

The intertwining of reputation and sharing – The significance of standardization in preparing research data and the impact of project organization

Saskia-Rabea Schrade¹

Leibniz Universität Hannover, Leibniz Center for Science and Society, 30156 Hanover, Germany

Abstract. Despite efforts to increase scientists' willingness to share research data political stakeholders and funding agencies, there is still a discrepancy between scientists' attitude toward data sharing and their actual practice. In a first step, this paper takes a close look at scientists' definition of research data and the influence of project organization on scientist's willingness to share data by analyzing interviews with scientists of three different disciplines. As the analysis shows, talking about "data sharing" should always happen in the context of data preparation and its various steps. Additionally, the influence of external factors such as a special form of project organization seems to be limited.

1 Introduction

Over the past few decades, the sharing of research data has become an increasingly common and important practice and requirement in scientific research. One factor of this development is, for example, the demand for data management as a requirement for funding or publication by funding agencies as well as publishers. Because public funding is one of the most important financial sources for scientists, it is common sense that at least the results of scientific research should be available to the public. However, the demand for publicly available research data and an increase in sharing research data is also on the rise. The importance of data sharing and openness with the outcome of data as a common good, at least within the scientific community, has long been known. In 1973, this was pointed out by Robert K. Merton, who introduced the norm of communism of the scientific ethos, which says that scientific goods should be of common ownership within the scientific community [1].²

Data sharing is a frequently investigated and examined topic, regardless of disciplinary borders. It poses many advantages, such as a decreased risk of repeating failed research and making replication studies easier [2, 3]. Thus, transparency and accuracy of research are

¹ saskia-rabea.schrade@lcss.uni-hannover.de

² First, Merton invented four norms: "communism," "universalism," "disinterestedness," and "organized skepticism" [1]. Those were later reviewed and expanded with a fifth norm, "originality," and "organized skepticism" was changed to "skepticism" [4].

improved. It also provides huge potential for further scientific research, as scientists are enabled to combine different datasets or look at shared research data from a different scientific perspective [5-7]. On the other hand, there are also impediments, barriers, and disadvantages anticipated with sharing research data. An unsatisfactory infrastructure (e.g., no hardware or software available for storing research data or a limited funding period for those tools) for data sharing is one of the main barriers [8, 9]. Scientists' critique also focuses on the consequences of sharing research data: an expected loss of possibilities to publish research results and, thereby, the loss of reputation. This problem is admittedly not novel to science and the scientific community. One of the five norms invented by Merton is the norm of originality: He stated that for the acknowledgement of scientific reputation, the priority of discovery and scientific knowledge that is considered original is imperative [4]. Current research shows that scientists' willingness to share data with other scientists has increased in recent decades; nevertheless, many scientists are still very hesitant, even though political stakeholders and funding agencies are eager to increase their willingness to share or make research data available using various approaches [10, 11]. To take a closer look at this discrepancy, this study starts at the beginning of data sharing: a definition of the term "data", based on scientists' descriptions and taking project organization into account.

2 Background

Research data is most important in the scientific field because scientists' funding depends significantly on their scientific output. As aforementioned, producing knowledge that is considered and acknowledged as "original" is particularly significant in science because the acknowledgement of scientific originality is crucial for the acknowledgement of scientific reputation [12-16]. To produce original knowledge, scientists must perform science first and publish their results before the acknowledgement of reputation; therefore, scientists are more likely willing to share data that has already been published [17]. This results in a limited willingness to share data that has not yet been published because it will be used to produce reputation via publications, and thus, sharing data or intermediate products could lead to a competitive disadvantage [18-20]. Many studies also show that highly competitive disciplines are related to a modest willingness or a highly mandatory way of sharing data [16]. Although scientists are aware that research data is imperative to their scientific career, they are often unaware of the data's value for scientists' own discipline and other disciplines [3, 6]. An important characteristic is that data sharing is carried out at different social levels. Therefore, data can, for example, be shared in a broad field such as politics or economics, or limited to a small number of scientists, such as a certain scientific community [21]. It also varies depending on the discipline's norms on data sharing and discipline-specific attitudes adopted by scientists [8, 22]. Thus, data is shared not only with other scientists but also with members of other social groups.

