

Research Data Management for Small Research Projects

A practical guide



German Data Forum (RatSWD)

Research Data Management for Small Research Projects

A practical guide

List of abbreviations

| | |
|------------|---------------------------------------------------------------------------|
| ALLBUS | German General Social Survey |
| CESSDA | Consortium of European Social Science Data Archives |
| CF | Campus Files |
| CIS | Centro de Investigaciones Sociológicas |
| DDI | Data Documentation Initiative |
| DFG | German Research Foundation |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| ELSST | European Language Social Science Thesaurus |
| ERIC | European Research Infrastructure Consortium |
| FAIR | Findable, Accessible, Interoperable, Re-usable |
| FDZ-BO | Research Data Center for Business and Organizational Data |
| FORS | Swiss Centre of Expertise in the Social Sciences |
| GDPR | General Data Protection Regulation of the EU |
| GESIS | GESIS – Leibniz Institute for the Social Sciences |
| GND | Integrated Authority File |
| ICPSR | Inter-university Consortium for Political and Social Research |
| ISSDA | Irish Social Science Data Archive |
| KonsortSWD | Consortium for the Social, Behavioural, Educational and Economic Sciences |
| LD SG | Data protection laws of the German Länder |
| NEPS | German National Educational Panel Study |
| NFDI | German National Research Data Infrastructure |
| PUF | Public Use Files |
| RDC | Research Data Centre |
| RDM | Research Data Management |
| SOCIOSITE | Social Science Information System |
| SOEP | Socio-Economic Panel |
| SUF | Scientific Use Files |
| TOM | Technical and organisational measures |

Contents

| | | |
|-------|--------------------------------------------------------------------|----|
| 1 | Aims and target audience | 7 |
| 2 | Basic principles | 7 |
| 3 | Research data management: step-by-step | 9 |
| 3.1 | Planning phase | 9 |
| 3.1.1 | Primary v. secondary data | 9 |
| 3.1.2 | Diversity of data | 10 |
| 3.1.3 | Issues relating to archiving and re-use of data | 10 |
| 3.1.4 | Data management plan | 11 |
| 3.1.5 | RDM and the contingency of research processes | 12 |
| 3.1.6 | Research ethics: aspects to consider when planning a project | 13 |
| 3.2 | Data collection phase | 14 |
| 3.2.1 | Complying with data protection regulations | 14 |
| 3.2.2 | Storing, backing up and securing research data | 19 |
| 3.2.3 | Organising data | 22 |
| 3.2.4 | Continuously documenting and contextualising data | 24 |
| 3.3 | Preparation and analysis phase | 27 |
| 3.4 | Publication / archiving phase | 29 |
| 3.4.1 | Advantages and requirements of data sharing | 29 |
| 3.4.2 | Finding a suitable data infrastructure | 29 |
| 3.4.3 | Data sharing as a collaborative task | 30 |
| 3.5 | Re-use of research data | 31 |
| 4 | Glossary | 34 |
| 5 | Bibliography | 36 |
| | Contributors | 41 |

List of illustrations

Figure 1: Lifecycle of research data 8

Figure 2: Advantages and disadvantages of frequently-used storage options..... 20

Figure 3: Regulations for use of the GESIS data archive 32

1 Aims and target audience

This guide aims to make it easier for researchers working on social, behavioural and economic sciences projects to get started in *research data management (RDM)*. It focuses on individual research projects or small-scale, temporary project associations that intend to build up a data store for their long-term use and/or to make it available, but which have not yet resorted to using institutional infrastructures. This brochure gives an overview, limiting its focus to data in the social, behavioural and economic sciences. It is application-oriented and does not claim to be comprehensive.

2 Basic principles

This guide aims to make it easier for researchers working on social, behavioural and economic sciences projects to get started in *research data management (RDM)*. It focuses on individual research projects or small-scale, temporary project associations that intend to build up a data store for their long-term use and/or to make it available, but which have not yet resorted to using institutional infrastructures. This brochure gives an overview, limiting its focus to data in the social, behavioural and economic sciences. It is application-oriented and does not claim to be comprehensive.

Depending on the discipline and the collection method, research data can take many different forms. According to the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), they include “for example, measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures and questionnaires. Compilations, software und simulations can likewise constitute a key outcome of academic research and are therefore also included under the term research data” (DFG, 2021, p. 1).

