Research Data in Scholarly Practices: Observations of an Interdisciplinary Horizon2020 Project

Madeleine Dutoit(✉)

Oslo and Akershus University College of Applied Sciences, Oslo, Norway
madeleine.dutoit@hioa.no

Keywords: Research data · Sharing · Interdisciplinary · Scholarly practices · Scholarly communication · Horizon2020

1 Research Topic and Questions

My research project belongs within Library and Information Science in the area of Scholarly Communication. It started in October 2016 and will continue for three years. The study focuses on researchers and research data - more specifically on research data sharing in the scholarly practices of an interdisciplinary research project mandated by a data policy. A Horizon2020-project including four different disciplines will be investigated. These EU-projects are by default obligated to develop data management plans (DMPs). Few studies have been done on this subject and the questions to answer are many. I have chosen to focus on the following questions: what does data mean to the different participating disciplines? How do the researchers work with the data of the project and how do they share the data between the represented disciplines? What are the effects of the data policy on the daily research work?

2 Introduction

Within the area of Scholarly Communication, many things are changing rapidly today. One of these changes concerns how research data is viewed and valued. Different stakeholders demonstrate an unprecedented level of interest in how researchers communicate their findings (Jubb 2013). Governments, universities and research funders, among them actors as the OECD and the EU, are currently formulating digital data policies that requires granted research projects to develop plans for their data management. Researchers receiving funding are expected to develop data management plans where data storage, handling and access is specified. Within scholarly communication research data as new actors are predicted to become "recognized as significant scholarly contributions in their own right" (Hey et al. 2009). Other authors terms this development as an institutionalization of open access to research data (e.g. Mauthner and Parry 2013) but it is perhaps early to use this term, since policies have not yet been in place or applied consistently long enough. In either case, researchers sharing data with one another is per se nothing new, but the external demands on researchers to do this sharing are. I see

J. Kamps et al. (Eds.): TPDL 2017, LNCS 10450, pp. 623–627, 2017.
DOI: 10.1007/978-3-319-67008-9_57

these policies as constituting a prominent example of this new interest coming from outside the researchers closest community, directing towards openness to research data. These policies are changing the conditions for and practices of data sharing.

Then what does research data mean and what are the drivers behind this development? To begin with, the concept of research data is complex and research seems only to agree in that no single definition is sufficient. Data have many kinds of value that varies widely over place, time and context (Borgman 2015). She, like other authors, agrees on that the sometimes more interesting question than what are data is *when* are data, claiming data to be "emergent, relational, and shaped by their use" (Haider and Kjellberg 2016). However, in order to analyse data in the context of scholarly communication, Borgman decides that a narrower approach to the concept will suffice which is why I settle with her definition of data as "representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship" (Borgman 2015). The general underlying motive for the aspirations of opening up access to data is the idea that accessible research data can contribute to benefit both research itself and the society in general. The driving arguments are mainly political, quality improvement of research through facilitation of transparency, and economical, enabling researchers and other interested to utilise the data for further research or innovation (NordForsk 2016).

I would here like to clarify what open data means in this discussion, since what is meant by openness and the degree of accessibility varies. Open can mean free and accessible data posted on a researchers personal website or published alongside a scholarly journal article as well as data deposited in a repository accessed only after registering and requesting it. In a report by The UK Royal Society speaks of "qualified openness" in the meaning of open data as "accessible, useable, assessable and intelligible" data (Royal Society 2012), implying that not all data are equally interesting or important. Though different approaches to openness, stakeholders agree that not all data can be made open, certain data will remain confidential for commercial, safety, privacy or security reasons.

3 Contributions

The results of this study will have both theoretical and practical implications. Learning more on how researchers from different disciplines collaborate on research data issues and how the data is shared with other project members, will increase our knowing on how knowledge is created jointly in temporary research projects. I hope to clarify the process of shaping common decisions on questions on research data and factors influencing this process. Additionally, with more knowledge of the epistemological bases of data in scholarly practices, the researchers' daily data "doings" and negotiations, and what data means to them, these behaviours can be better understood. Practically this new knowledge can be used when developing and forming services adapted to support the needs of research groups.

Work remains to be done in order to reach the economic and political goals set up for open data sharing. There is a gap between the aspirations of openness of research

data and the actual sharing being done. The above mentioned data policies rarely describe data as the complex heterogeneous phenomenon it actually is, moreover varying from discipline to discipline, but rather simplifies without recognizing the many obstacles present before a realization of these demands is possible. Some problems that have to be resolved are technical issues, others are infrastructural. Many authors indicate that these problems are the easier ones to solve. What is needed and is crucial, but more difficult to achieve, is to change the scholarly cultures and practices of researchers (Hey et al. 2009). And studies on disciplinary cultures show that we are facing multiple cultures (e.g. Kim and Stanton 2016).

In order to support the process of changing scholarly practices and develop infrastructure supporting data sharing, further research on how research data in scholarly practices looks like within interdisciplinary collaborations. Studies on the practices of scholars related to research data will "be imperative to improve our understanding of both the epistemological bases and the actual practices that arise from new forms of collaboration and novel approaches to data management" (Palmer and Cragin 2008).

