for anthropological research: data management, ethics, and integrity. *Ethnography*, *20*(2), 170–174. https://doi.org/10.1177/1466138119843312

Dilger, H., Pels, P., & Sleeboom-Faulkner, M. (2019). Guidelines for data management and scientific integrity in ethnography. *Ethnography*, *20*(1), 3–7. https://doi.org/10.1177/1466138118819018

Elias, P. (2014). A European perspective on research and big data analysis. In J. Lane, V. Stodden, H. Nissenbaum, & S. Bender (Eds.), *Privacy, big data, and the public good: frameworks for engagement* (1st. ed., pp. 173–191). Cambridge University Press.

Eschenfelder, K. R., & Shankar, K. (2020). Of seamlessness and frictions: transborder data flows of European and US social science data. In A. Sundqvist, G. Berget, J. Nolin, & K. I. Skjerdingstad (Eds.), *Sustainable Digital Communities: 15th International Conference, iConference 2020*, Boras, Sweden, March 23–26, 2020, Proceedings (Vol. 12051, pp. 695–702). Springer International Publishing. https://doi.org/10.1007/978-3-030-43687-2

Ess, C., & Hård af Segerstad, Y. (2020). Everything old is new again. In Å. Mäkitalo, T. E. Nicewonger, & M. Elam (Eds.), *Designs for experimentation and inquiry: approaching learning and knowing in digital transformation* (pp. 179–196). Routledge.

European Commission. (2016). *Guidelines on FAIR data management in H2020.*

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. (2020). *CORDIS: Projects and results*. https://cordis.europa.eu/projects/en

European Research Council (ERC). (2017). *Guidelines on implementation of open access to scientific publications and research data—In projects supported by the European Research Council under Horizon 2020*. http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

First Archivist Circle. (2007). Protocols for Native American archival materials. https://www2.nau.edu/libnap-p/protocols.html

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, *7*(4), 185–200. https://doi.org/10.1007/s10676-006-0001-7

Floridi, L. (2013). *The ethics of information* (1st. ed.). Oxford University Press.

Guillemin, M., Barnard, E., Allen, A., Stewart, P., Walker, H., Rosenthal, D., & Gillam, L. (2018). Do research participants trust researchers or their institution? *Journal of Empirical Research on Human Research Ethics*, *13*(3), 285–294. https://doi.org/10.1177/1556264618763253

Hardy, L. J., Hughes, A., Hulen, E., & Schwartz, A. L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. *Journal of Empirical Research on Human Research Ethics*, *11*(2), 191–198. https://doi.org/10.1177/1556264616636233

Havemann, J., & Bezuidenhout, L. (in press). *Harnessing the open science infrastructure for an efficient African response to COVID-19*. https://doi.org/10.5281/zenodo.3733768

HEW Advisory Committee on Automated Data Systems. (1973). *The Code of Fair Information Practices* (p. 2). U.S. Department of Health, Education and Welfare.

Hofstede, G. (1984). *Culture's consequences: international differences in work-related values* (abridged ed., Vol. 5). SAGE Publishing.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations*. McGraw-Hill Professional Publishing.

Inness, J. C. (1992). *Privacy, intimacy, and isolation*. Oxford University Press.

Jackson, B. (2018). The changing research data landscape and the experiences of ethics review board chairs: implications for library practice and partnerships. *The Journal of Academic Librarianship*, *44*(5), 603–612. https://doi.org/10.1016/j.acalib.2018.07.001

Jorge, V. de A., & Albagli, S. (2020). Research data sharing during the Zika virus public health emergency. *Information Research*, *25*(1), 20.

Kim, Y. (2015). Social scientists' data sharing behaviours: investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, *35*(4), 408–418.

