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 frictions 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 using themes following exploratory sequential design methods. After three rounds of data collection, the entire corpus of data were connected and analysed thematically according to integrated analysis.*
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, *et al*,, 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, *et al.*, 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 on research data services. 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).

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 research 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.

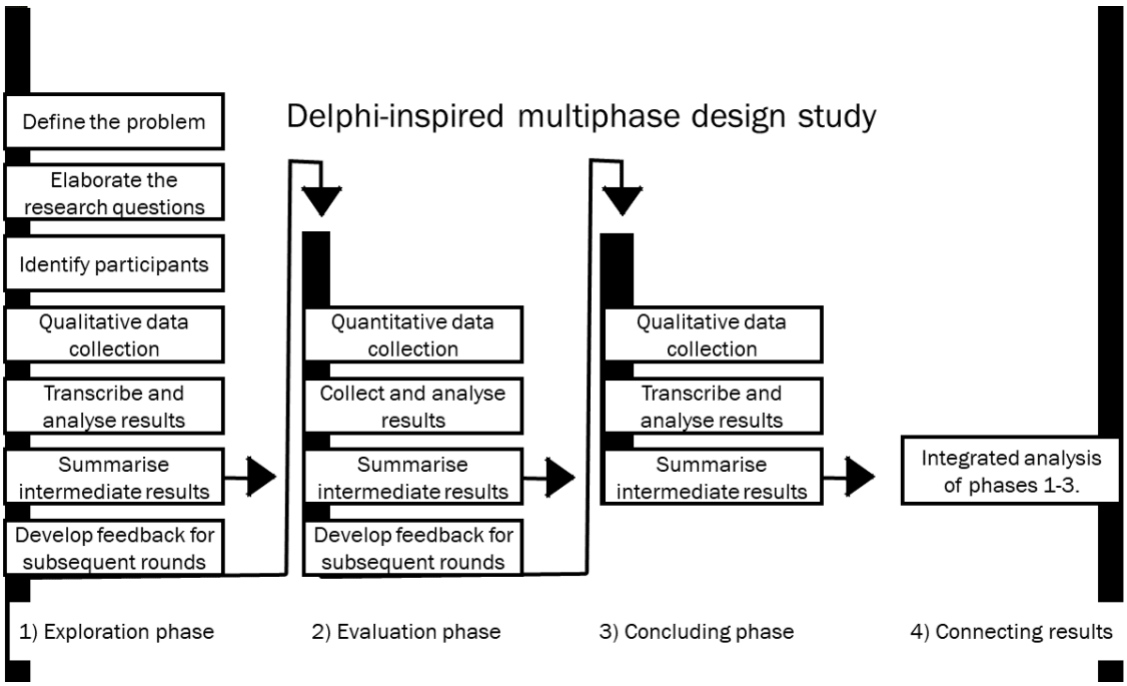

Figure 2: The research design (Kvale and Pharo, 2021)

In the first phase, conducted at the beginning of 2018, the interviews were approximately one hour long, yielding a total of twenty-four 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 ten 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 twelve hours of recordings and ninety-eight 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 as that; it takes so little to destroy that trust, and by that, remove the will to fund’. (IN3)
Cost profit	Thought or experiences regarding the cost and profit aspect of data archiving	

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 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	'This is fairly new at the university, and 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?".	'Quality assurance must be a 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

reproducible'. (RO1)	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)	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. Before 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.

