

Dealing with Privacy

Personal Privacy from a Research Data Management Perspective

Live Kvale¹[0000-0001-5264-0382] and Peter T. Darch²[0000-0002-0648-4611]

¹ Oslo Metropolitan University, Oslo, Norway

² University of Illinois at Urbana-Champaign, Urbana, IL, USA

live.kvale@oslomet.no

ptdarch@illinois.edu

Abstract. Sharing research data is increasingly required by key stakeholders, such as funding agencies and journals. However, sharing human subjects data poses multiple issues around protecting these subjects' privacy. This poster examines these issues through a Delphi study, comprising interviews and questionnaires, of stakeholders (n=24) involved in data curation and sharing in Norway. Two particular contexts are considered: 1) privacy in international research collaborations and 2) maintaining research subjects' trust in the researcher. The findings presented in this poster show tensions between maintaining subjects' privacy, maintaining trust, and advancing research. These tensions are complicated further by conflicting perspectives on privacy held by different stakeholders. Researchers and other stakeholders must balance these aspects throughout the data lifecycle, from data collection to decisions about sharing. All stakeholders involved in data sharing should pay attention to research subjects through dialog and negotiation.

Keywords: Data Sharing, Data Curation, Research Data, Privacy, Open Data.

1 Introduction

Digital research methods and data-driven research are increasingly becoming the norm in many scientific disciplines [1]. This transition involves multiple stakeholders, including data management professionals and infrastructure providers. Funding agencies increasingly require research data to be shared with other researchers [2]. Concerns regarding protection of personal privacy¹ frequently appear to be the elephant in the room in the drive for data sharing, often present but rarely addressed explicitly.

The next section contains background information and theoretical perspectives, which is followed by our research question. Thereafter we present our methods, before Findings and Conclusions.

¹ The concept of privacy is complex. In this poster we define privacy as respect for an individual's right to control over data about themselves

2 Background

Sharing of human subjects research data can pose many challenges for protecting subjects' privacy. While giving research subjects complete control over how data about themselves are subsequently used is seen as the ideal [3], privacy may not always be respected by researchers, potentially compromising trust in researchers [4] and the willingness of people to participate in future studies [5]. Another issue is that researchers may mistakenly believe that removal of directly identifiable information will protect privacy [6] and that participants do not need to be informed that data about them may be shared. However, even anonymized data are liable to be re-identified, as has happened in multiple cases [7, 8].

Addressing these challenges is complicated by a lack of agreement among research stakeholders about what privacy is, and what constitutes researchers' responsibilities to their subjects. Floridi describes two traditions of understanding privacy in the western world [4]. One focuses on consequences, namely that privacy is useful for preserving human dignity. The second argues that privacy needs to be respected because of each person's right to bodily security and property. These perspectives are enshrined in principles governing human subjects research, as described in the Belmont and Menlo reports [9,10]. This lack of agreement is compounded when research data are shared across national and cultural boundaries. One scenario where sharing data internationally is regarded as a necessity is in the case of global health emergencies, such as the recent Ebola crisis. However, a lack of shared norms can create barriers to sharing [11].

3 Research question

Various stakeholders make decisions and take actions that affect how, whether, and under what circumstances research data may be shared – and thus whether research subjects' privacy is maintained. Understanding the perspectives of these different stakeholders, and how these perspectives affect data sharing, motivates the following research question:

How do stakeholders address privacy with respect to research data sharing?

4 Method

The first author conducted a Delphi study, characterized by the usage of an expert group and data collection in multiple rounds focusing on 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 contain three phases: 1) the exploration phase, 2) the evaluation phase, and 3) the concluding phase [12]. In each phase data are collected and analyzed and the intermediate results are used in the development of the next phase of data collection. When the data collection was completed, a fourth integrated analysis of all the data was conducted. Inspired by a multiphase design mixed method study [13],

the first and third phases involved interviews and the second phase comprises a questionnaire (Fig. 1).