Kvale, L. (2021a). Using personas to visualize the need for data stewardship. *College & Research Libraries*, *82*(3). https://doi.org/10.5860/crl.82.3.332

Kvale, L. (2021b). *Data from a three-phase Delphi study used to investigate knowledge infrastructure for research data in Norway, KIRDN_Data v.2* [Data set]. Oslo Metropolitan University. http://doi.org/10.5281/zenodo.5582714

Kvale, L., & Darch, P. T. (2020). *Dealing with privacy—Personal privacy from a research data management perspective* [Poster presentation]. IConference 2020 Proceedings, Borås, 2020. http://hdl.handle.net/2142/106564

Kvale, L., & Pharo, N. (2021). Understanding the data management plan as a boundary object through a multi-stakeholder perspective. *International Journal of Digital Curation*, *16*(1). https://doi.org/10.2218/ijdc.v15i1.729

Landeta, J., Barrutia, J., & Lertxundi, A. (2011). Hybrid Delphi: a methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting and Social Change*, *78*(9), 1629–1641. https://doi.org/10.1016/j.techfore.2011.03.009

Lane, J., Stodden, V., Nissenbaum, H., & Bender, S. (Eds.). (2014). *Privacy, big data, and the public good: frameworks for engagement* (1st. ed.). Cambridge University Press.

Lee, J.-S., & Jeng, W. (2019). *The landscape of archived studies in a social science data infrastructure: Investigating the ICPSR metadata records*. *Proceedings of the Association for Information Science and Technology*, *56*(1), 147–156.

McDonald, M., Townsend, A., Cox, S. M., Paterson, N. D., & Lafrenière, D. (2008). Trust in health research relationships: accounts of human subjects. *Journal of Empirical Research on Human Research Ethics*, *3*(4), 35–47. https://doi.org/10.1525/jer.2008.3.4.35

Melinder, A., & Milde, A. M. (2016). Samarbeid, samtykkeerklæring og deling av data [Collaboration, consent forms and sharing of data]. In V. Enebakk, H. Ingierd, & N. O. Refsdal (Eds.), *De berørte etter 22. Juli; Forskningsetiske perspektiver* [Research ethical perspectives]. Cappelen Damm Akademisk. https://press.nordicopenaccess.no/index.php/noasp/catalog/download/6/15/57-1

Modjarrad, K., Moorthy, V. S., Millett, P., Gsell, P.-S., Roth, C., & Kieny, M.-P. (2016). Developing global norms for sharing data and results during public health emergencies. *PLOS Medicine*, *13*(1), e1001935. https://doi.org/10.1371/journal.pmed.1001935

National Science Foundation. (2011). *Chapter VI—Other post award requirements and considerations*. https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

Nissenbaum. (2010). *Privacy in context: technology, policy, and the integrity of social life*. Stanford University Press.

OECD. (1980). *OECD guidelines on the protection of privacy and transborder flows of personal data*. http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

OECD. (2013). *The OECD privacy framework* (p. 154). OECD. https://www.oecd.org/internet/ieconomy/privacy-guidelines.htm

OECD. (2017). *OECD guidelines on measuring trust*. OECD. https://doi.org/10.1787/9789264278219-en

Palmer, C. L., & Cragin, M. H. (2008). Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, *42*(1), 163–212.

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: relationships, activities, drivers and influences. *PLOS One*, *9*(12), e114734. https://doi.org/10.1371/journal.pone.0114734

The Norwegian Personal Data Act. (2018). Lov om behandling av personopplysninger [The Norwegian Personal Data Act], LOV-2018-06-15-38. https://lovdata.no/dokument/LTI/lov/2018-06-15-38 (Archived by the Internet Archive at https://web.archive.org/web/20211221135801/https://lovdata.no/dokument/LTI/lov/2018-06-15-38)

Rappert, B., & Bezuidenhout, L. (2016). Data sharing in low-resourced research environments. *Novation*, *34*(3–4), 207–224. https://doi.org/8109028.2017.1325142

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. http://data.europa.eu/eli/reg/2016/679/oj/eng (Archived by the Internet Archive at https://web.archive.org/web/20211221111954/https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng)

Research Data Alliance. (2020). *RDA COVID-19 recommendations and guidelines on data sharing*. https://doi.org/10.15497/rda00052

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd. ed.). SAGE Publishing.