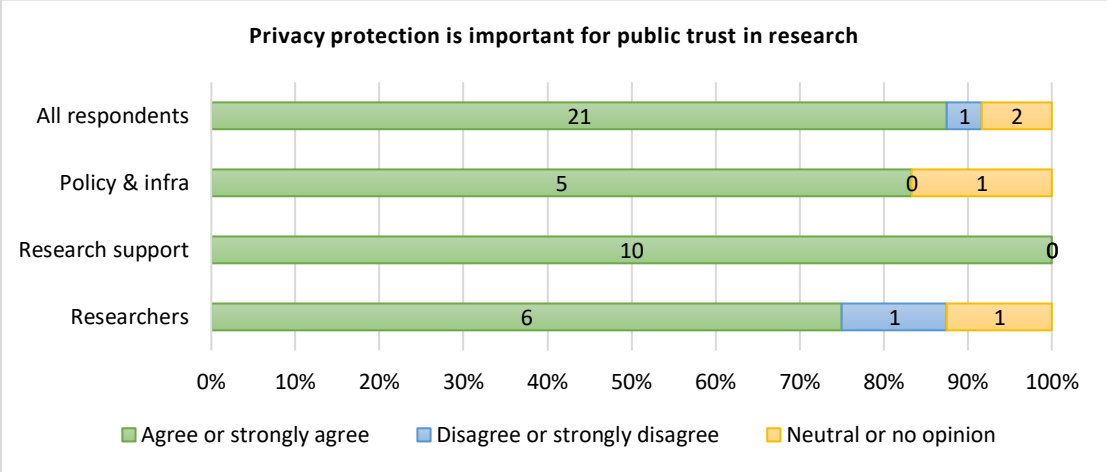

Figure 3: Privacy protection and public trust in research; responses from the questionnaire

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 sharing 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 what was considered sensitive 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).

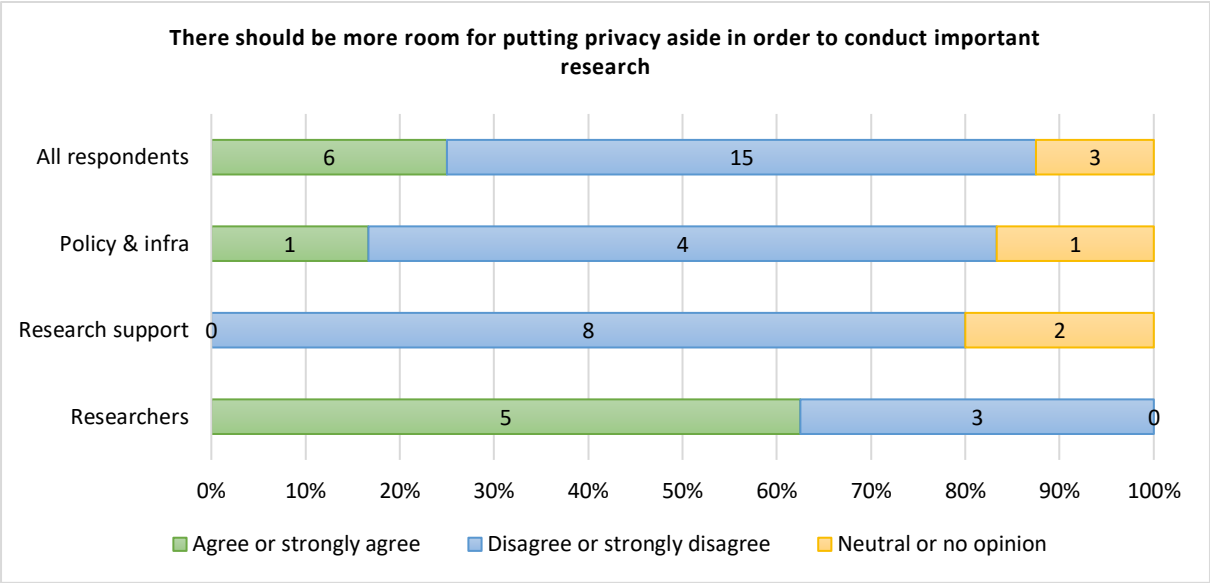

Figure 4: Putting privacy aside to conduct research

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).

