FORS GUIDES

Data Sharing in the Social Sciences

Marieke Heers[1] iD

[1] FORS

Abstract:

In social science research there is a push towards open research and open data. Open Data implies that researchers make their data available to other researchers so that the data can be re-used. This Guide describes the requirements and provides researchers with guidance on how and where to share social science research data with focus on the Swiss research environment.

How to cite:

The FORS Guides to survey methods and data management

The [FORS Guides](#) offer support to researchers and students in the social sciences who intend to collect data, as well as to teachers at university level who want to teach their students the basics of survey methods and data management. Written by experts from inside and outside of FORS, the FORS Guides are descriptive papers that summarise practical knowledge concerning survey methods and data management. They give a general overview without claiming to be exhaustive. Considering the Swiss context, the FORS Guides can be especially helpful for researchers working in Switzerland or with Swiss data.

1.   DATA SHARING AND OPEN RESEARCH DATA

During the last decade, there was a change in the research environment and research culture: An important characteristic of this change was a push towards open research and open data. Open data implies that researchers make their data available to other researchers so that the data can be re-used. Overall, in the social sciences there is a growing demand for more transparency throughout the research cycle (see for example van der Zee & Reich, 2018). This is also due to the observation that many published results are not reproducible (Camerer et al., 2016; Open Science Collaboration, 2015). In this context, the replication crisis has played a major role (Freese & Peterson, 2017; Nature, 2018; see also Heers, 2021).

Against this background, researchers are asked to make available the data they collect in the context of their projects. These requests come from research funders[1] and scientific journals, but also from peers. Especially amongst younger researchers there seems to be a growing understanding that research data – and related materials – should be reusable and be made available. That also explains the popularity of platforms such as Github[2] and Open Science Framework[3].

The general assumption is that research data should be made available to other researchers at the end of a research project, unless there are ethical, legal or contractual reasons not to do so – the mantra is "as open as possible, as closed as necessary". Research data might be re-used by the original research team as well as by other researchers. Data sharing is relevant for both quantitative and qualitative data. This Guide provides a general overview of data sharing in the social sciences, and is written is written with the assumption that most social science data are not fully anonymised. Big data and social media data are not the focus of this Guide.

2.   MOTIVATIONS FOR DATA SHARING

There are several motivations for and advantages of data sharing. These refer to science and society more generally, but there are also benefits at the level of the individual researcher. Most importantly, data sharing allows for transparency, reproducibility, and verification of research findings (Tedersoo et al., 2021). This is crucial for producing robust research evidence and for scientific progress. Second, by sharing their data, researchers can make better use of the data they have produced. This is because when preparing data for sharing, and carefully noting and describing the research steps, researchers are forced to argue well why specific decisions are taken. If data are collected with the idea that they will be shared, this might have implications for the data collection and related documentation. Researchers can also make better use of the data if these are shared via an appropriate repository (such as SWISSUbase[4]), since the data are preserved for the long-term. Data sharing also reduces the burden on respondents. Lastly, sharing data is important to meet the requirements of some scientific funders and journals (Logan et al., 2021). Many funders have made data sharing a requirement for obtaining research funding. In Switzerland, the Swiss National Science

---

[1] For example: https://www.snf.ch/de/dMILj9t4LNk8NwyR/thema/open-research-data
[2] https://github.com/
[3] https://osf.io/
[4] https://www.swissubase.ch/en/

Foundation (SNSF) now explicitly asks researchers to share their data from projects for which they received funding[5].

Considering the motivations at the level of the individual researcher, sharing data makes research work and results more visible. This valorizes and acknowledges the work on data production and data management (including anonymization and documentation). Several studies have shown that scientific articles, for which the data are published, are cited more than articles for which the data are not available (Colavizza et al., 2020; Drachen et al., 2016; Piwowar et al., 2007). Sharing research data increases the quality of data and research. If, from the beginning of the project, researchers plan to share their data, this increases the quality of the data and research as they tend to better document their work and key decisions. Throughout the project, when researchers have the idea of data sharing in mind, they are, for example, motivated to describe the data collection instruments. By sharing in an appropriate repository, researchers can make their data FAIR (findable, accessible, interoperable, reusable) (Wilkinson et al., 2016), a requirement that is required from funders. The FAIR principles are described in more detail in Wilkinson et al. (2016). Moreover, it encourages new collaborations and avenues of research.