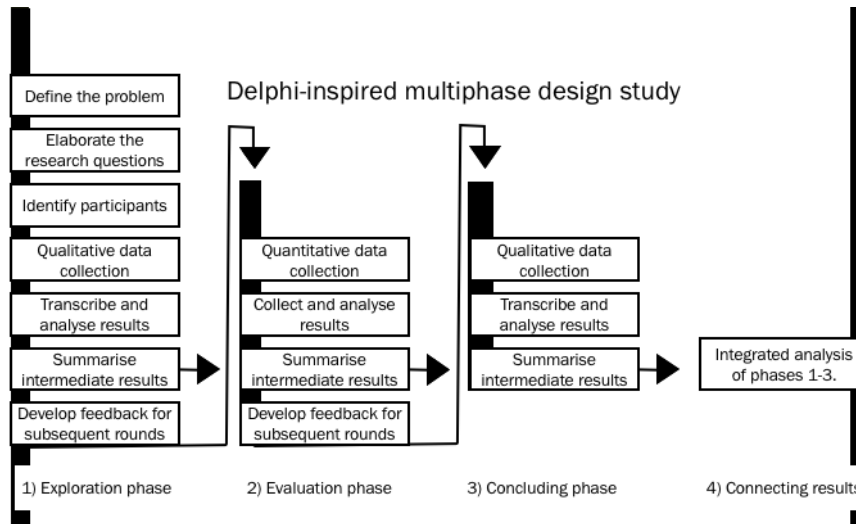


Fig. 1. The research design [14]

A group of 24 experts took part in the study (Table 1), comprising policymakers, national infrastructure providers, researchers, and research support staff from four Norwegian universities. Participants were chosen based on their involvement in developing policies, infrastructure or other research data support.

Research support staff were recruited to cover the full range of research support services (Library, IT, and research office). Researchers were recruited from different disciplinary backgrounds (humanities, sciences, and social sciences).

In the first phase the interviews were approximately one hour long, open and exploratory. Participants were asked about issues they encountered during all stages of the research data lifecycle. In the second round of data collection, participants answered a questionnaire where they were asked to share their opinion on different statements originating from the analysis of the first round of interviews. The third round involved interviews, following up on issues discovered in round one and two. The results presented in this poster are from all three rounds of data collection.

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

Table 1. The participant group

5 Findings

The conflict of sharing versus privacy is often a core ethical challenge of research data sharing, as reflected in the following quotation from an interviewee:

When it comes to storing of data and management of data, I believe that apart from the purely technical question there is a fundamental conflict between different values, the need for high quality scholarship and personal privacy (RLJ)

We investigated how this conflict is negotiated by various stakeholders in two contexts: international research collaboration, and researcher-participant relations.

5.1 Privacy in international research collaborations

Some interviewees in international collaborations encountered tensions between their understandings of privacy and perspectives of collaborators from other cultures. These tensions created barriers to data sharing.

LM worked on an international data collection where sharing of data was challenged by different perceptions of which parts of the interview material that 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, I remember finding it a bit strange, that this was considered sensitive information by the Japanese researchers and could not be made available. So, the way I understand it the Japanese data became almost useless if you need to anonymize according to the Japanese norms where food preferences were considered sensitive information, as a result that data could not be made available. While that is of course not the standard for the Norwegian data we collected in the same project. So, I have seen how different cultures apply different understandings of what is sensitive, and in this case, we had less data available for reuse. (LM)

In this case, negotiation to create a common understanding among researchers did not happen. Instead, researchers held to their own definitions of privacy. As a result, the data collected by the Japanese researchers was not shared with other researchers.

RGV encountered a similar situation. However, it was partially resolved due to having research partners from local universities on site, and using their knowledge in ethical reflections concerning the participants:

We came to Bangladesh and did these studies there. We realized that even if we had only one informant in a village, then not all of the village, but 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'. And things like this, which means sharing relative sensitive information with others look different in a western context than in many other cultural contexts where you don't have the individual based, but the group based. I find this to be a task for empirical ethics to understand how this works and what are the privacy limits, what can we allow and what is abuse? And sometimes we must be aware not to fall in implicit power structures where respondents just rely on what their superior expects, but yes there are different structures, and we need much more reflections. (RGV)

These examples illustrate the need to understand privacy as a cultural phenomenon. For RGV, a researcher based in Norway, her cultural perspective on personal privacy as an issue for the individual only became a challenge when doing research in a different culture. Instead, research should be grounded in an understanding of privacy, which should be supplemented with cultural data that can also permit more group-based understandings as well. RGV also described challenges when it came to storing, depositing, or deleting the data from the project. As she had partners in Bangladesh, the original recordings and transcripts were kept in Bangladesh, while the Norwegian researchers worked on translated transcripts from the interviews. Subsequently the re-

quirements from Norwegian data protection services² of deleting the original recording did not apply as these were kept in Bangladesh and never transferred.