Finally, the unawareness of services offered for data management and preparation and the lack of skills in those areas have also been repeatedly investigated [3, 5-7, 24]. The lack of standardization in data sharing and the complex notion of the term "data" is another subject of inquiry. It results in a diversity of data formats that are used, an often undefined idea of "data," "metadata," and their meaning, as well as different requirements regarding what data should be shared and when it would end with an overload of useless or meaningless data [7, 9]. A problem that stems from the organization of science itself is the (lack of) funding infrastructures for sharing, which results in an unstructured landscape of ways and places to save, store, or share data, as many studies have shown [3, 8, 9]. Although standardization seems to be a key factor in the documentation of data and data sharing, almost no standards exist. Standardization is an imperative factor in data sharing, but also to data itself. The reason for this assumption lies in the degree of standardization of methods used during the process

of data preparation. A low degree of standardization increases the value of the method, as it is then hardly possible to recreate without further information or instruction. A strong degree of standardization, on the other hand, reduces the value of the method used, as it can be easily reproduced. Therefore, scientists should be more hesitant to share, for example, methods that they created themselves and that are not yet highly standardized. Furthermore, current research suggests also taking a disciplinary perspective in the context of standardization into account, as the development of different standards might be necessary [25]. This study's aim is to examine scientists' method of data preparation in three different disciplines and to define the term "research data" according to their descriptions. Thereby, the standardization of research data will be analyzed by also including the possible influence of disciplines.

3 Theoretical Framework

According to Merton, the priority of discovery and scientific knowledge that is considered original is important for scientists because it is crucial for the acknowledgement of scientific reputation [4]. According to Pierre Bourdieu, scientific reputation is important for scientists' careers, as it can be used as scientific capital to accumulate, for example, further financial funding or to gain positions within the scientific field [15, 26]. Therefore, research data is the base or fundament of scientists' careers and it is necessary to understand how scientists use research data via sharing or providing access to generate and accumulate scientific capital and scientific reputation. A first step is to define research data according to scientists' own descriptions.

Following Bourdieu's field theory, each field—in this case, I will talk of science as a field—has its own set of rules ("nomos") that are unique to the field itself [27]. The "nomos" represents the logic of the field's practice, by means of which order is created in it. The creation of order is achieved through a common understanding, a consensus, about which practice is sought in the field. Actions of the field members (e.g., scientists) can be structured or guided by this logic. Because the members of the field are guided by this logic, as they recognize the rules and agree to them ("illusio"), a certain form of regularity or expectation of their actions within the field can be expected [27]. This creates a structure of legitimate patterns of action in the field, although it is dynamic and can be renegotiated. The experiences that the researchers gain in dealing with other members of the field (e.g., other researchers) are incorporated into the "habitus" of the individual members and will influence their future actions.³ By influencing future actions, the field logic can also be changed, provided that the actions are recognized as legitimate by other field members.

This concept is of interest as an attempt is made by political members, funding agencies, and publishing houses to change the logic within the scientific field by increasing the currently low level of willingness to share or provide research data through incentives and demands, as illustrated previously. If scientists agree to this interference, the actions connected to the interference can be then integrated into the "nomos" of the scientific field and thus change the logic of the field. Eventually, scientists' methods and willingness of sharing or providing access to data should change and increase.

4 Method and Design

This study uses a qualitative approach. The empiricism was conducted in collaborative research centres ("Sonderforschungsbereiche, SFB" and "SFB/Transregio, TRR"), a research

³ In short, the "habitus" is a collection of dispositions, through which behavioral and practical forms are generated [28].

