Handling of Research Data in DFG Individual Grants – An Analysis of Proposals in the Social and Behavioural Sciences

Research funding organisations are increasingly seeking to promote professional research data management and create incentives for researchers to share their data. The DFG also invites applicants to describe in the proposal how they intend to handle the data generated in a project. At an interdisciplinary roundtable discussion in spring 2018 on "Research Data Management in the Social and Behavioural Sciences: Problems and Need for Action in a DFG Context", the DFG examined the information provided in proposals. This infobrief summarises the types of research data management that were discussed.

1 Development of the discourse on research data management and reason for the study

For some years now, both nationally and internationally, research organisations and institutions have been stepping up their efforts to design solutions for improved research data management and create incentives for researchers to share their data. In 2010, the Alliance of German Science Organisations called for the "long-term archiving of and open access to data from publicly funded research" (Alliance of German Science Organisations, 2010, p. 2). The same need has been expressed by the scientific community itself. Empirical surveys show that as a result of different disciplinary cultures, empirical approaches and incentive structures, data sharing is practised to highly varying degrees (Borgmann, 2012, p. 13; Tenopir et al., 2011, pp. 2–3). In the social and behavioural sciences, these efforts are encouraged by the results of large-scale replication studies in experimental psychology (Open Science Col-

laboration, 2015) and experimental social science (Camerer et al., 2018), which called attention to quality issues in these fields.

According to a resolution by the DFG Senate on the handling of research data in October 2015, "[t]he long-term archiving and accessibility of research data [...] contributes to the traceability and quality of scientific work and enables researchers to carry on work begun by others" (DFG, 2015, p. 1).[1] Against this background, learned societies and DFG review boards discussed standards for research data management appropriate to different subject areas. One example is the German Psychological Society (DGPs), which in 2016 formulated a set of guidelines on handling research data (see Schönbrodt et al.,

---

1 Here research data is described as „an essential foundation for scientific work". Attention is also drawn to the heterogeneity of research data: „The diversity of this data reflects the wide range of different scientific disciplines, research interests and research methods. Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work. Methodical forms of testing such as questionnaires, software and simulations may also produce important results for scientific research and should therefore also be categorised as research data." (DFG, 2015, p. 1).

2017). In addition to two roundtable discussions organised by the DFG, workshops were also held in individual disciplines.

In spring 2018, on the initiative of the Economics review board, a wide-ranging roundtable discussion took place in Berlin with representatives of the social and behavioural sciences. The aim of the discussion was to identify shared problems and lines of discussion relating to research data management and define the resulting shared need for action.

To place the discussion on an empirical basis, a content analysis was carried out on statements relating to research data management in proposals submitted to the DFG. This infobrief presents the key results of this analysis.

2   Structure and Content of the Study

The sample for the evaluation of DFG proposals includes new and renewal proposals in the research grants programme, submitted in 2016 to the review boards for Linguistics, Educational Research, Psychology, Social Sciences and Economics. This range of subject areas is drawn from the primarily empirical areas of the social and behavioural sciences. As the majority of proposals in law do not generate empirical data, these were not included in the analysis. Linguistics, on the other hand, does involve strongly data-based research approaches, so this humanities subject was also included. The evaluation relates to a 25 percent random sample of the 1,108 proposals and therefore a sample size of N=275. The distribution between the five participating subject areas can be seen in Table 1.

To what extent is the topic of research data management addressed in DFG proposals? Item 2.4 in the Proposal Preparation Instructions invites applicants to describe how data will be handled and explain "if and how these will be made available for future reuse by other researchers" (DFG, 2018, p. 5). This applies to all projects which systematically produce research data or information. The instructions also refer to standards, examples of best practice and the option of requesting with a project "project costs [...] associated with making research data available for future reuse" (DFG 2018, p. 5). To answer the questions described, the sections of text under this item of the proposal were extracted for the 275 research grant proposals in the random sample and systematically analysed using the qualitative analysis software MAXQDA.[2] A coding scheme was developed on the basis of the "Basic Information on Research Data Management" published by the German Data Forum (RatSWD, 2016). This paper comprises "Guidance on the proposal and review of data-generating and