5.2 Trust in the researcher

When researchers are working on studies that involve long-term engagement of participants, even over decades, maintaining trust between participants and researchers is critical. However, data sharing may undermine this trust if it compromises participant privacy. RGD describes this issue in the following quotation:

I see now that we ask for more input from their side in the various part of the research and [my colleague] was amazed how committed they are and how much trust they put here. So, I think that this is the key here, and we always have to do everything to maintain the trust [...] But of course there can always be issues like if you don't take the ethics into account the research could even go much further, [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. (RGD)

RGD describes how the researcher limits what they ask to protect their relationship with the participant. She is surprised about how many details the participants reveal about their lives, later referring to cases where participants share more about their medical history than they are asked. By sharing these details, participants trust researchers to protect their confidentiality privacy and not to be negligent with the data. Another interviewee, RIL, describes how she balances the sharing of stories and the integrity of the participants when she analyzes every quote she uses in her book:

In this study, I knew that the [context] through so many years would make it possible for the participants to recognize each other. So, when the book was published, I reviewed it thoroughly, considering every quote and whether this would feel intimidating for anyone. If so, I changed the name to "a man" or "a woman" [...] Which I believe was successful. I met one participant that was unhappy about one quote which she herself had interpreted in a negative way, but other participants told me it was incredible how I managed to write it without exposing anyone, at the same time as they recognized themselves. So, I believe that regarding the ethical commitments held by the researcher it is possible to use your head to think when something is offensive and not. (RIL)

RIL is aware that she cannot achieve full anonymity for the participants. Her active reflections on which quotes to use, which data to share, and how to use them becomes more important as anonymity becomes increasingly difficult to guarantee. Both RGD and RIL focus on dialog with participants: they know their participants and want to maintain their trust by protecting their confidentiality privacy and by not asking for

² <https://nsd.no/personvernombud/en/index.html>

more than they will use in the research project. These researchers are engaged in negotiation and continuous reflection on how to balance respect for the individual and their privacy with advancing their research.

6 Discussion and Conclusions

Sharing of research data containing personal and/or sensitive information is challenging. Cultural understandings of privacy vary, and they affect how data can be shared in international research collaborations. In order to respect the individual participant the researchers should not enforcing their own understanding of what is sensitive information but reflecting on possible power structures and the cultural context of the participant.

Researchers working with committed participants in longitudinal studies both emphasized how the participants' trust in both the researcher as an individual and the research context is crucial for data and research quality. The researcher handles this by protecting the identity of the participants, not just in terms of anonymity, but also in regard to what information they share. This reflects Floridi's description of the relational trust existing in the sharing of private information [4] Through direct contact with participants, the researchers focus on respecting the individual and balancing this in their work.

Personal privacy is about respecting the individual, i.e. the research participants in the context of research. In digital research, more stakeholders are involved in data management, all of who need a sensitivity towards data privacy and the research participants. If data sharing is to take place, respecting the individuals that share their knowledge and their perception of what is sensitive and private must have priority. Our next step is to explore further the tension between practice and the application of legal and policy frameworks, aiming at developing guidelines for how and when personal data can be reused while maintaining respect for individual participants.

References:

1. Hey, T., Tansley, S., & Tolle, K. (eds.): *The Fourth Paradigm: Data-Intensive Scientific Discovery* (1 edition). Redmond, Washington (2009).
2. European Research Council (ERC): *Guidelines on Implementation of Open Access to Scientific Publications and Research Data—In projects supported by the European Research Council under Horizon 2020*. (2017).
3. 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 (2016).
4. Floridi, L.: The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology*, 7(4), 185–200 (2005).
5. Deshmukh, P., & Croasdell, D.: HIPAA: Privacy and Security in Health Care Networks. *Information Ethics: Privacy and Intellectual Property*, 219–238. <https://doi.org/10.4018/978-1-59140-491-0.ch013> (2005).

6. Wjst, M.: Caught you: Threats to confidentiality due to the public release of large-scale genetic data sets. *BMC Medical Ethics*, 11(1), 21 (2010).
7. Zimmer, M.: “But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325 (2010).
8. Zeller, T.: AOL Moves to Increase Privacy on Search Queries. *The New York Times* (2006, August 22)
9. U.S. Department of Health, Education, and Welfare.: *The Belmont Report* (1979).
10. U.S. Department of homeland security.: *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. (2012).
11. Modjarrad, K., Moorthy, V. S., Millett, P., Gsell, P.-S., Roth, C., & Kieny, M.-P.: Developing Global Norms for Sharing Data and Results during Public Health Emergencies. *PLOS Medicine*, 13(1), e1001935 (2016).
12. Ziglio, E.: 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* (pp. 3–33). Jessica Kingsley, London (1996).
13. Creswell, J. W., & Plano Clark, V. L.: Choosing a Mixed Methods Design. In *Designing and Conducting Mixed Methods Research* (pp. 53–106). SAGE Publications Ltd, Thousand Oaks (2011).
14. Kvale, L., & Pharo, N.: Understanding the data management plan as a boundary object through a multi-stakeholder perspective. Manuscript submitted for publication (2020).