3.   WHAT TO SHARE?

When researchers decide to share their data, an important question is: What to share? This question should already be asked (and at least partly answered) early in the project. That allows for preparing the data and documentation throughout the research process and contributes to the overall quality of the data and documentation. It should be emphasized that the researcher or the research team know their project best, and that, based on this, they are most competent in deciding what to share. Data service experts can help and guide researchers in the decision-making process. What should be shared depends on many factors. The most important ones are presented in what follows. Here it should be emphasized that both quantitative and qualitative data can be shared.

3.1   DATA THAT ARE PUBLICLY FUNDED

Data that are publicly funded should be shared. This has become a requirement of many research funders, including the SNSF. Amongst others, this is clarified in the Guidelines concerning the Data Management Plan (DMP).[6]

3.2   DATA UNDERLYING PUBLICATIONS

More and more scientific journals ask researchers to share the data used in publications and, in the past years, have developed policies thereon. When submitting a manuscript, it is becoming a norm that researchers have to fill in a "data availability statement". This is usually an indication where the data are available, how they can be accessed, and under what conditions. With this, scientific journals want to contribute to the transparency of the research process. This also increases the quality of publications and gives more credit to the work related to the data production process. For example, on the website of the *European*

---

[5] https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies. Some exceptions exist.
[6] https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies

Educational Research Journal as part of the submission criteria the following is stated with regard to Research Data[7]:

> *"The journal is committed to facilitating openness, transparency and reproducibility of research, and has the following research data sharing policy. [...]*
> *Subject to appropriate ethical and legal considerations, authors are encouraged to:*
>   - *share your research data in a relevant public data repository*
>   - *include a data availability statement linking to your data. If it is not possible to share your data, we encourage you to consider using the statement to explain why it cannot be shared.*
>   - *cite this data in your research"*

Other journals have similar requirements. While some journals encourage researchers to share their data, others have stricter requirements. The fact that articles with deposited data are cited more often than others (Colavizza et al., 2020; Drachen et al., 2016; Piwowar et al., 2007) is a major incentive for researchers to share their data.

3.3   ADDITIONAL INFORMATION ON DATA PROCESSING AND ANALYSIS

In addition to sharing the data produced as part of research projects, researchers should also share the documents produced as part of the data analyses, such as syntaxes and scripts, also referred to as replication material. This allows for more transparency in research and gives other researchers the possibility to validate and build on their colleagues' work. For a more detailed description of the motivations and requirements for sharing replication materials, please consult the respective FORS Guide (Heers, 2021). The FORS replication service allows researchers to easily share their replication materials.[8]

4.   RE-USE PURPOSES OF SHARED RESEARCH DATA

There are a number of purposes for which data can be of interest for further use. Often, data can be used in different ways which go beyond the use they were initially collected for. These use cases include the following.

Data can be used for replications of the initial analyses (Heers, 2021). Similarly, it can be used for amplified analyses, that is, a comparison or combination of several datasets. Data that are shared (and properly preserved) are valuable for carrying out longitudinal research. Data can then also be matched with data resulting from more recent data collections. The major use of shared data is probably further analyses, i.e. an in-depth examination of a question or aspect of the data that has not been fully analyzed in the primary study. Data that were collected on unique occasions and which collected non-reproducible materials are of high value. For example, data collected during the first Covid-19 wave. Later, it will be hard to collect data on such events without accepting problems related to re-call bias. Data that are also very valuable for sharing are on to hard-to-reach populations, such as on the elderly (for example, The Survey of Health, Ageing and Retirement in Europe (SHARE))[9]. Getting access to these populations is hard and this is often a burden for respondents. This burden should be kept low; hence, once data have been collected on these populations, they should be made available so

---

[7] https://journals.sagepub.com/author-instructions/EER#ResearchData
[8] https://resources.swissubase.ch/replication/
[9] For details see http://www.share-project.org.

that they can be fully exploited. Another example is data collected in schools. Data collection takes away time from instruction, parental permission has to be requested, and, therefore, the data should be as useful as possible[10].

Shared data are also a very valuable resource for teaching, in particular for methodology. Students enjoy using "real" data. Also, for Bachelor or Master theses, available high-quality data are appreciated.