---

2   The analysis did not take into account whether the research projects actually generate data or use secondary data.

Table 1: Random sampling by review board		
Review board	**Number of proposals**	**25% sample**
104 - Linguistics	159	39
109 - Educational Research	144	36
110 - Psychology	294	73
111 - Social Sciences	307	76
112 - Economics	204	51
	1,108	**275**

Data basis and source:
New and renewal proposals for research grants in selected subject areas received in 2016.

data-using research projects" and addresses aspects of data backup, archiving, dissemination and protection, research ethics[3] and the methodical, technical and bibliographical documentation of data. These aspects were supplemented by aspects of research data management obtained inductively from the proposals themselves. These additions include the use of standards or guidelines of learned societies, the description of the dataset and/or data evaluation and the publication of project results.[4] After the text was coded, the codes were descriptively evaluated.
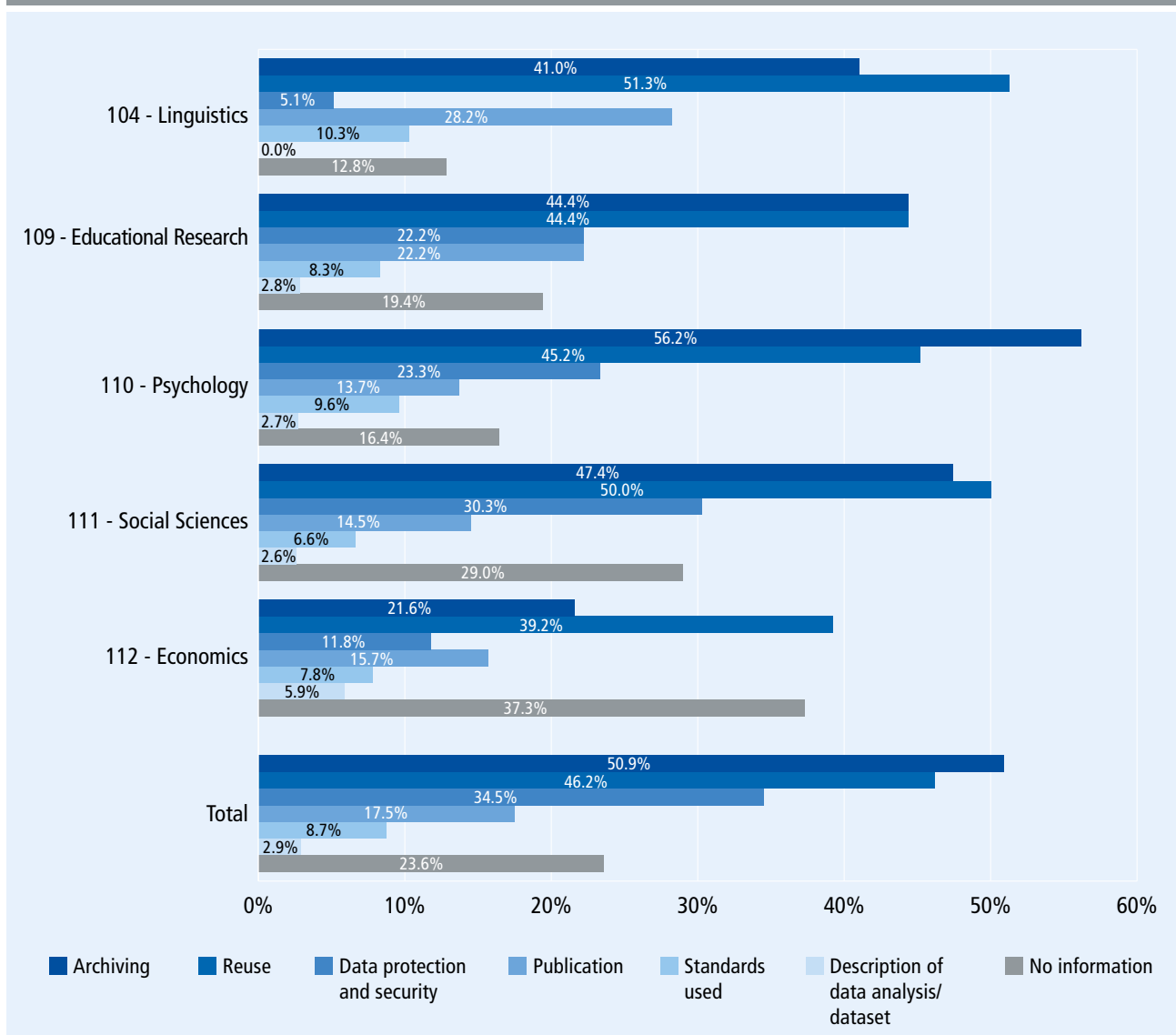
3 Results of the content analysis

