



The connection of open science practices and the methodological approach of researchers

Isabel Steinhardt¹ · Mareike Bauer² · Hannes Wünsche² · Sonja Schimmler²

Accepted: 1 September 2022 / Published online: 13 September 2022
© The Author(s) 2022

Abstract

The Open Science movement is gaining tremendous popularity and tries to initiate changes in science, for example the sharing and reuse of data. The new requirements that come with Open Science poses researchers with several challenges. While most of these challenges have already been addressed in several studies, little attention has been paid so far to the underlying Open Science practices (OSP). An exploratory study was conducted focusing on the OSP relating to sharing and using data. 13 researchers from the Weizenbaum Institute were interviewed. The Weizenbaum Institute is an interdisciplinary research institute in Germany that was founded in 2017. To reconstruct OSP a grounded theory methodology (Strauss in *Qualitative Analysis for Social Scientists*, Cambridge University Press, Cambridge, 1987) was used and classified OSP into open production, open distribution and open consumption (Smith in *Openness as social praxis*. First Monday, 2017). The research shows that apart from the disciplinary background and research environment, the methodological approach and the type of research data play a major role in the context of OSP. The interviewees' self-attributions related to the types of data they work with: qualitative, quantitative, social media and source code. With regard to the methodological approach and type of data, it was uncovered that uncertainties and missing knowledge, data protection, competitive disadvantages, vulnerability and costs are the main reasons for the lack of openness. The analyses further revealed that knowledge and established data infrastructures as well as competitive advantages act as drivers for openness. Because of the link between research data and OSP, the authors of this paper argue that in order to promote OSP, the methodological approach and the type of research data must also be considered.

Keywords Open science · Open science practice · Research data · Research data infrastructure

✉ Isabel Steinhardt
isabel.steinhardt@uni-paderborn.de

Extended author information available on the last page of the article

1 Introduction

Open Science (OS) is seen as the new paradigm that will change science fundamentally (Breznau 2021; Grand 2015; Mosconi et al. 2019). It “advocates for more public and accessible science, and has progressively encompassed new researchers’ practices and identities that go beyond the idea of digital science towards open and social activities” (Raffaghelli and Manca 2019). With this paradigm shift, scientific practices also need to change accordingly to fulfill the requirements that come with OS. Open Science Practices (OSP) can be defined as practices improving “the openness, integrity and reproducibility of research by preventing research misconduct or reducing questionable research and/or reporting practices.” (Banks et al. 2019, p. 258). This includes practices such as making research data freely accessible or pre-registering study designs prior to data collection (Banks et al. 2019, p. 257). Third-party funders are also increasingly demanding the publication of data for reuse, such as the EU in the Horizon 2020 funding line. Due to the new requirements that come along with Open Science and especially with Open Data, researchers are often confronted with different challenges, for example, the protection of personal rights and anonymisation issues, licensing rights and the long-term archiving of data. If these and other challenges are not clarified, Open Data is not per se the best option.

In the opening and re-use of data, the question of OS currently culminates, as data is the nucleus of research. Accordingly, studies quantify how many researchers share their data (Banks et al. 2019), or investigate what are the reasons for sharing or not sharing data (Houtkoop et al. 2018). The observation is that commitments for data sharing are high but few researchers share their data on request (Stieglitz et al. 2020). When asked about barriers to data sharing, 61% of the researchers responded that it is not a common practice in their field (Houtkoop et al. 2018). Researchers, despite their knowledge of OSP, are often unsure about the implementation of these practices (Banks et al. 2019; Levin et al. 2016) especially with regard to data sharing. Studies show that the degree of adaptation of the OSP of data sharing is dependent on disciplinary epistemologies, ontologies, theories and methods (Mosconi et al. 2019; Pampel and Dallmeier-Tiessen 2014; Pearce et al. 2010).

The interviews were originally conducted with the goal of gathering requirements from Weizenbaum Institute scientists in order to create a data infrastructure. However, the interviews were so comprehensive that the authors decided to use grounded theory methodology (Strauss 1987) to analyse the interviews in more depth. Grounded theory is particularly suitable when deeper layers of meaning are to be reconstructed, as is the case with practices. We initially approached the interviews openly in order to extract as much information as possible from the data. In the ongoing coding process we focused on the research practices connected with data and OSP of data use and sharing. We reconstructed the strong connection between practices and the methodological approach of researchers. The methodological approach is socialised and incorporated (Bordieu 1977, 1984) during studies and research projects and depends on the research question as well as the data. For the final coding (selected coding), we used the classification open production, open distribution and open consumption—by Smith and Seward (Smith and Seward 2017).

Our study allows us to show the researchers different practices in relation to data use and sharing and how those practices relate to their methodological approach. The practices are closely related to the methodological approach and less to the discipline. Accordingly, the challenges identified in the study are also related to methodological issues and to the incorporation of methodological practices rather than, for example, facilities provided by the institution, such as a data infrastructure. We show that practices are stable even in

interdisciplinary institutions that endorse Open Science. Also, our study has an exploratory character, our results indicate that OSP can only be implemented over a long period of time and measures for promoting Open Science should take into account the specifics based on the methodological approach.

Our results show these methodological specifics in detail together with the reconstructed practices and the challenges that the researchers face. In the next section, we explain the theory of practice that underlay our research along with the methods used. Next, we discuss our main results followed by conclusion.