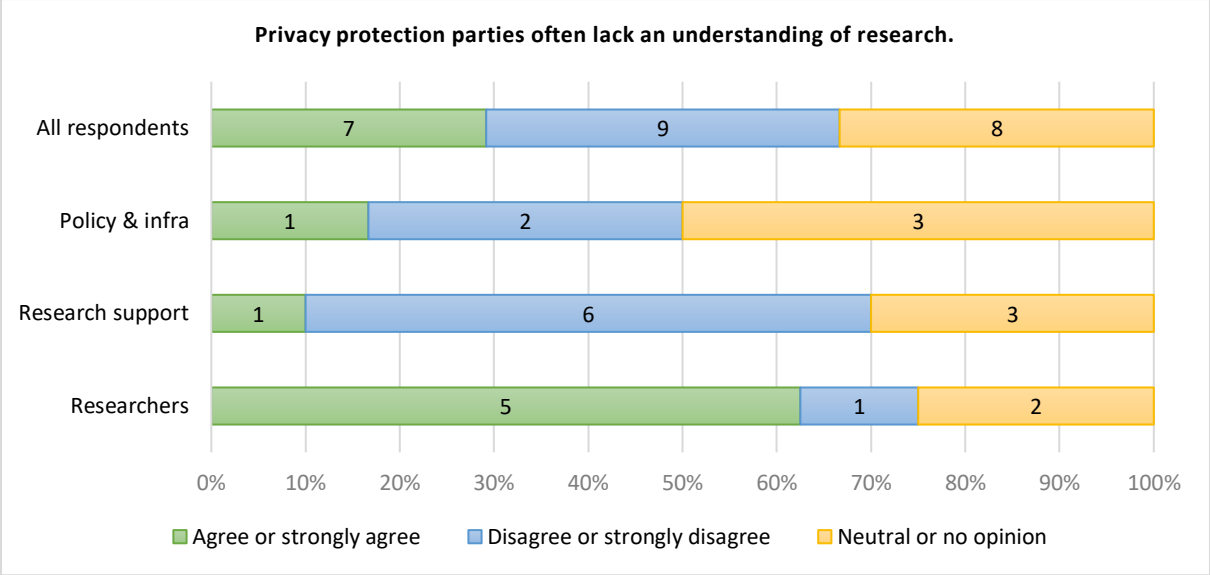

Figure 5: Stakeholders' views on privacy protection parties and their understanding of research

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.

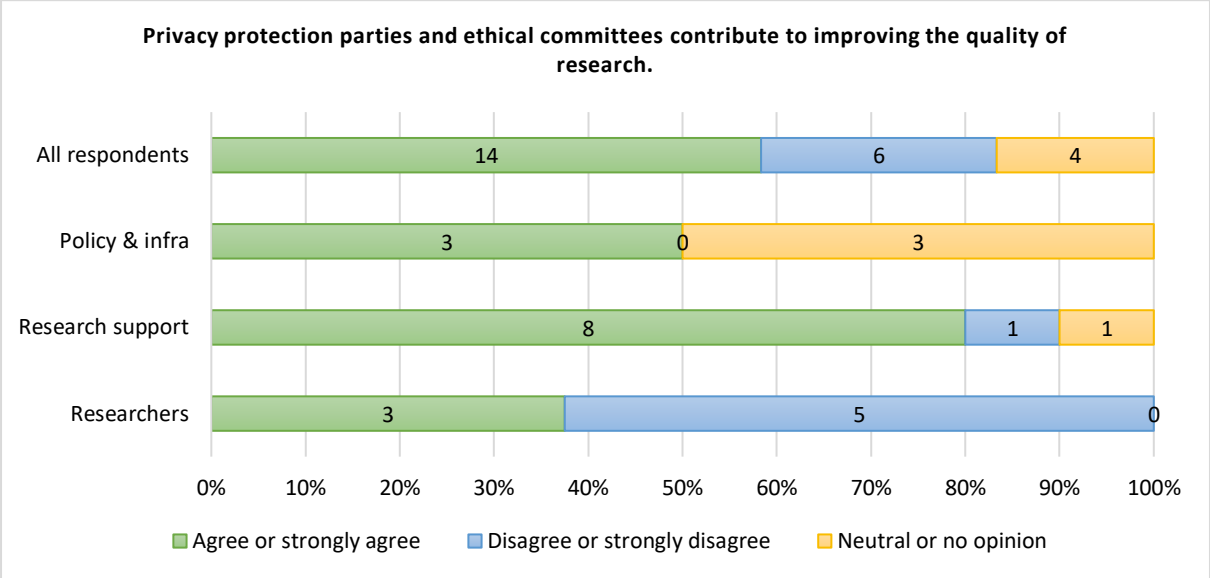

Figure 6: Stakeholders' views on privacy protection parties and their contribution to research quality

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.