In Figure 1, the different categories of data handling are itemised. Around one quarter of proposals (24 percent) contain no information

---

3   As research ethics aspects are not addressed under item 2.4 in DFG proposals, they were not taken into account.

4   According to the Proposal Preparation Instructions, the description of the dataset and information about data analysis and the publication of project results should be included under item 2.3, „Work programme including proposed research methods". However, since these aspects are also discussed under the item on data handling, these categories were included in the analysis.

Figure 1:
Information provided on data handling by review board

	Archiving	Reuse	Data protection and security	Publication	Standards used	Description of data analysis/ dataset	No information
104 - Linguistics	41.0%	51.3%	5.1%	28.2%	10.3%	0.0%	12.8%
109 - Educational Research	44.4%	44.4%	22.2%	22.2%	8.3%	2.8%	19.4%
110 - Psychology	56.2%	45.2%	23.3%	13.7%	9.6%	2.7%	16.4%
111 - Social Sciences	47.4%	50.0%	30.3%	14.5%	6.6%	2.6%	29.0%
112 - Economics	21.6%	39.2%	11.8%	15.7%	7.8%	5.9%	37.3%
Total	50.9%	46.2%	34.5%	17.5%	8.7%	2.9%	23.6%

Data basis and source:
Content analysis of information provided under item 2.4 of proposal. All proposals in random sample (N=275), multiple mentions possible. „No information" indicates data handling not specifically addressed.

under item 2.4. An examination of the proportions in the different subject areas reveals a heterogeneous picture. While the proportion of linguistics proposals with no information on research data management is fairly low at around 13 percent, over 37 percent of applicants in economics and 29 percent of those in the social sciences provide no information as to how they intend to handle research data. Educational research and psychology fall in between with 19 percent and 16 percent respectively.

Overall, the types of information provided most frequently are on data archiving (51 percent) and reuse (46 percent). The same result applies to the individual subject areas. Proposals in linguistics and economics more often contain information on reuse than on archiving. Of all the applicants who comment on the reuse of research data, only a very few (around 2 percent) are expressly opposed to the data being reused.[5] Data protection and security issues were addressed in just under 35 percent of the proposals examined. There are clear differences between the different subject areas (e.g. 5 percent in linguistics and 30 percent in the social sciences), which are presumably due to the use of different types of data.[6]

The publication of the project results does not constitute a data management plan in itself, but planned publication was mentioned in 17.5 percent of proposals. This applies to over a fifth of proposals in linguistics and educational research. In psychology, social sciences and economics, 15 percent of proposals or fewer refer to the publication of research results.

In response to the question on the handling of research data, only a small number of applicants (between 6 percent and 10 percent) referred to the use of standards and guidelines from learned societies or other research organisations. 5 percent of applicants refer to the DFG guidelines while approximately 2 percent mention the guidelines of the German Psychological Society (DGPs). Only a few proposals refer to the standards of the American Psychological Association (APA) and the guidelines of GESIS, the German Data Forum (RatSWD) and the Educational Research Data Association (Verbund FDB).