2 Theory of practices and open science practices

Openness has established itself as a keyword in the demand for more transparency and reproducibility of research. Yet there is little consensus on the definition of openness and how to practice it in science (Levin et al. 2016, p. 126). As Bowman and Keene (2018, p. 364) state: “(...) this open approach to science moves us from a ‘trust me, I’m a scientist’ to a ‘here, let me show you’ position with regards to our joint and individual empirical efforts.” Accordingly, other authors theoretically describe OSP (Banks et al. 2019), by discussing the challenges and advantages they pose for the researchers (Allen and Mehler 2019; Pampel and Dallmeier-Tiessen 2014). However, few articles investigate the actual practices employed by researchers and their incorporated (embedded) values, norms and patterns of action (Levin et al. 2016). The majority of these articles focus on one discipline (Levin et al. 2016) or pre-OS practices in interdisciplinary contexts (Mosconi et al. 2019).

Although digital technologies and their use is omnipresent (Grand et al. 2016), research practices only change slowly (Costa 2014; Hagger 2022). Additionally as Pearce et al. (2010) explored, there is no “homogeneous form of ‘scholarship’ within academia” (Pearce et al. 2010, p. 37). Scientific practices are connected to disciplinary cultures (Becher 1981) as well as academic tribes and territories (Becher and Trowler 2001) and their incorporated (embedded) values, norms and patterns of action, thought and perception inherited in one’s own social context (Bourdieu 1977, 1984). Schatzki (1996) defines practices as “temporally unfolding and spatially dispersed nexus of doings and sayings”. The awareness of practices is a part of the process of socialisation into a disciplinary and/or academic culture, when individuals incorporate the typical social structures and practices of the disciplines (Schneijderberg 2018). Practices are embedded in structures, such as institutional frameworks and institutional cultures. If these structures change to more openness, the practices of individuals who work within these structures can be changed by repetition (Schäfer 2016). Accordingly, if the practices of the members of the institution are to change, the institution has a key role to play. Interdisciplinary institutes—such as the Weizenbaum Institute—face the challenge of combining disciplinary socialisation (outside the institute) with its own OS guidelines. At the same time, an interdisciplinary context can have positive effects on scientific practices. Mosconi et al. (2019) argue: “an examination of an environment where researchers come from a variety of different disciplinary origins, have heterogeneous knowledge, skills, and have different mundane practices in respect of choices about how to organize, store and represent data, ought to be fruitful.”

In order to answer the question about the practices of the researchers at the Weizenbaum Institute, we conducted a grounded theory study in which we did a secondary analysis of interviews. A detailed insight into the methodological procedure is given in the next section.

3 Sample and method

The Weizenbaum Institute founded in 2017 makes a good case for analysing OSP because of its structure and mission. By investigating the official resources of the Weizenbaum Institute, such as its mission statement and annual reports, we find that there is a great support for OS. Open-mindedness and participation are two of the five principles of work in the mission statement. The statement names Open Science as well as Open Access, Open Data, Open Source and Citizen Science explicitly as targeted practices.

The Weizenbaum Institute was formed in 2017 by five universities—Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin, Universität der Künste and the Universität Potsdam—as well as two non-university research institutes—Fraunhofer FOKUS and Berlin Social Science Center (WZB)—in the Berlin area. The Weizenbaum Institute includes twenty research groups, all concerned with research on the impact of digital transformation on modern societies. The research groups all consist of interdisciplinary teams and include different methodological approaches.

Additionally, the Weizenbaum Institute will soon provide its own repository for publications and research data, the Weizenbaum Library, to open up the institute's research. This step is necessary since it is one crucial factor for openness in science, as Levin et al. (2016) show in their study. Furthermore, "Perhaps the greatest bilateral technical barrier facing both the providers and users of open science practices is the lack of a supporting infrastructure." (Banks et al. 2019, p. 265). The statement is generally agreed on, yet the difference between provided technologies and infrastructure should be noted. Infrastructures are defined by their entanglement with established practices. It is through the embeddedness of technologies into practice that they are infra—unseen—structures, consequently, becoming visible upon breakdown (Star and Bowker 2002). OSPs therefore need supporting technologies to evolve into infrastructures through established use. For example, a data repository that is hardly used, either for depositing or downloading, will not be missed if offline. The lack of integration in existing practices would prevent it from becoming part of a supporting structure.

The set up of the Weizenbaum Library was the starting point of our study. We wanted to involve researchers and members of the institute right from the start in the development of the research data infrastructure. In order to do so we conducted interviews with members of the Weizenbaum Institute regarding the implementation of the Weizenbaum Library, a repository for research data and publications. We invited all members of the institute to participate in interviews. The interviews were intended to gather requirements for a research data infrastructure at the Weizenbaum Institute, and implicitly to identify existing practices. In total, 13 members of the Weizenbaum Institute participated (approximately 10% of the scientific staff) from 12 of the 19 research groups, who are not involved in the implementation of the infrastructure. The interviews were conducted between March 2019 and May 2019. When looking at the self-selected sample, the positions of the interviewees range from student researchers to research group leaders. As an interdisciplinary institute, researchers with different disciplinary backgrounds took part in the interviews, including sociology, political science, computer science, communication science, psychology, philosophy, media studies or combinations of them. In addition to the disciplinary diversity, different data types used became apparent (see next section).