program funded by the German Research Foundation (“Deutsche Forschungsgemeinschaft, DFG”). The choice of this special kind of research program lies in its own characteristics: Due to the possibility of funding an infrastructure project for data management, an external effect can be included into the analysis from the beginning. Furthermore, the projects’ uniformity in their expiry helps to allow for better comparability. Researchers are also allowed the privilege of scientific freedom, as scientists are free to choose the topic, structure, and organization of the project they want to work on. As they are called collaborative research centres, it is expected that scientists work in a collaborative way that may also lead to an increased habit of data sharing. The research program itself is structured by a major topic under which many smaller projects involve their own parts of the main research topic. The projects are not meant to be interdisciplinary, but nevertheless, they can be created in an interdisciplinary manner. Out of these centres, just one project was picked to make sure that the disciplinary aspects of the smaller project could be included into the analysis. The program itself is funded by the DFG over a period of 4 years, which can be extended in intervals of 4 years up to a total of 12 years and applied for by a whole university (SFB) or two/three universities (TRR) and not individual researchers. By January 2020, a total of 192 SFB and 83 TRR programs were funded by the DFG with a total of 797 Mio. euro.⁴

The sample selection was guided by five criteria: first, to include projects with an infrastructure project as well as those without an infrastructure project. Therefore, it is possible to analyze if an infrastructure project it influences scientists’ decision on whether or not to share research data. The second one was the discipline: Biology, computer science, and neuroscience were chosen as competition, the kind and amount of research data generated during the research process, as well as legal issues regarding the issue of data protection might slightly differ in their occurrence and importance across those disciplines. Third, the project’s starting year was decisive, as it was crucial that some research data had been generated within the project already. The fourth criterion was the status of the interviewee within the project but also regarding academic career, as this might influence scientists’ understanding of research data and their willingness to share it. Lastly, it was important that there had already been publications in the context of the project and its research aim. Based on the approach of theoretical sampling by Glaser and Strauss [29], the aforementioned criteria were developed and adapted in the course of the research process. After conducting the first interviews and briefly analyzing them, some of the already mentioned limitations and criteria were reconsidered to adjust the sample’s composition. The selection of further interviewees was therefore made according to those evolving criteria. In 2018, a total of 21 expert interviews were conducted with a structured guideline.⁵ The interviews were conducted in German; for a better understanding, relevant statements and quotes were translated into English for this paper.⁶ The guideline was divided into different topics containing such sections as focusing on scientists’ definition of research data and descriptions of their practice of data sharing, as well as their experiences with the infrastructure project. The interviews were recorded, transcribed, and afterward analyzed.

The interview analysis is based on qualitative content analysis (QCI) as well as on the approach of theoretical coding by Corbin and Strauss [30, 31]. In this context, the first

⁴ For further information, please visit www.dfg.de/en/research_funding/programmes/coordinated_programmes/collaborative_research_centres/index.html

⁵ Due to problems in recruiting a sufficient number of SFB members in biology, one member of a non-DFG-funded project was recruited. While the project line of the investigated project does not offer the possibility to apply for a specific infrastructure project, the project idea itself was the development and implementation of an infrastructure for the biological scientific community.

⁶ On the problem of translating interview transcripts and analyzing those transcripts, see Wettemann [32].

conducted interviews were read in an open approach line by line.⁷ During this approach, statements and phenomena that seemed to be of relevance were first highlighted and then annotated with definitions, concepts, and comments, so-called “codes.” As more interviews were conducted and more codes were generated during the process of reading and annotation, in a second step, those codes were grouped and clustered around recurring or relevant phenomena to form an initial coding system. The generated codes and transcribed interviews were then transferred into the software MaxQDA to carry out further coding and development of a final coding system. The last step of the analysis will be carried out in the following chapter by analyzing the codes and highlighted statements and phenomena, which will then be discussed with regard to the aforementioned theoretical framework.

5 Results

With the analysis still in progress, the presentation of results will focus on scientists’ definitions and descriptions of research data, the process of preparation, the effect of the special form of project organization in the form of an “SFB” on their willingness to share or make research data available, and a first glance at which state of preparation of the conducted research data is shared and with whom.

5.1 Research Data – Descriptions and Definitions

The term “research data” will be constructed based on scientists’ descriptions and statements. The analysis of the conducted interviews suggests that data exists mostly in three states of preparation: *raw data*, *prepared and/or partially analyzed data*, and *results data*. The processes of *data collection*, *data preparation*, and *data processing/evaluation* will be analyzed, too. Thus, it is possible to explain why the use of research data is sometimes limited to specific states of preparation or why sometimes research data does not exist in all states of preparation.