To provide a more detailed insight into the range of responses relating to data protection and security and also the archiving of research data, the information provided on these aspects is broken down below (Figure 2). In total, 12 percent of proposals that contained statements on data protection and data security provided no concrete details. In 83 percent of cases, data protection was to be ensured by anonymising the research data. 22 percent of proposals mentioned seeking the participants' consent for data protection purposes. In only 12 percent of proposals where data protection and security were mentioned was the storage of the original data discussed.
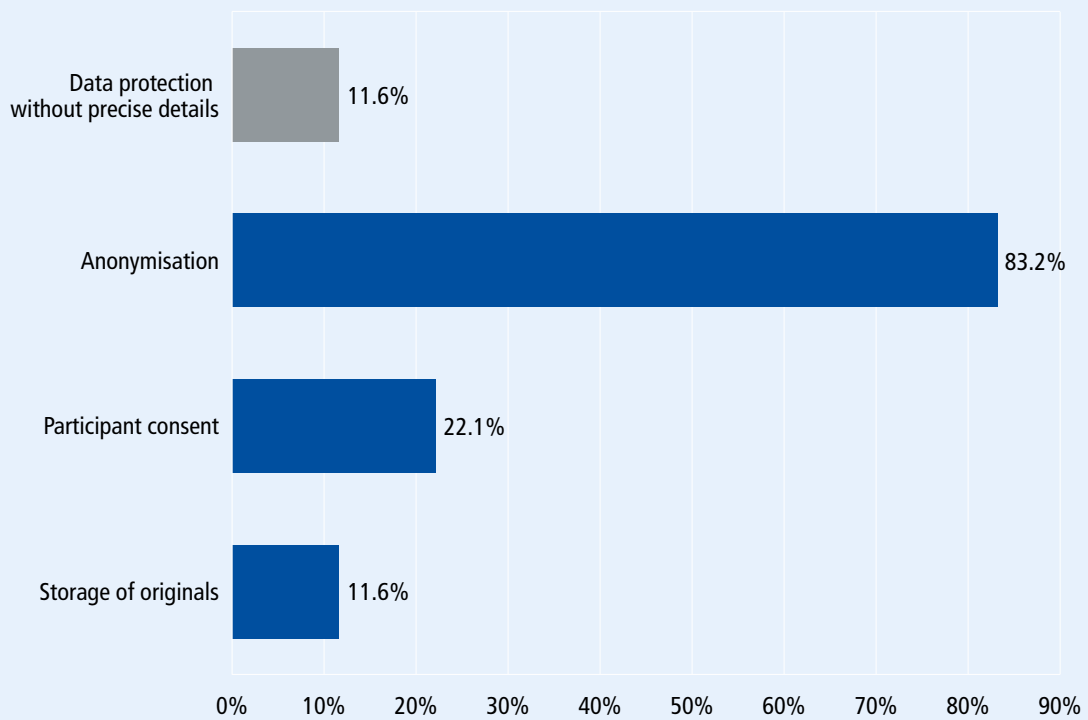
In terms of data archiving, 21 percent of proposals that mentioned this topic provided no further details (Figure 3). In 37 percent of proposals, it is specified that the data is to be archived on a university server. In 35 percent of proposals, external archives or research data centres are considered for archiving purposes. 13 percent of proposals that provided details on archiving plan to use an online repository and 10 percent mention archiving the data processing scripts.

Overall, in the cases where aspects of data protection and archiving are taken into account with respect to data handling, further details are generally provided. Although archiving is mentioned more frequently in proposals than data protection, in a fifth of cases the information provided on archiving is not concrete.