Since the interviews were primarily intended to determine the requirements for creating a research data infrastructure, various aspects of practices came to the fore during the interviews and manifested in the coding process. The interview guideline contained

18 pre-formulated questions clustered into different topics. The first set of topics dealt with the idealistic motivation of the interviewees and included questions such as “Which of these data should be published in general?” and “What would motivate you to provide data to the public apart from your publications?”. The next four questions dealt with the attitudes toward and use of research data platforms in general and to a research data platform at the Weizenbaum Institute. In addition, another two sets of topics (each containing three questions) enquired about the types of data the researchers are working with. At the end of each interview, the participants were given the opportunity to add further comments. Instead of making concrete demands on the research data platform, the participants reported on practices of scientific work. These included, for example, practices of research, scientific exchange or data preparation. As Maiwald (2003) and Leger et al. (2018) indicate, our study confirms that practices can also be reconstructed from interviews. We therefore derived concrete requirements from the practices described. Instead of just identifying the requirements, we wanted to analyse the described practices and the relationship between OSP, the methodological approach and used data within this article.

To analyse the interviews we used Strauss (1987) Grounded Theory, in which the use of prior theoretical knowledge is explicitly allowed. Strauss (1987) assumes that researchers have prior theoretical knowledge that cannot be abandon. All four authors contributed to the development of coding steps. Due the extensive back and forth communication between the authors, the coding process was time consuming. However, this method and the discussions enabled a deep engagement with the collected data. We used the software MAXQDA to analyse the interviews. In the first step, the open coding, we reviewed the data to find practices described by the interviewees. We coded the interview passages and combined them into concepts. In the second step, axial coding, we developed categories from the concepts. One identified category contains practices regarding the research process, meaning that participants described how they search for literature or data. Practices concerning the provision of data could be identified as well as practices regarding the administrative part of science. Another set of practices were clustered under the term “interaction”. This category describes practices of cooperation and collaboration. The practices clustered under the category “interfaces” describe how the participants use scientific tools and existing platforms such as ResearchGate.

We recombined these categories in the third step, the selected coding. For selected coding, we used the classification of OS practices—open production, open distribution and open consumption—by Smith and Seward (2017). Smith and Seward (2017) have studied practices of openness as a social praxis and have divided it into the three social processes: open production, open distribution and open consumption (Smith and Seward, 2017, p. 5). These open processes have four main elements: the open practices, the structuring characteristics of these practices, the context of the open process and finally the consumption or the outcome of the open process (Smith and Seward 2017, p. 8f). On this basis they developed a typology of openness as a social praxis (see Table 1).

Table 1 Three main types of open processes. Source: (Smith and Seward 2017, p. 8)

Open process	Open practice
Open production	Peer production, crowdsourcing
Open distribution	Sharing, republishing
Open consumption	Retain, (re)use, revise, remix

During the selected coding, the strong connection between practices and methodological approach and data types appeared, which we examine in more detail below. For an even deeper insight, we have made the anonymized interview transcripts and a detailed description of how the interviews were conducted available open access (Bauer and Wünsche 2022).

4 Results

The first result was a basic commitment to OS at the level of talk. The main narrative was sharing own data and/or accessing other data as an ethical commitment to society. If society finances the research, the data should also be publicly available: *“I find this connection very important—in fact, public research with public funds must also be publicly accessible.”* (interview 1, l.24–27, own translation) With regard to the level of practices, great differences and only a few OSP could be found. This confirms the findings of Vanpaemel et al. (2015) that the normative understanding of science (support of OS) and the performed scientific practice differ. Thus, an open understanding of science can exist without sharing own data and/or accessing other data (Tenopir et al. 2011).

As we mentioned above, we applied the three types of OSP by Smith and Seward (2017), which include here open production, open distribution and open consumption, to analyse the practices of our interviews. Open production enables voluntary and nondiscriminatory participation in research processes. Open distribution is based on free and non-discriminatory sharing of digital content. Open consumption provides the possibility of free use, as for instance: retaining, reusing, revising, remixing, and redistributing of shared materials (Wiley 2014).

Overall, we did not find open production practices in our interviews. We found a form of informal distribution of data at the Weizenbaum Institute between research groups *“We also have a small cooperation with another group at the Weizenbaum Institute and we use their data and they use our data. It was kind of difficult to share our and use their data.”* (interview 10, l.29–31, own translation) This form of distribution can be viewed as informal OSPs that do not require institutional data infrastructures. Data is shared via mobile data carriers or sent via email. In the sense of Smith and Seward (2017), this form of distribution is not an OSP, since one of the basic requirements of their definition is the free and non-discriminatory sharing of data.

For open distribution and open consumption, we found a close link between practices and self-understanding/self-construction of methodological identities. These referred to distinctions between qualitative and quantitative research and data, which is often ambiguous and is discussed controversially (Allwood 2012; Morgan 2018). In addition, two new types of data are viewed separately, source code and social media data, which cannot be clearly assigned to an epistemology. In the following, we will therefore speak of research approaches. In this way, we take into account the participants' self-attribution to one research approach and in addition it is possible to reconstruct four types of researchers, who create their epistemology on the basis of the data types they use. To avoid redundancies in describing, the practices we present open distribution and open consumption together. Accordingly, we have formed four types: qualitative approach, quantitative approach, source code and social media data.