Due to the logic of the research process, the analysis will start with the *data collection* and the *raw data* generated during this process. Looking into the interviews, it becomes clear that raw data usually means research data that is the fundament or basis of the planned research project and process. It can be obtained by analyzing pre-existing research data or newly created by the researchers. It is characteristic for raw data not to have been processed thoroughly yet. The only steps carried out during this process are checking and storing the research data, putting together relevant databases, adding further data, or removing redundant data. Interviewees from biology or computer science in particular stated that raw data also includes data that has been checked by the scientists regarding its content and accuracy. In the field of biology, interviewees stated that data that needs further attention by the scientist is also considered raw data. They mostly described a process of adding further data or aggregating (concatenate) raw data. Particularly in biology, the generated raw data is checked for accuracy before further preparation. As the next quote shows, computer scientists also evaluate the raw data they have generated with regard to the research project and then decide whether the raw data should be stored for further processing and usage as research data:

Then you take out the datasets that are obviously rough, smooth them out; that is basically raw data. So, with raw data I don’t mean what is on the PC after the experiment, but what ends up in SPSS. This is the raw data. Or in the machine-learning program.

[PhD Student, computer science]

⁷ This process, as well as the first development of a coding system, was carried out together with Prof. Dr. Eva Barlösius.

Taking a closer look at the significance of the data collection shows that in general, this process is a fundamental step of the research process. Without this step, carrying out research is impossible. Furthermore, this process and its results are inextricably linked, as the following quote shows:

The samples, you see, the samples, it's not just a sample for nothing; it means the whole experiment. It's part of the experiment, I mean.

[Postdoc, biology]

With regard to data collection, it becomes clear in the interviews that the various methods used to collect or generate data show different degrees of standardization. The result is that the importance of data collection is interpreted differently by the interviewed researchers. These differences are most evident between biology and computer science: An interviewed PhD student in biology explained that this is assessed as a scientific performance, because an experiment has to be developed or adapted and carefully thought through in advance. According to an interviewed professor in computer science, on the other hand, the replication of experiments is simple, because the procedure can be largely derived from the publications without any problems. However, it seems like the described difference is not based on the disciplines but rather on the research project and interest itself, because in some projects, a standardized approach or method can be used, whereas in other projects, a new one has to be created. The difference here is the consideration of the data collection either as a part of the scientific research process or as an activity that is only necessary to advance the research process and could also be carried out under supervision without scientific expertise. However, a pronounced standardization of the data collection does not necessarily mean a simple or quick procedure, as the interviewees stated.

The second part of the research process is the *data preparation*, during which *prepared data and/or partially analyzed data* is generated. In this process, information and data will be added or removed, and the dataset will be prepared further. Unlike in the first step with adding and removing data, this step is carried out according to rules to avoid being assessed as data manipulation. Nevertheless, the interviewees described data preparation as a crucial part of the research process, as data might be “worthless” in its raw form without any preparation:

So, the data itself, that is, the measured figures, of course also always include the logs of what you were doing, so to speak. Without all the logs of what went on in the experiment, the data is of course useless.

[Professor, neuroscience]

Thus, the development and assignment of metadata is important to the scientists; especially in biology, the addition of so-called “alignments” is highly relevant for further processing and preparation of data. The removal of data is also an integral part of this step but is limited to data that scientists deem not useful or incorrect. Descriptions of this process were mainly found in biology and neuroscience interviews. For computer scientists, anonymizing raw data for further processing is a vital step in the preparation of research data.

As with the raw data, the prepared data and/or partially analyzed data generated in this process still exist in various data and/or saving formats. In this process, some differences between the disciplines become visible. Interviewees in computer science focused mainly on specific expertise and logic necessary to process and analyze data. Without this specific expertise or logic, it is impossible for other scientists to use the generated data. Hence, it is most crucial to them to process data comprehensibly and clearly for other scientists by adding further information:

Sometimes there are also imponderables; you don't want to discuss them unconditionally, because [they irritate] the doctors [rather] than do much more than help them, and so on. Yes. So, that's not because we want to withhold it from you, but there is just a lot of knowledge in it, no, and yes, that's why we are reluctant to give it out sometimes and even if so, then only commented or

something. So, we like to show them, but just pass them on, "Please, here you have the spectra," without annotation or anything, no.