---

5   In a few cases, limiting remarks were also made about the possibility of reusing the data (on request; in some circumstances; for cooperation partners only).
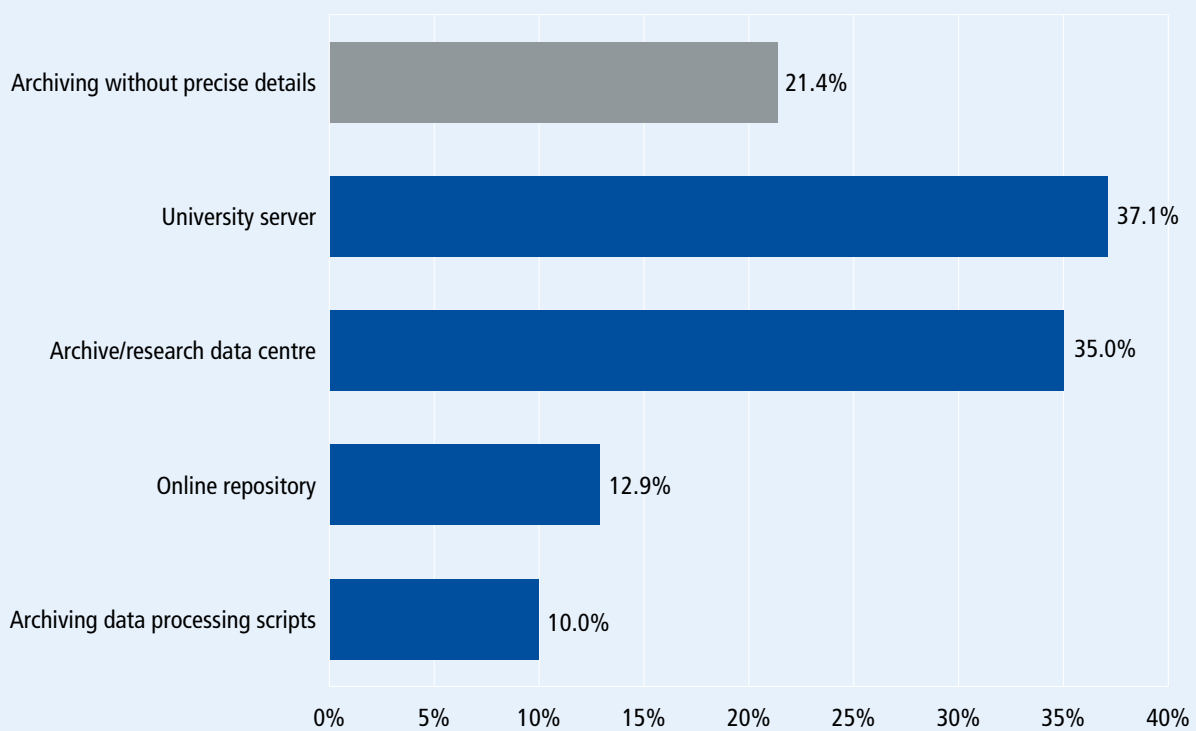
6   While linguistics research tends to use data in the form of text corpora, in social sciences projects the usual form is survey data. These different kinds of data require different levels of attention to data protection and security.

Figure 2:
Information provided on data protection and security from all proposals where mentioned

Data protection without precise details	11.6%
Anonymisation	83.2%
Participant consent	22.1%
Storage of originals	11.6%

Data basis and source:
Content analysis of information provided on data protection and security under item 2.4 of proposal. Proposals in random sample that included relevant information (N=95), multiple mentions possible.

Figure 3:
Information provided on archiving from all proposals where mentioned

Archiving without precise details	21.4%
University server	37.1%
Archive/research data centre	35.0%
Online repository	12.9%
Archiving data processing scripts	10.0%

Data basis and source:
Content analysis of information provided on archiving under item 2.4 of proposal. Proposals in random sample that included relevant information (N=140), multiple mentions possible.

4   Conclusions

In their proposals, applicants provide different kinds of information on data handling with various levels of abstraction. Overall, a heterogeneous understanding of this aspect of the proposal can be observed. The difference ranges from applicants who provide no information at all in this section or only mention the publication of results to those who provide very detailed information. Some applicants describe a detailed data management plan which covers the whole data life cycle, while others aim to comply with standards or endeavour to make the data available in a subject-specific online repository[7].