5.   REQUIREMENTS FOR DATA SHARING

Sharing data requires that adequate, understandable, and well-organized documentation is made available together with the data. This allows other researchers to make sense of the data. The data must also be cleaned. Moreover, valid consent for sharing must have been obtained. In accordance with the informed consent, the data must be sufficiently anonymized. There also needs to be a definition of the access conditions. Upon sharing the data (and related materials) researchers need to clearly define who can have access to the data and how. In this sense, the data producer stays in control of the data. Repositories such as SWISSUbase guide researchers with this[11]. Before depositing data, researchers also have to consider intellectual property issues: Is the researcher allowed to share the data? When sharing data, researchers should also carefully think about the format of the data. Ideally the data should be shared in an open-source format. Finally, the FAIR principles must be considered (Wilkinson et al., 2016). Here, it is important to emphasize that it is mostly repositories and not individual researchers that make research data FAIR. Therefore, researchers should deposit their data in a repository that follows the FAIR principles. For the social sciences in Switzerland, this is the case for SWISSUbase. A detailed description of the FAIR principles can be found in the Appendix. By applying the FAIR principles, researchers ensure that the data can be used for secondary analyses and that they are preserved for the long-term. The Swiss National Science Foundation SNSF encourages researchers to adhere to the FAIR principles and provides an overview of what each of the four principles implies for researchers[12].

6.   DATA SHARING THROUGHOUT THE RESEARCH PROCESS

From the above, it has become clear that data sharing should be considered from the beginning of the project and be worked on throughout the project. Figure 1 summarizes the most important steps throughout the research process that later facilitate data sharing. During the planning stage, informed consent must be obtained. On top of asking research participants their consent for taking part in the study, they should also consent for their data to be archived and shared, and potentially used for teaching. In the data processing stage, anonymisation of the data should be carried out as stated in the consent form[13]. While in figure 1 anonymisation is mentioned, it should be said that often it is not possible (or intended) to fully anonymise the data. Researchers can then de-identify the data, i.e. different approaches can be employed to

---

[10] Examples of large and publicly available educational data sources are PISA (https://www.oecd.org/pisa/) and AES (https://www.uegk-schweiz.ch/).
[11] https://resources.swissubase.ch/help/user-guide/
[12] https://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf
[13] For more practical guidance on these topics check the numerous FORS Guides:
https://forscenter.ch/publications/fors-guides/

reduce the personal information in a dataset. For example, information can be aggregated. For a detailed treatment of these issues, please consult the [FORS Guides](#).

Finally, if applicable, during the dissemination stage, researchers can consider restricting the access to the data. Whether this should be done or not depends on the data. For example, if the data do not contain any personal information, then this is not necessary. If, due to reasons related to confidentiality or ethical reasons, the full dataset cannot be shared, researchers should think about a partial dissemination of the material.

Figure 1. Data sharing throughout the research project.
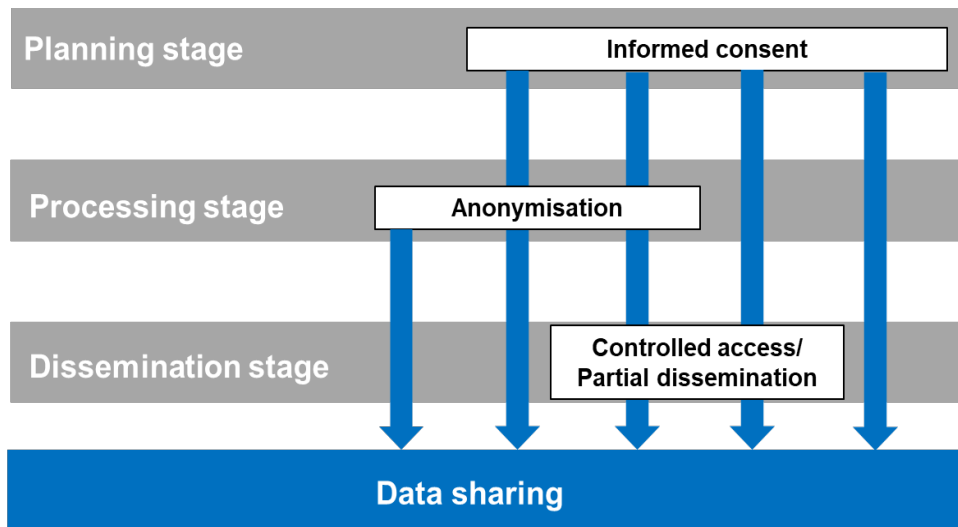
Source: Data management support team FORS.

7. WHERE TO ARCHIVE AND SHARE DATA?

Data should be shared in trustworthy and non-commercial repositories that enable the FAIR principles. Different types of archives and repositories exist, the most common ones are domain-specific repositories, institutional repositories, general purpose repositories, archives referring to specific research projects and journal supplementary material services. Researchers should privilege discipline-specific repositories or institutional repositories. In particular, domain-specific repositories can guide researchers, since they usually have data experts in that particular domain. Moreover, they have domain-specific metadata that help researchers structure their work.