4.1 Qualitative approach: researchers who do not reuse and share data

Researchers in our interviews that describe themselves as qualitative researchers mention the following data types: documents, interviews, video recordings and transcribed audios. The methods they use for analysing data are, for example, qualitative content analysis or discourse analysis. We reconstructed that qualitative researchers do not draw on existing repositories for qualitative data in their research practices. Furthermore, we found no practices of sharing and/or reusing of data. We identified three reasons: uncertainties and missing knowledge, data protection and competitive disadvantages.

4.1.1 Uncertainties and missing knowledge

The missing knowledge relates to the practices of data collection: For interviews and observations, subjects must give their consent to data collection and processing. For qualitative research in Germany, it is established practice to obtain consent only for one's own research project, but not for the subsequent use of the data (Steinhardt 2018). One reason for this is the assumption, which has not been investigated so far, that consent is more likely to be refused when participants are asked for subsequent use of their data. On the other hand, most researchers do not know how qualitative data can be sufficiently anonymised and contextualised to still be of interest for other researchers.

This led to the situation that the qualitative researchers in our sample described practices of non-sharing, referring to "purpose limitation": Sharing of qualitative data is not possible due to data protection and anonymisation regulations, installed in the consent agreement. *"we are not allowed to publish. (...) we are also not allowed to publish, because we also have a purpose limitation."* (interview 3, l.218–219, own translation). Consequently, in the description of their own practices, it became clear that qualitative researchers do not attempt to remedy the uncertainties that arise.

4.1.2 Data protection

A second motive brought forth are uncertainties that had arisen because of the European "General Data Protection Regulation (GDPR)". These uncertainties lead to the fact that sharing qualitative data is generally neglected: *"I think this is one reason why many people don't publish their stuff. All those legal uncertainties."* (interview 5, l.24–25, own translation). Besides the general lack of knowledge about consent agreements, the discussion about new enforced data protection regulation seems to foster established practices as legally safe.

4.1.3 Competitive disadvantages

The third reason why qualitative data is not shared is the fear of competitive disadvantages: *"And if two scientists have the same idea and one has good data, the other has no data, then it is not worth sharing. Of course this is not very desirable and laudable, such a behaviour, but I think it is also human, that some people get this idea."* (interview 5, l.28–31, own translation). In this context one researcher described that in their disciplinary socialisation it is advised to be especially careful with one's own data: *"then I deposit it and people now also use my data somehow to get faster or get results, so that in the end I*

have to quote other people who have taken over my work, because they were simply faster. (...) But also because I was taught to be careful with your data.” (interview 9, l.148–151, own translation).

The fear of losing advantages in the scientific competition and the fear of reputational damage, due to a generally negative attitude towards openness in the scientific community, cause rejection for data sharing. In addition, some interviewees noted: The disclosure of data also makes one’s own scientific work vulnerable and thus does not necessarily serve one’s own scientific reputation.

4.2 Quantitative approach: researchers who reuse data but do not share data

Researchers in our interviews describing themselves as quantitative researchers define quantitative data as data and data sets that are collected and evaluated using quantitative methods. This includes measurement results, statistics, survey and panel data.

Unlike qualitative researchers, quantitative researchers reuse data, mainly from large panel and survey studies. In Germany, there are large state-financed panel studies that can be used for secondary research. These data are made available through data infrastructures set up for this purpose (e.g. at GESIS). These infrastructures have existed for decades and lecturers frequently use the panel data in quantitative methodological courses, e.g., by sociologists and educational scientists. Thus, a high degree of awareness about these repositories and large-panel data sets exists and the socialisation in the use of it starts early, because lecturers use the panel data for statistical method courses.

In addition to the large panel data, the repositories also archive small studies for re-use. However, the interviewees did not seem to know other data for re-use, apart from the data of panel studies. Accordingly, the practice of searching for smaller data sets that could be of interest for one’s own research does not exist. We also could not find any practices of sharing data: *“Well, my own research there, of course, I produce quite a lot of data, but most of it is just lying around.” (interview 8, l.13–15, own translation).* Reasons why researchers with a quantitative approach do not share data are: competitive disadvantages, vulnerability and costs.

4.2.1 Competitive disadvantages

It is evident that OSP are fraught with fear of experiencing disadvantages by opening the data. In accordance with qualitative researchers, fear of competitive disadvantages is mentioned as reason for not sharing one’s own data: *“Others certainly have some strategic research concerns about publishing data, where one might want to draw results oneself, which should not be circulated before or even after, yes (...)” (interview 2, l.13–15, own translation).*

4.2.2 Vulnerability

Furthermore, the opening of the data makes one’s own research vulnerable to claims of scientific misconduct. By reproducing studies, errors can be detected and research can be improved. So far, however, there is no culture of constructive criticism, which is why opening data is associated with a risk: *“(…) that is such a part of the reproduction crisis, that as soon as one makes research data publicly available one makes oneself criticisable in a certain way, because other people can understand one’s own research, so to speak, and*

then possibly find errors." (interview 2, l.15–19, own translation). Especially the aspect of vulnerability shows that OSP is synonymous with a change in scientific culture.