[PD, biology]

The interviewees in neuroscience emphasized the process of data preparation by describing in incredible detail the amount of work necessary to transfer raw data into results data. They are also the only ones who described the use of mostly standardized methods in this process.

The last step is *data processing/evaluation*, in which *results data* is finally generated. This step is usually a quantifying process that is generally done in a statistical form. The format of the data generated in this process is therefore less varied than in the previous steps. Across disciplinary borders, scientists stated that for this procedure, a lot of effort, expertise, and knowledge is required:

Because, the real effort is not to produce the data. Maybe it is annotating the data. But the real effort is to analyze the data in a meaningful manner.

[Postdoc, biology]

At the end of this process, the statistical analysis often ends with transferring statistical results into diagrams, figures, and illustrations or other forms of presentation that are used in publications. This step is usually highly standardized, as overall, the analyzed disciplines' researchers did not mention various formats of data. This last step finally enables the interpretation of published results data by such persons without the needed specific expertise.

5.2 The effect of a special form of project organization

Considering the organization of a research project as an SFB and, if necessary, an additionally raised infrastructure project, the interviews suggest that scientists are fully aware of the basic concept: to achieve a research goal in cooperation with the other project members. Therefore, new findings or discoveries are shared among the members of an SFB. Additionally, findings and discoveries that might be helpful to make overall progress are shared, too. In general, the interviewees described a higher openness and willingness to operate more openly within the SFB. Nonetheless, this openness is restricted to specific states of preparation of the data, as mostly results data and conclusions or findings are shared with the other scientists:

Yes, of course. If there are new findings that are of interest to the SFB, it is the goal of the sharing that everyone gets to know it and can contribute if possible to push it forward.

[Postdoc, biology]

Besides sharing data and findings, the scientists also share their expertise to help and complement each other. However, this openness and helpfulness seems to be limited to sharing expertise, results data, and findings. When the interviewees were asked to describe how they share raw data or rules that might be associated with the sharing of raw data, they were not able to describe any. Their explanations were associated with group affiliations such as working groups or a laboratory (lab). Furthermore, it seems like raw data or prepared/partially analyzed data is not shared within the whole SFB, whereas it is shared with cooperation projects within the SFB. This practice is mostly justified by the lack of interest of the other colleagues within the SFB. The interviews also show that existing infrastructure projects are not or are only rarely used by the interviewed researchers. A neuroscientist explained that other projects often have no need for the data they would like to share:

Or our idea at the very beginning was that everything, all the data we have about the words, that we would like to make that available, the MEG things rather not because that, or in the sense of "What helps them in projects that actually work in semantics?" [For those who] very often deal with individual words and examples for certain categories as the main point of interest [...] we have something that [...] might be helpful. But then it didn't seem to be the need.

[Postdoc, neuroscience]

Descriptions of sharing data other than results data and conclusions or findings were limited to statements that highlighted cooperation with the SFB.

Furthermore, the existence of an infrastructure project does not seem to affect if and how research data is shared within the SFB, as the interviewees mostly illustrated ways of sharing outside the infrastructure project (e.g., e-mail, USB sticks, external hard drives). Often, the existence of the funded infrastructure project is not known to the scientists. If it is known, most interviewees could describe what it is or should be used for. Nevertheless, often, it is not used for that purpose or even used at all, as scientists seem to organize the process of sharing in their own projects by themselves:

I don't know what's the infrastructure project's purpose in our SFB, yes. Most certainly it is not pivotal for data sharing, but each project is responsible itself; therefore, that is not correct.

[Professor, computer science]

This result reproduces one of the main critiques in data sharing, as previously mentioned: the weak standardization in enabling data sharing and data storage due to the lack of or limited funding for necessary infrastructures. Rather, the scientists themselves (have to) develop their own standards and methods for this procedure.