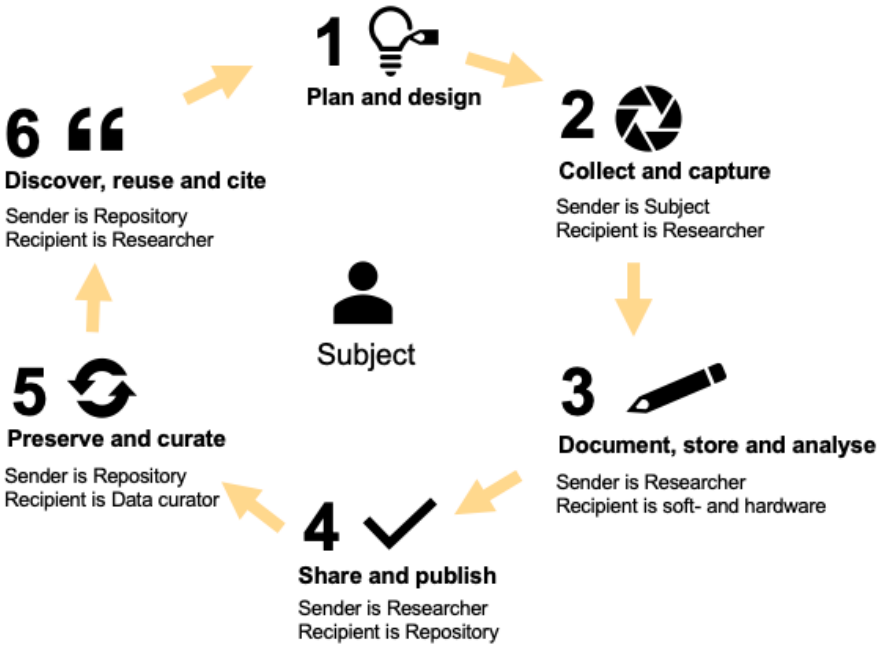

Figure 7: Personal privacy in the research data life cycle

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 before 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 (Viberg Johansson, *et al*., 2021; 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' [replace 'stakeholder'] 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.	Inform and assist researchers in privacy protection including principles of shared stewardship.
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.

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 (Archived by the Internet Archive at https://archive.org/details/PlanS_Principles_and_Implementation)

Corti, L., Van der Eynden, V., Bishop, L., & Woollard, M. (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 (Archived by the Internet Archive at https://archive.org/details/h2020-hi-

oa-data-mgt_en)
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 (Archived by the Internet Archive at https://archive.org/details/h2020-hi-erc-oa-guide_en)
First Archivist Circle. (2007). Protocols for Native American archival materials. https://www2.nau.edu/libnap-p/protocols.html (Archived by the Internet Archive at https://archive.org/details/print-protocols)
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 (Archived by the Internet Archive at https://archive.org/details/award-and-administration-guide)

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 (Archived by the Internet Archive at https://archive.org/details/oecd-guidelines-on-the-protection-of-privacy-and-transborder-flows-of-personal-data-oecd)

OECD. (2013). *The OECD privacy framework* (p. 154). OECD. https://www.oecd.org/internet/ieconomy/privacy-guidelines.htm (Archived by the Internet Archive at https://archive.org/details/oecd-legal-0188-en)

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 (Archived by the Internet Archive at https://archive.org/details/ma-24-1-3)

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 (Archived by the Internet Archive at https://archive.org/details/24-academic-librarians)
The Global Indigenous Data Alliance, G. (2019). *CARE principles for Indigenous data governance*. https://www.gida-global.org/care (Archived by the Internet Archive at https://archive.org/details/care-principles-of-indigenous-data-governance-global-indigenous-data-alliance)
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].