When choosing a repository, researchers should make sure that the candidate repository offers long-term preservation and persistent identifiers. Moreover, repository must accept the respective data formats and data types (e.g., videos or brain scans). It should also give visibility to the data and conduct some quality control of data and documents. There should be a catalogue for discovery and sufficient dissemination capacity. Finally, the repository should allow researchers to control the access to their data.

8. DATA SHARING IN THE SWISS SOCIAL SCIENCES

The main Swiss research funder, the SNSF, requires researchers to share data that have been collected as part of funded projects. More information on their open data policy can be found

on their website.[14] There are several benefits of sharing data through a repository. A major advantage is that repositories, such as SWISSUbase, take care of key practices to make data FAIR. This includes practices relating to metadata, file formats, data access and persistent identifiers. Repositories also take care of the promotion and dissemination of the data.

Overall, data sharing in the Swiss social sciences has become a common practice. The largest repository is SWISSUbase, which contains more than 800 datasets from all social science disciplines, including education, psychology, and economics. Despite the overall growing importance and practice of data sharing in social science research carried out in Switzerland, there are considerable differences across disciplines, with researchers from some disciplines being more reluctant to share data than others. So far, most shared datasets are quantitative data, but the sharing of qualitative data is becoming more common.

The FORS Data Service advises encourages the sharing of data via SWISSUbase. Like most repositories, SWISSUbase has contracts for depositors and for users. With respect to data sharing, it offers different options, including Creative Commons licenses. For the social sciences, SWISSUbase offers a specific contract that allows researchers to retain control over their deposited data, as they can define their own access restrictions, access conditions and embargos. SWISSUbase provides services that make data FAIR and has services and practices with respect to documentation, metadata, file formats, and data access. All data deposited in SWISSUbase are stored on secure Swiss servers, and the FORS Data Service experts support researchers with any questions related to the deposit and sharing of data. Qualitative and quantitative data can be shared in SWISSUbase. A full list of repositories can be found on Re3data.org.

9. RECOMMENDATIONS FOR RESEARCHERS

Recommendation 1 – Start thinking about data sharing at the start of the project. This also includes proper documentation of the project and the data. If good data management practices are applied throughout the project, then sharing is the relatively easy final step.

Recommendation 2 – Make the data underlying scientific journals available. This will not only contribute to scientific transparency but might also increase the number of citations of your work.

Recommendation 3 – Check out SWISSUbase and the services it offers for the Swiss social science research community.

10. RECOMMENDED FURTHER READINGS

- CESSDA Data Management Expert Guide - Online tool developed by the Consortium of European Social Science Data Archives (CESSDA) to help researchers make their data FAIR: https://dmeg.cessda.eu/

- The Open Research Data Policy of the Swiss National Science Foundation: https://www.snf.ch/de/dMILj9t4LNk8NwyR/thema/open-research-data

---

[14] https://www.snf.ch/de/dMILj9t4LNk8NwyR/thema/open-research-data

- Logan et al. (2021) provide a nice overview of the current state of data sharing in educational sciences. They also provide responses to the researchers' most common concerns related to data sharing.

- Tedersoo et al. (2021) have conducted an analysis of how data sharing practices differ across scientific disciplines and heterogeneities in motivations for data sharing across motivations.

- Wilkinson et al. (2016) provide an excellent overview of the FAIR principles.

REFERENCES

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., & Chan, T. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433-1436. https://doi.org/10.1126/science.aaf0918

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PloS one, 15*(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

Drachen, T. M., Ellegaard, O., Larsen, A. V., & Dorch, S. B. F. (2016). Sharing data increases citations. *LIBER Quarterly: The Journal of the Association of European Research Libraries, 26*(2), 67-82. https://doi.org/10.18352/lq.10149

Freese, J., & Peterson, D. (2017, 2017/07/31). Replication in Social Science. *Annual Review of Sociology, 43*(1), 147-165. https://doi.org/10.1146/annurev-soc-060116-053450

Heers, M. (2021). Replication in the Social Sciences. *FORS Guides, 16*, 1-7. https://doi.org/10.24449/FG-2021-00016

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data Sharing in Education Science. *AERA Open, 7*, 23328584211006475. https://doi.org/10.1177/23328584211006475

Nature. (2018). Learning from replication. *Nature Human Behaviour, 2*(9), 601-601. https://doi.org/10.1038/s41562-018-0441-1

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

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PloS one, 2*(3), e308. https://doi.org/10.1371/journal.pone.0000308

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific data, 8*(1), 192. https://doi.org/10.1038/s41597-021-00981-0

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open, 4*(3), 2332858418787466. https://doi.org/10.1177/2332858418787466

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t

Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*, 160018-160018. https://doi.org/10.1038/sdata.2016.18