6 Discussion

It has previously been explained that the term “research data” is highly complex and often, little attention is paid to the importance of the different states of preparation. Instead, mostly the difference in sharing published and unpublished data is highlighted. Additionally, the individual components of the process research data must go through are hardly examined. The analysis carried out in this study leads to the assumption that the different states of preparation, the different steps in the process and the individual components of these steps are highly relevant for sharing research data or making it available to other scientists. In the interviews, three different states of preparation could be identified, each of them associated with a step in the preparation process. The different processes themselves are assessed as almost as important as the research data generated because of the degree of standardization of the methods used in the process. In the sample, some scientists defined methods also as research data that should be shared with further information and instructions if possible, as they are not highly standardized and therefore hard for others to recreate. Additionally, they were less willing to share it or make it available to other scientists without checking its accuracy beforehand. Surprisingly, these differences do not depend on the disciplines but rather more on the research project itself. Furthermore, the standardization of methods used can be interpreted as an indicator of the intensity of a research topic: A highly standardized set of methods implicates that the scientists agree on the approach to certain steps of the data collection and preparation process. In addition, because raw data is checked and incorrect data is removed, not all the generated data is present in every state of preparation. Therefore, it seems logical that not all data is shared with other scientists. The interviewees also described the step of adding and removing data as vital. Hence, I come to the conclusion that when looking at “research data” and “meta data” in the context of “data sharing,” it is necessary to look not only at data generated during the research process, but also at the process of preparation and the methods themselves, as they are highly valued and shared or made available by the scientists, too.

The analyzed interviews also show that researchers share their research data little or not at all with other research projects of the SFB. The analysis also suggests that specific research data is not shared with all the individual project groups, but rather with cooperating projects of the SFB. Moreover, the infrastructure project is often unknown or not used by most of the interviewees. Regarding Bourdieu’s “nomos” and the scientists’ “habitus,” the habit of sharing research data using infrastructure projects does not seem to be part of the “habitus”

or “nomos” yet. Attempts to influence these concepts from outside the field of science and its logic seem to be hardly successful. Nonetheless, as scientists stated in the interviews, sharing and making research data available is part of the practice in the field, even though it is limited to specific projects or colleagues.

As shown also, data can be shared in and between interdisciplinary projects. Besides data, expertise is also highly valued and shared with other scientists. Therefore, it seems possible to transfer research data or expertise generated in one research area to another area of research to use it for further generating reputation. This leads to the assumption that research data itself is used as a capital that is invested within the scientific field to accumulate further scientific reputation.

7 Conclusion

In further research, the significance of the three different states of preparation and the individual steps of this process should be investigated more thoroughly, as this study leaves open questions regarding the value and use of the different states of preparation. A short glimpse at the conducted interviews hints that all of them are used for different purposes, at different times during the research process, and in different social groups.

As was shown very briefly in this study, data can be shared or made available at different points during the research process. This topic should be investigated further by taking a closer look at which kind of data is shared at which point during the research process. Because the timing of sharing or making data available during the research process can vary, how scientists justify their decision regarding a specific timing should also be analyzed. Additionally, it will be important to differ between published and unpublished data, as the significance of research data in the context of accumulating scientific capital in this special form of project organization has to be looked into more thoroughly.

Furthermore, the influence of the various other effects from outside the scientific field might be worth a closer look to find out how and which external factors are incorporated into the fields' logic because legal regulations in particular cannot be bypassed.

To tackle those topics, the theoretical approach should be adjusted by adding an approach that looks into the importance of social groups in the process of data sharing.

Funding and Acknowledgements

The project is funded by the German Research Foundation (DFG), project title “Zum Zusammenhang von disziplinären Originalitätskonzepten und handlungspraktischen Orientierungen für das Teilen von Daten” (No. 347305329).

I gratefully acknowledge the participants in my interviews as their explanations and descriptions are the fundament of this study. Furthermore, I am thankful for Eva Barlösius' help and suggestions in this research. Without the support of these people, this study would have been impossible.