4 Theoretical Framework

For this project the theoretical approach is based on practice theory. Although there is no such thing as a unified practice theory or single practice-based approach, its origin coming from different intellectual backgrounds, the way of how to view organizational knowing is central and unifying. This challenges today's assumptions of knowledge based on rationalistic and cognitivist learning (Cox 2012) in considering knowing as "situated in the system of ongoing practices of action, as relational, mediated by artefacts, and always rooted in a context of interaction" (Nicolini et al. 2003). The philosopher Schatzki describes a practice as an "array of human activity" and as "bodily doings and sayings" (Schatzki et al. 2001). They are routine-based activities and things said or unsaid, and they materialize un-reflected knowledge. For studies in LIS, this approach can therefore serve as a useful instrument when examining the social aspects of scholarly practices, the tacit knowledge of researchers and their approach in information use or information sharing.

Based in practice theory, moving away from the individualistic focus, the scholar will be studied mainly as a member of a community. She will be seen as a carrier of a practice, neither autonomous nor dependent of social culture (Cox 2012). The results will thus be analysed seeing data activities less governed by individual needs and more as having a social nature: as an information activity that is woven through social practices.

5 Methodology

Three different qualitative methods for studying research data in scholarly practices have been chosen since qualitative methods are well suited to describe phenomena in context and provide an interpretation that leads to a greater understanding of the phenomena (Justesen and Mik-Meyer 2012). The methods I find most suitable are interviews, focus

groups and participant observations. Researchers within an identified research project funded by the EU Research and Innovation programme Horizon 2020 will be studied. This project has been chosen firstly since projects within Horizon 2020 from January 2017 by default are part of the Open Data Pilot and thereby "must deposit your data in a research data repository where they will be findable and accessible for others" (European Commission 2016). In order to specify data collection, handling, sharing and curation, the projects are suggested to develop data management plans. Secondly, the project was chosen since it is constituted of researchers representing four different disciplines. The researchers thus have different traditions of handling data and their data will differ. Different scholarly practices will meet when project participants will discuss and agree on data management issues.

I have chosen to observe one single Horizon2020-project for this study in order to make it fit my time-frame and to be able to go deep into the subject; I want to search for "thick descriptions". The disadvantage with this choice is that it is difficult at this stage to say how representative my study will be. However, it would be quite difficult to find projects identically constituted in order to make an exact comparison. Naturally it would be very interesting to investigate several projects of this kind, to make a largely scaled qualitative study of this kind. Unfortunately, this will have to be for others to realize.

Semi-structured individual interviews as well as focus groups will be conducted with a yet unknown number of the researchers in order to find information of data in their scholarly practices, of cultures and norms, that can answer my research questions. Focus groups will be used as a "parallel force" to the interviews allowing me to observe the process of the participants managing their role both as an individual as a representative of the collective (Barbour 2013). Additionally, groups can prompt talk and interacting in offering other audience than the researcher (Macnaghten and Myers 2007) and points of view are argued for or defended in dialogue with the other participants (Tenopir et al. 2011). This information hopefully allows me to identify or grasp the data in the scholarly practices of the group and what is considered general behavior and unusual.

References

Barbour, R.S.: Analysing focus groups. In: Flick, U. (ed.) The SAGE Handbook of Qualitative Data Analysis. SAGE Publications, London (2013)

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

European Commission: What is the Open Research Data Pilot? (2016). https://www.openaire.eu/opendatapilot. Accessed 2 June 2017

Cox, A.M.: An exploration of the practice approach and its place in information science. J. Inf. Sci. **38**, 176–188 (2012)

Haider, J., Kjellberg, S.: Data in the making. In: Rekers, J.V., Sandell, K. (eds.) New Big Science in Focus: Perspectives on ESS and MAX IV. Lunds universitet, Lund (2016)

Hey, A.J.G., Tansley, S., Tolle, K.M.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond (2009)

Jubb, M.: Introduction: scholarly communications – disruptions in a complex ecology. In: Judd, M., Shorley, D. (eds.) The Future of Scholarly Communication. Facet Publishing, London (2013)

Justesen, L., Mik-Meyer, N.: Qualitative research methods in organisation studies. Hans Reitzels Forlag, København (2012)
Kim, Y., Stanton, J.M.: Research Article. J. Assoc. Inf. Sci. Technol. **67**, 776–799 (2016)
Macnaghten, P., Myers, G.: Focus groups. In: Seale, C., Gobo, G., Gubrium, J.F., Silverman, A.D. (eds.) Qualitative Research Practice. Sage, London (2007)
Mauthner, N.S., Parry, O.: Open access digital data sharing: principles, policies and practices. Soc. Epistemol. **27**, 47–67 (2013)
Nicolini, D., Gherardi, S., Yanow, D.: Knowing in organizations: a practice-based approach. M.E. Sharpe, Armonk (2003)
Nordforsk. Open Access to Research Data: Status, Issues and Outlook, Oslo (2016)
Palmer, C.L., Cragin, M.H.: Scholarship and disciplinary practices. Ann. Rev. Inf. Sci. Technol. **42**, 165–212 (2008)
Schatzki, T.R., Knorr-Cetina, K., Savigny, E.V.: The practice turn in contemporary theory. Routledge, New York (2001)
Royal Society: Science as an open enterprise. The Royal Society Science Policy Center (2012)
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M.: Data sharing by scientists: practices and perceptions. PLoS One **6**, e21101 (2011)