4.2.3 Costs

One further reason for the refusal to share own data are the costs. For providing data, resources are needed (e.g. personnel and/or time) to manage this additional workload. According to this understanding, openness and accessibility are perceived primarily as a workload and thus as a financial burden: *"This mainly means money, because an institution or somebody has to take money in hand to achieve standardisation, in terms of human resources: scientific staff, staff positions at institutes. These are a number of questions that still go hand in hand with the 'good will' (sic) we are now putting everything on the net and everything is available for free"* (interview 8, l.21–25, own translation). Quantitative researchers also see no benefit for themselves in sharing data. Nevertheless, they point out that ideally, it should be done, but the fact that there is no benefit, gives them no reason to do so: *"For me, this is really a question of money, because I do not profit from the fact that others can work with the data. Of course I think it's great. I think it's very right in terms of scientific ethics and it meets many standards, but of course it doesn't meet our publication requirements, which is a hindrance."* (interview 8, l.67–70, own translation).

4.3 Source code: researchers who reuse and share source code

Another group of interviewees were researchers who are developing and experimenting with digital tools, in particular program source code. In contrast to qualitative and quantitative research, these researchers are organised differently. Originating in software development, reuse of existing source code and sharing of own source code is a well-known practice. They are aware of repositories for source code, which are very well integrated in their research process. While reusing source code is a lived practice among researchers, sharing of source code is seen as useful but not overly a common practice. We could identify three reasons for openness or lack of openness: competitive advantages, knowledge and established data infrastructures and costs.

4.3.1 Competitive advantages

The writing of source code is perceived as less efficient without recourse to source code from other developers. Accordingly, the use as well as the provision of source code is common practice among developers and is socialised from the very beginning. This can be seen in many practical programming classes, where students learn how to use tools such as Git.

Sharing of source code and test data is seen as specifically important, in order to trace back the research process, to enable transparency and reproducibility: *"Well, the thing is research is also not a snapshot 'I did this, now it's done and over', but there is always a traceable development through the area, which is read over and over again (...)"* (interview 7, l.139–141, own translation) In more and more sub-communities, after completion of the project, it is even expected that the project source code and test data is shared in addition to the publication itself in order to allow for a thorough comparison. *"Oh, yes, because otherwise you can't go forward. [personal information]. And as said, again, my research is useless if the others cannot relate their own research to it."* (interview 7, l.106–108, own translation).

For those who do not share their source code, there is awareness of the benefits of sharing source code. One of the researchers states: *“In the past you had to know the people and then they would finish and send it to you, at some point, and today you might want to get to the point where you can download it immediately and all that. So I would see this kind of benefit and the other way around would be for me, too, if others can use it is nice.”* (interview 13, l.170–174, own translation).

4.3.2 Knowledge and established data infrastructures

A new research project usually starts with a search for other software projects that have done similar work. One researcher described his/her research process like this: *“Because you build your—the very first thing you have to do is, you search your related work to see if people have done it or not. And people have certainly done it in one form or another and you want to compare your approach to say why you are better.”* (interview 7, l.29–31, own translation).

Data infrastructures such as GitLab or GitHub have been implemented early on, allowing for collaboration. Consequently, these repositories are standards for the community and favoured in comparison with newly established solutions: *“Why can’t I just upload my [data type 20] to [repository 2] and my paper to [tool 2] with some additional data? And then simply, why reinvent the wheel when the solution already exists?”* (interview 13, l.107–110, own translation).

One challenge that arises is that these tools were intended for collaboration rather than for archiving. Consequently, they were designed differently, and features such as persistent identifiers and long-term preservation are neglected in these systems. Another challenge is that related research data, such as publications, models or test data, are published in other repositories and therefore links between these digital objects are not persistent or non-existent. *“So now there is no need (incomprehensible) to collect or be on the same platform but that somehow associations are put together and that it is searchable and findable in the long run. That is the problem I have.”* (interview 7, l.23–26, own translation).

4.3.3 Costs

Nevertheless, the reuse of source code and the combination of different code snippets is often a challenge. The source code might be, for instance, outdated, of low quality, or not well documented. *“Well, in [discipline 2] there is so much sharing and so on anyway, otherwise it is not possible. And then there is the [datatype 16], which I just told you about, on top of it. Then you put another [data type 16] on top of it, so everything is somehow [repository 2], [repository 2], [repository 2].”* (interview 13, l.58–61, own translation). This means that sharing source code is time consuming, especially when some effort is put into preparing it well. *“But this takes work”* (interview 13, l.174, own translation).

4.4 Social media data: researchers who reuse social media data but do not share data

Another group of interviewees are researchers that are exploring social media data. In our interviews, the used social media data mentioned mostly comes from Twitter, as it is the only social media platform that shares data openly. For these researchers, using data is a

prerequisite, while sharing data is underdeveloped mainly due to legal concerns. We could identify one reason for lack of openness: uncertainties and missing knowledge.

4.4.1 Uncertainties and missing knowledge

Due to legal requirements of social media companies, researchers do use social media data, but do not pass on their data to third parties or make it available for subsequent use: “*But I don’t know where and can I publish data interactions for example from (incomprehensible) for I know they have some requirements and probably if I use some data, you know, then I am probably not allowed to publish it somewhere else. And it is always very difficult and finally you don’t publish anything because you have no time to figure out how it works.*” (interview 10, l.2125). The interviews show that there is a will to share, but the legal requirements are so unclear that they shy away from sharing the data: “*I am just not sure about publishing bigger data in all that because I am not sure that we are allowed to do this. So, if I am sure that I am allowed to use it and I don’t violate any API requirements (...) then yes.*” (interview 10, l.103–105).