References

1. R. K. Merton, *The Sociology of Science: Theoretical and Empirical Investigations*. (Chicago, IL: University of Chicago Press, 1973)
2. A. E. Thessen, D. J. Patterson, *ZooKeys* **51(150)**, 15 (2011)
3. C. Wiley, W. H. Mischo, *ISTL* (**95**) (2016)

4. R. K. Merton, *Entwicklung und Wandel von Forschungsinteressen. Aufsätze zur Wissenschaftssoziologie*, 258 (Frankfurt/M.: Suhrkamp, 1985). (First published 1957 in English: ASR **22(6)**, 635)
5. S. Irwin, M. Winterton, Timescapes Working Paper Series **4** (2011)
6. M. A. Haendel, N. Wasilevsky, J. A. Wirz, PLOS Biology **10(5)** (2012)
7. C. J. Volk, Y. Lucero, K. Barnas, Environmental Management **53(5)**, 883 (2014)
8. C. Tenopir, E. E. Dalton, S. Allard, M. Frame, I. Pjesivac, B. Birch, D. Pollock, K. Dorsett, PLOS ONE **10(8)** (2015)
9. C. Borgman, JASIST **63(6)**, 1059 (2012)
10. S. J. Ceci, ST&HV **13(1/2)**, 45 (1988)
11. B. Fecher, S. Friesike, M. Hebing, S. Linek, A. Sauer mann, Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, 246 (2015). <https://www.ratswd.de/publikationen/working-papers/2015> [Accessed: April 20, 2020]
12. T. S. Kuhn, Science **136(3518)**, 760 (1962)
13. T. S. Kuhn, *Die Entstehung des Neuen. Studien zur Struktur der Wissenschaftsgeschichte*, 308 (Frankfurt/M.: Suhrkamp, 1988)
14. P. Bourdieu, SSI **14(6)**, 19 (1975)
15. P. Bourdieu, *Science of Science and Reflexivity*. (Cambridge Polity Press, 2004)
16. E. Barlösius, ST&HV **44(6)**, 915 (2019)
17. E. Barlösius, F. Knoke, M. Pook-Kolb, in *Handbuch: Interpretativ forschen*, L. Akremi, N. Baur, B. Traue (Eds.), 133 (Weinheim: Beltz, 2018)
18. E. Leahey, ST&HV **33(5)**, 605 (2008)
19. C. Haeussler, L. Jiang, J. Thursby, M. Thursby, RP **43(9)**, 1621 (2013)
20. B. Fecher, S. Friesike, M. Hebing, S. Linek, Palgrave Commun **3**, 1 (2017)
21. I. Stamm, FQS **19(1)**, Art. 16 (2018)
22. Y. Kim, J. M. Stanton, asis&t **67(4)**, 776 (2016)
23. J. Hartter, S. J. Ryan, C. A. Mackenzie, J. N. Parker, C. A. Strasser, PLOS Biology **11(9)** (2013)
24. D. MacMillan, JAL **40(5)**, 541 (2014)
25. B. Fecher, *Eine Reputationsökonomie: Der Wert der Daten in der akademischen Forschung*. (Wiesbaden: Springer VS, 2018)
26. P. Bourdieu, *Homo academicus* (Frankfurt am Main: Suhrkamp, 1991)
27. P. Bourdieu, *Praktische Vernunft. Zur Theorie des Handelns*. (Frankfurt/M.: Suhrkamp, 1998)
28. P. Bourdieu, *Sozialer Sinn. Kritik der theoretischen Vernunft*. (Frankfurt/M.: Suhrkamp, 1987)
29. B. G. Glaser, A. L. Strauss, *The Discovery of Grounded Theory. Strategies for qualitative research*. (Chicago: Aldine, 1967/1998)
30. J. Corbin, A. Strauss, *Basics of Qualitative Research. Grounded Theory Procedures and Techniques*. (California: Sage Publication, Inc., 1990)
31. M. Schreier, FQS, **15(1)**, Art. 18 (2014)
32. U. Wettemann, in *Qualitative Interviewforschung in und mit fremden Sprachen*, J. Kruse, S. Bethmann, D. Niermann, C. Schmieder (Eds.), 101 (Weinheim und Basel: Beltz Juventa, 2012)