The extent to which the quality of the research data management plan is considered in the review process and influences the funding decision is not covered in this analysis; this is a topic for further analyses. In the review and evaluation of funding proposals submitted to the DFG, data handling is assigned different degrees of relevance depending on the subject area. The very high number of applicants who provided no information under item 2.4, and the fact that to date there have been only a few references to standards, make clear that the cultural shift from professional data documentation and preparation towards reuse and data sharing is progressing at different rates within the various subject areas. The DFG is therefore calling on the "various sections of the scientific community to reconsider their handling of research data and develop appropriate guidelines for the discipline-specific use of such data and, if appropriate, open access to it."[8] Currently, for the social and behavioural sciences there are subject-specific guidelines for applicants to the review boards on Linguistics (DFG Review Board on Linguistics, 2017), Educational Research (Stanat 2015; here currently only on the handling of quantitative research data) and Economics (DFG Review Board on Economics, 2018). There also exist statements from a number of learned societies, for example the German Psychological Society (Schönbrodt et al., 2017) and the German Educational Research Association (GERA, 2017), which set out recommendations for handling research data, and more general guidelines, such as the German Data Forum Orientation Guide (RatSWD, 2016).

The use of subject-specific repositories, in particular, has the advantage that it ensures long-term archiving, findability and preparation of the data. Information Infrastructures for Research Data is a separate DFG programme that helps researchers to establish such repositories, data service centres and information platforms and lay the foundations for improved data handling, for example, with the inclusion of advice or basic and advanced training. These repositories benefit in turn from the demand from researchers. The decision made recently by the Joint Science Conference (GWK) on the establishment of national research data infrastructures (NFDI) is another policy step in the direction of standardised research data management. Within the framework of NFDI, the funded consortia are to deliver (subject-specific) infrastructures for archiving, ordering, processing and provision to facilitate the scientific reuse of data (GWK, 2018).

References

Alliance of German Science Organisations (2010): "Principles for the Handling of Research Data" (in German).
www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_156.pdf

Borgmann, Christine (2012): The Conundrum of Sharing Research Data. In: Journal of the American Society for Information Science and Technology 63, 6, p. 1059–1078.

---

7   re3data (www.re3data.org) is a directory of research data repositories, part-funded by the DFG and used worldwide. The DFG portal (http://risources.dfg.de/) also lists various types of research infrastructures in Germany and, currently, some 100 research data repositories.

8   www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data

Camerer et al. (2018): Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015, Nature Human Behavior 2, p. 637– 644.

DFG Review Board on Linguistics (2017): Guidelines for building language under German law, CLARIN Legal Issues Committee (CLIC), White Paper Series. www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_review_board_linguistics_corpora.pdf

DFG (2018): Proposal Preparation Instructions – Project Proposals. DFG form 54.01 – 09/18. www.dfg.de/formulare/54_01/54_01_en.pdf

DFG (2015): Guidelines on the Handling of Research Data, approved by the DFG Senate on 30 September 2015. www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_research_data.pdf

German Data Forum (RatSWD) (2016): "Basisinformationen zum Forschungsdatenmanagement. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte".

German Educational Research Association (GERA) (2017): Statement by GERA on the Archiving, Provision and Reuse of Qualitative Research Data in Educational Research (in German). www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2017.09_Archivierung_qual._Daten.pdf

DFG Review Board on Economics (2018): Managing Research Data: What does Review Board 112 "Economic Sciences" expect of applicants? www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/fachkollegium112_forschungsdatenmanagement_en_1811.pdf

GWK (2018): Information infrastructures/NFDI (in German) www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/

Open Science Collaboration (2015): Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Schönbrodt, Felix; Gollwitzer, Mario; Abele-Brehm, Andrea (2017): "Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien", on behalf of the Council of the German Psychological Society (17/09/2016), in: Psychologische Rundschau 68 (1), p. 20–35. https://epub.ub.uni-muenchen.de/37158/1/10.1026_0033-3042_a000341.pdf

Stanat, Petra (2015): Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft" der DFG, Erziehungswissenschaft, 2, pp. 75 –90. https://doi.org/10.3224/ezw.v26i1.19514

Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev Umur; Wu, Lei et al. (2011): Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. https://doi.org/10.1371/journal.pone.0021101

(Version dated: 05/03/2019)