5 Discussion

With regard to methodological approach and type of data, uncertainties and missing knowledge, data protection, competitive disadvantages, vulnerability were uncovered as well as costs as reasons for the lack of openness. The study further uncovered knowledge and established data infrastructures as well as competitive advantages as drivers for openness:

Uncertainties and Missing Knowledge: Uncertainties in implementing OSP arise for both qualitative researchers and researchers working with social media data in terms of how data can be shared, opened, and (re)used. These uncertainties result not just from the ambiguities regarding the meaning and practice of openness (Levin et al. 2016, p. 126), but as well from a lack of knowledge regarding the legal framework. *Data Protection:* Concerns about data protection were expressed by those who use social media data and interviewees who describe themselves as qualitative researchers. These included insecurities regarding the opening and sharing of one’s own research data as well as the use of other data. In case of qualitative data, concerns regarding the protection of the collected data and the anonymisation of data were especially present. When using social media data there are mainly legal concerns regarding the distribution of the used data. This draws on the fact that the legal requirements of social media companies are very opaque. As already mentioned, uncertainties regarding the anonymisation of data were especially present among qualitative researchers. However, these uncertainties result not only from the legal framework, but also from a lack of knowledge when it comes to practices of sharing. When it comes to open consumption of data, qualitative researchers show little knowledge of existing repositories and databases. Therefore, the consumption of data is limited. Quantitative researchers also do show knowledge gaps regarding repositories and databases. A limited amount of large panel data is known whereas smaller repositories and databases are not part of the research process. *Competitive Disadvantages:* However, another concern regarding the practices of open production and distribution mentioned by researchers with qualitative or quantitative approaches can be described as competitive disadvantages. Although the researchers recognise the values and norms of OS in theory, they themselves do not see OS as an enrichment in their own research. Rather the fear of a loss of advantages is mentioned

when it comes to open production and open distribution. *Vulnerability*: At the same time, there is a concern among researchers with a quantitative approach to make themselves vulnerable by disclosing their own data and to become subject to unwanted evaluation of their research. *Costs*: Additionally, for researchers with a quantitative approach and researchers who work with source code, openness is perceived as an extra workload. Despite recognising the value of OS in theory, the implementation of OS in practice is described as a time consuming and financial burden and therefore not feasible in practice. *Knowledge and Established Data Infrastructures*: Among researchers working with source code, the implementation of OSP is more widespread. For them, open practices are not only normatively anchored, but part of their scientific socialisation. *Competitive Advantages*: The interviewees that work with source code did neither express fear of competition nor loss of advantages. Rather open distribution and open consumption are seen as important regarding the traceability of research. Furthermore, they see research as a social process that can only be realised through (data) exchange with others.

Nevertheless, with regard to open practices, our research shows that the type of research data takes a central role and not disciplinary affiliations. Open practices such as sharing and opening data are not practiced among researchers with a quantitative or qualitative approach and are therefore associated with insecurities. Due to the institutional context, all of the interviewees expressed a positive attitude towards OS. However, this positive attitude has not yet been reflected in the practice of the researchers unless they are socialised to open practice due to the type of research data.

Smith and Seward (2017) point out that the type of content, the technological infrastructure and contextual factors, such as individual skill sets and access to technology, are crucial for the implementation of open distribution and open consumption of data (Smith and Seward 2017, p. 23). Our research implies that insecurities regarding the legal framework especially in terms of data protection, missing knowledge regarding the anonymisation and contextualisation of data, personnel and financial resources, the fear of unwanted quality control, a loss of advantages and vulnerability as well as a missing recognition of OS as an enrichment for one's own research also prevent OSP.

6 Conclusion

For more and more research institutes, OS and infrastructures supporting openness of research data and materials currently are discussed topics. Nevertheless, institutional change is slow and little is known of existing obstacles preventing OS to become an everyday practice. As a recently founded institute, which investigates the digital transformation, the Weizenbaum Institute is strongly committed to OS, not just through research on openness in various areas of society but also through facilitate openness of its own research. To see how OS is perceived and enacted in this open-positive research environment interviews with 13 researchers of the Weizenbaum Institute on data sharing and related digital infrastructures in science were conducted.

As previous research suggests, the socialisation of researchers plays a decisive role closely linked to the disciplinary and institutional context (Becher 1981; Becher and Trowler 2001). But as the results of the interviews show, the socialisation of researchers is influenced not only by the discipline and the research environment, but also by the methodological approach and the type of data. As shown, open practices are implemented differently and the barriers vary depending on these categories. Problems and

concerns in transiting to OSP are mainly linked to the type of data produced and consumed by the researchers. Within socialisation processes, individuals incorporate the prevailing social structures and practices (Schneijderberg 2018).

Although all of the interviewed researchers described a positive attitude towards OS, the practice of open production, as defined by Smith and Seward (2017), is not part of the research process at the Weizenbaum Institute. The practice of open consumption was implemented by researchers who work with source code and partly by quantitative researchers and researchers that work with social media data. Open distribution was only partially implemented by researchers that work with source code. As open distribution and consumption are a central part of the research process for source code researchers the implementation of these practices was partially observed. A partial implementation of open consumption was observed among quantitative researchers due to their socialisation and among social media researchers due to the specificity of the data type.