Scroggins, M. J., Pasquetto, I. V., Geiger, R. S., Boscoe, B. M., Darch, T., Cabasse-Mazel, C., Thompson, C., Golshan, M. S., & Borgman, C. (2019). Thorny problems in data (-intensive) science (p. 10). *UCLA: Center for Knowledge Infrastructures*. https://escholarship.org/uc/item/31b1z69c

Shah, N., Viberg Johansson, J., Haraldsdóttir, E., Bentzen, H. B., Coy, S., Mascalzoni, D., Jónsdóttir, G. A., & Kaye, J. (2021). Governing health data across changing contexts: a focus group study of citizen's views in England, Iceland, and Sweden. *International Journal of Medical Informatics*, *156*, 104623. https://doi.org/10.1016/j.ijmedinf.2021.104623

Shankar, K. (1999). Towards a framework for managing electronic records in scientific research. *Archival Issues*, *24*(1). https://minds.wisconsin.edu/bitstream/handle/1793/45890/MA24_1_3.pdf?sequence=3&origin=publication_detail

Solove, D. J. (2002). Conceptualizing privacy. *California Law Review*, *90*, 71.

Solove, D. J. (2010). *Understanding Privacy* (2/28/10 ed.). Harvard University Press.

Tapio, P., Paloniemi, R., Varho, V., & Vinnari, M. (2011). The unholy marriage? Integrating qualitative and quantitative information in Delphi processes. *Technological Forecasting and Social Change*, *78*(9), 1616–1628. https://doi.org/10.1016/j.techfore.2011.03.016

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2013). Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, *1*(39), 70–78.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. J., & Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly*, *27*(1), 23–44. https://doi.org/10.18352/lq.10180

Tenopir, C., Allard, S., Pollock, D., Hughes, D., Lundeen, A., Sandusky, R. J., & Baird, L. (2019). Academic librarians and research data services: attitudes and practices. *IT Lib: Information Technology and Libraries Journal*, *1*. https://itlib.cvtisr.sk/wp-content/uploads/docs/24_academic%20librarians.pdf

The Global Indigenous Data Alliance, G. (2019). *CARE principles for Indigenous data governance*. https://www.gida-global.org/care

Thouvenin, F., Burkert, H., & Hettich, P. (2016). *Remembering and forgetting in the digital age – a position paper*. *Information Research*, 21(1). http://informationr.net/ir/21-1/memo/memo2.html#.YcSCmC8w2gQ

U.S. Department of Health, Education, and Welfare. (1979). *The Belmont report* (p. 10).

U.S. Department of Homeland Security. (2012). *The Menlo report: ethical principles guiding information and communication technology research*.

Viberg Johansson, J., Shah, N., Haraldsdóttir, E., Bentzen, H. B., Coy, S., Kaye, J., Mascalzoni, D., & Veldwijk, J. (2021). Governance mechanisms for sharing of health data: an approach towards selecting attributes for complex discrete choice experiment studies. *Technology in Society*, *66*, 101625. https://doi.org/10.1016/j.techsoc.2021.101625

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: a content analysis of libraries' websites. *College & Research Libraries*, *78*(7). https://doi.org/10.5860/crl.78.7.920

Ziglio, E. (1996). The Delphi method and its contribution to decision-making. In M. Adler & E. Ziglio (Eds.), *Gazing into the Oracle—The Delphi method and its application to social policy and public health* (pp. 3–33). Jessica Kingsley.

Zureik, E., & Stalker, L. H. (2010). The cross-cultural study of privacy: problems and prospects. In E. Smith, D. Lyon, Y. Chan, E. Zureik, & L. H. Stalker (Eds.), *Surveillance, privacy, and the globalization of personal information: international comparisons* (pp. 9–30). McGill–Queen's University Press. https://www.jstor.org/stable/j.ctt1q606m

Zureik, E., Stalker, L. H., & Smith, E. (2006). *Background paper for the globalization of personal data project international survey on privacy and surveillance* [The Globalization of Personal Data Project, Queen's University].