The obstacles linked to the different data types vary. Researchers working with qualitative or quantitative data both formulate concerns regarding competitive disadvantages, revealing competitive practices as a counterpart to open practices. Furthermore, data protection obligations as well as uncertainties and missing knowledge regarding the anonymisation and contextualisation of data were brought forth in the context of qualitative data, production cost and vulnerability as a hindering factor for sharing quantitative data. Researchers working with social media data expressed concerns about the illegality of sharing data, therefore taking distance to provide their data to others.

In reconstruction of existing research practices and their relation to OSP, this paper's research shows that the provision of technologies and infrastructures for OS is not sufficient. Providing tools for OS is a prerequisite but does not result in open practices. In order to change existing practices and promote OSPs, it takes time and an active accompaniment of the process. As a starting point, this requires the provision of financial resources that go beyond providing technical infrastructures. To successfully implement open practices, it is also necessary to make available specific instructions and best practice examples.

The described uncertainties and the existing lack of knowledge could be reduced by advisors regarding the research data infrastructure and the legal concerns, as one of the interviewees suggested: *"So, of course I have all the legal questions, what I am allowed to publish and what not. And then, I don't know if the platform will contain that, but an advice center or contact person would be good, who can help ((laughs))"* (interview 5, 1.62–65, own translation). However, the specific considerations that depend on the research approach must be taken into account. Consequently, not just one contact person, but one contact person for each approach may be required. As practices change slowly the implementing open practices could be expedited through these contact persons and measures such as workshops and training courses. Especially for interdisciplinary institutes, such as the Weizenbaum Institute, where different scientific socialisations and research approaches meet, the active and accompanying support for this process seems useful in order to promote the general facilitation of OSPs.

In addition, role models exemplifying open practices could promote the socialisation process in regards to OS. As open practices are social practices (Smith and Seward 2017, p. 3) the institutional normative orientation towards OS is not sufficient. In order to implement open practices they must be exemplified. Accordingly, role models are needed which not only commit themselves to OS normatively, but also through their scientific practice. Since most of the researchers at the Weizenbaum Institute are at the beginning of their scientific career, thus scientists who are in the process of scientific socialisation, it is even

more important to exemplify open practices. Only if open practices become part of the scientific socialisation will a process a long term change towards OS be possible.

Future research on the acceptance and implementation of open science practices should therefore not only focus on the technical requirements but also consider the scientific socialisation of researchers—that is, not only the disciplinary context but also the methodological approaches and data types used. Since this study focused on one specific scientific institution, future research should examine OSP in terms of methods and data types across different institutions to gain further insights into how open science practices can be implemented in research processes.

Funding Open Access funding enabled and organized by Projekt DEAL. This work has been partially funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DI117 (Deutsches Internet-Institut).

Availability of data and material (data transparency) Anonymized interview transcripts, the interview guidelines, the declaration of consent, and a detailed data report are published at Zenodo: <https://doi.org/10.5281/zenodo.5971467>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Consent to participate All interview participants signed a declaration of consent.

Consent for publication See consent to participate.

Ethics approval See consent to participate.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References





- Allen, C., Mehler, D.: Open science challenges, benefits and tips in early career and beyond. *PLoS Biol.* **17**(5), 1–14 (2019). <https://doi.org/10.1371/journal.pbio.3000246>
- Allwood, C.M.: The distinction between qualitative and quantitative research methods is problematic. *Qual. Quant.* **46**(5), 1417–1429 (2012). <https://doi.org/10.1007/s11135-011-9455-8>
- Banks, G.C., Field, J.G., O'Boyle, H., Landis, R.S., Rupp, D.E., Rogelberg, S.G.: Answers to 18 questions about open science practices. *J. Bus. Psychol.* **34**(3), 257–270 (2019). <https://doi.org/10.1007/s10869-018-9547-8>
- Bauer, M., Wünsche, H.: Interviews zu Forschungsdateninfrastrukturen und digitalen Praktiken offener Wissenschaft am Weizenbaum-Institut (1.0) [Data set]. Zenodo (2022). <https://doi.org/10.5281/zenodo.5971468>
- Becher, T.: Towards a definition of disciplinary cultures. *Stud. High. Educ.* **6**(2), 109–122 (1981). <https://doi.org/10.1080/03075078112331379362>

- Becher, T., Trowler, P.: *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. Open university Press (2001)
- Bourdieu, P.: *Outline of a Theory of Practice*. Cambridge Studies in Social and Cultural Anthropology, Cambridge University Press (1977) <https://doi.org/10.1017/CBO9780511812507>
- Bourdieu, P.: *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press (1984)
- Bowman, N.D., Keene, J.R.: A layered framework for considering open science practices. *Commun. Res. Rep.* **35**(4), 363–372 (2018). <https://doi.org/10.1080/08824096.2018.1513273>
- Brezna, N.: Does sociology need open science? *Societies* (2021). <https://doi.org/10.3390/soc11010009>
- Costa, C.: The habitus of digital scholars. *Res. Learn. Technol.* (2014). <https://doi.org/10.3402/rlt.v2i10.21274>
- Grand, A.: Open science. *J. Sci. Commun.* (2015). <https://doi.org/10.22323/2.14040302>
- Grand, A., Holliman, R., Collins, T., Adams, A.: We muddle our way through : shared and distributed expertise in digital engagement with research. *J. Sci. Commun.* (2016). <https://doi.org/10.22323/2.15040205>
- Hagger, M.S.: Developing an open science 'mindset'. *Health Psychol. Behav. Med.* **10**(1), 1–21 (2022). <https://doi.org/10.1080/21642850.2021.2012474>
- Houtkoop, B.L., Chambers, C., Macleod, M., Bishop, D.V.M., Nichols, T.E., Wagenmakers, E.: Data sharing in psychology: a survey on barriers and preconditions. *Adv. Methods Pract. Psychol. Sci.* **1**(1), 70–85 (2018). <https://doi.org/10.1177/2F2515245917751886>
- Leger, M., Panzitta, S., Tiede, M.: Daten-Teilen? Digitale Selbstvermessung aus praxeologischer Perspektive, transcript-Verlag, pp. 35–60. (2018) <https://doi.org/10.14361/9783839439579-002>
- Levin, N., Leonelli, S., Weckowska, D., Castle, D., Dupr'e J.: How do scientists define openness? exploring the relationship between open science policies and research practice. *Bull. Sci. Technol. Soc.* **36**(2), 128–141 (2016). <https://doi.org/10.1177/0270467616668760>
- Maiwald, K.: Methodenwerkstatt stellen Interviews eine geeignete Datenbasis für die Analyse beruflicher Praxis dar?: Methodologische Überlegungen und eine exemplarische Analyse aus dem Bereich der Familienmediation. *Sozialer Sinn* **4**(1), 151–180 (2003). <https://doi.org/10.1515/sosi-2003-0110>
- Morgan, D.L.: Living within blurry boundaries: The value of distinguishing between qualitative and quantitative research. *J. Mixed Methods Res.* **12**(3), 268–279 (2018). <https://doi.org/10.1177/1558689816686433>
- Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., Korn, M., Pipek, V.: Three gaps in opening science. *Comput. Supported Cooper. Work (CSCW)* **28**(3–4), 749–789 (2019). <https://doi.org/10.1007/s10606-019-09354-z>
- Pampel, H., Dallmeier-Tiessen, S.: *Open Research Data: From Vision to Practice*, Springer, pp. 213–24 (2014) https://doi.org/10.1007/978-3-319-00026-8_14
- Pearce, N., Weller, M., Scanlon, E., Kinsley, S.: Digital scholarship considered: How new technologies could transform academic work. In *Education* (2010) <https://ineducation.ca/ineducation/article/view/44>
- Raffaghelli, J., Manca, S.: Is there a social life in open data? the case of open data practices in educational technology research. *Publications* **7**(1), 9 (2019). <https://doi.org/10.3390/publications7010009>
- Schatzki, T.: *Social Practices - A Wittgensteinian Approach to Human Activity and the Social*. Cambridge University Press (1996)
- Schneijderberg, C.: *Promovieren in den Sozialwissenschaften. Eine sozialisationstheoretische Erschließung des Forschungsfeldes Promotion*. Springer VS (2018) <https://doi.org/10.1007/978-3-658-19739-1>
- Schäfer, H.: *Praxistheorie: Ein soziologisches Forschungsprogramm*. Transcript-Verlag (2016) <https://doi.org/10.14361/9783839424049>
- Smith, M.L., Seward, R.: Openness as social praxis. *First Monday* (2017). <https://doi.org/10.5210/fm.v22i4.7073>
- Star, S.L., Bowker, G.C.: *How to Infrastructure*, SAGE Publications, pp. 151–162 (2002) <https://doi.org/10.4135/9781848608245.n12>
- Steinhardt, I.: Open Science-Forschung und qualitative Methoden – fünf Ebenen der Reflexion. *Medien-Pädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung* **32**, 122–138 (2018). <https://doi.org/10.21240/mpaed/32/2018.10.28.X>
- Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., López, A., Rehwal, S.: When are researchers willing to share their data? – impacts of values and uncertainty on open data in academia. *PLoS ONE* (2020). <https://doi.org/10.1371/journal.pone.0234172>
- Strauss, A.L.: *Qualitative Analysis for Social Scientists*. Cambridge University Press (1987) <https://doi.org/10.1017/CBO9780511557842>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M., Frame, M.: Data sharing by scientists: Practices and perceptions. *PLoS ONE* (2011). <https://doi.org/10.1371/journal.pone.0021101>

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., Storms, G.: Are we wasting a good crisis? the availability of psychological research data after the storm. *Collabra* (2015). <https://doi.org/10.1525/collabra.13>
Wiley, D.: Open Content. (2014) <https://openedreader.org/chapter/opencontent/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Isabel Steinhardt¹  · Mareike Bauer²  · Hannes Wünsche²  · Sonja Schimmler² 

Mareike Bauer
mareike.bauer@fokus.fraunhofer.de

Hannes Wünsche
hannes.wuensche@fokus.fraunhofer.de

Sonja Schimmler
sonja.schimmler@fokus.fraunhofer.de

¹ University of Paderborn, Paderborn, Germany

² Fraunhofer Institute for Open Communication Systems (FOKUS), Berlin, Germany