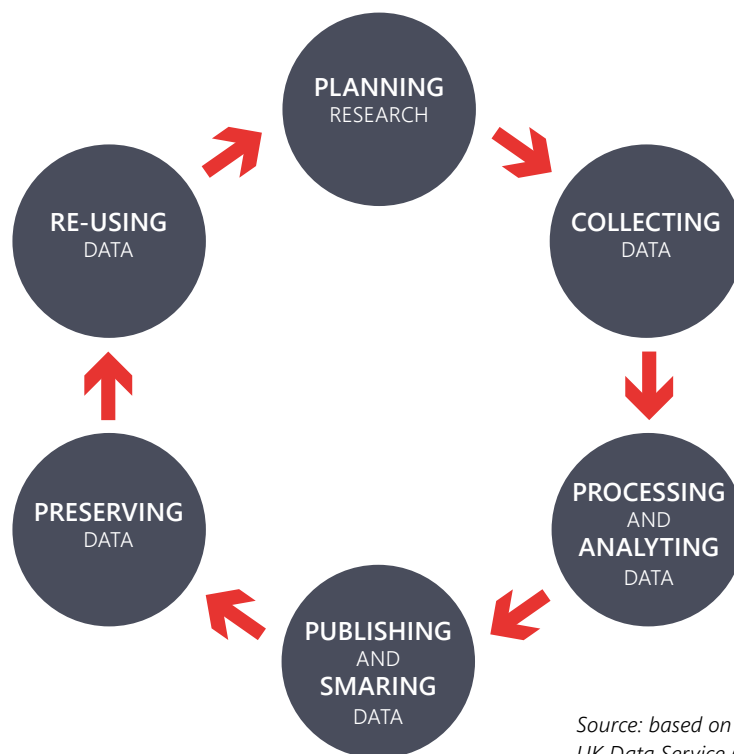
Research data should be handled in accordance with the *FAIR principles*, i.e., if possible, they should be *findable, accessible, interoperable* and *re-usable* (Wilkinson et al., 2016). The four FAIR dimensions represent a continuum rather than either-or options (cf. also Netscher, 2020). In contrast to Open Data (whereby all data are available to everyone), implementing the FAIR principles may also mean that the existence of data is findable using metadata, but access to that data is only permitted to a precisely defined group of potential users subject to certain conditions. Equally, legal and ethical reasons may militate against making the data openly available. Generally, FAIR data should be “as open as possible and as closed as necessary” (European Commission, n.d.)

Well-documented and comprehensible RDM is necessary in order to implement the FAIR principles as well as for other purposes. RDM covers the entire process undergone by research data during a project – from planning the study through data collection, processing and analysis, storing data and metadata, archiving and publishing them, if applicable, up to re-using and if appropriate, deleting the data.

This process can be illustrated by the *lifecycle of research data* (see ► Figure 1) that presents the ideal-typical process and is oriented to a great extent on the management of quantitative data. Deviations from this process are conceivable and sometimes reasonable, under certain conditions. For example, during the planning phase researchers may determine whether suitable data to answer their research question already exist and are accessible, in which case some process stages may be skipped. Equally, an iterative approach may be useful in the collection and analysis of data – as envisaged in Strübing’s *Qualitative Sozialforschung* (2018), for example – or alterations in the data collection process based on pretests. Equally, the publication of data may be restricted when (effective) →anonymisation¹ of qualitative data reduces the data’s interpretability to such an extent that profitable re-use is no longer possible. We have based this guide on an ideal-typical structure, despite its limitations.

¹ Terms marked with an arrow are explained in the glossary (appendix).

Figure 1: Lifecycle of research data



Source: based on *Forschungsdaten.info* (2022a) and the UK Data Service (n.d.-a)

Implementing an active RDM in accordance with the FAIR principles has many advantages. These include both utilitarian arguments from the scientific community's viewpoint and rational arguments from the individual researcher's perspective.

The scientific community considers that generally, academic knowledge increase can be better efficiently achieved, the more it is based on existing knowledge and thus also on existing data. In addition to this aspect of efficiency, however, the focus is on quality assurance of the research processes and findings and thus also of the research data. Active RDM is an essential precondition for reproducibility and traceability of research findings. Sharing research data can reduce the number of data collection activities, thus making more sparing use of survey respondents' willingness to participate, which is a collective good.

Individual researchers can increase their visibility and that of their research by using RDM and implementing the FAIR principles. In addition, work that is otherwise "lying fallow" can contribute to the researcher's reputation through publication and potential citations of the data and the relevant documentation. It becomes easier to initiate collaborations and cooperation runs more smoothly (e.g., thanks to comprehensible documentation and traceability of the data collected). The risk of data loss or unauthorised access to data can also be minimised. Experience with Open Science may also be a criterion for appointing staff in some disciplines, such as psychology and the economic sciences. In some cases, available data can be re-used in university teaching and in degree dissertations. Finally, most third-party funding providers require a data management plan (DMP) as part of the project proposal and expect that data collected using public funds will generally be published or made available to the scientific community.

Active RDM and implementing the FAIR principles, however, also always entails complying with ethical principles and legal regulations. It should not be confused with sharing research data at any price.

3 Research data management: step-by-step

3.1 Planning phase

Before and at the start of a project a range of conceptional issues, research strategies and organisational issues must be clarified based on the intended research objectives, all of which influence the RDM in one way or another. This section covers these aspects.



PLANNING
RESEARCH

3.1.1 Primary v. secondary data²

The first question concerns the type of empirical data required to address the research objectives: should new data be collected or can existing datasets be re-used? While to a certain extent the usual practice is to evaluate parts of national surveys or panels (e.g., in Germany, SOEP, ALLBUS, NEPS) and tailor them to each project's particular objectives, in →qualitative research in particular it is more usual to generate new data material oriented on the current research objectives, i.e., to assume that the field to be examined is specific in nature.

To clarify whether it is necessary to generate fresh data material for a particular project, the researcher requires an overview of the available empirical material from third parties in the target field. This information can be gained from a range of →research data centres (RDCs)³ that →host and →curate a large number of studies, in some cases related to particular topic fields or particular data types. The German National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI)⁴ is currently working to construct central portals that will facilitate the search for relevant secondary datasets and access to the material.

In projects that exclusively use →secondary data from large national or international research data repositories, a number of tasks associated with project related RDM are superfluous or become easier. The tasks involved in collecting fresh data, such as establishing →informed consent from survey participants, as well as the processing and anonymisation or →pseudonymisation of the data material, have already been done. In addition, the decision on whether or not make one's own material available for re-use by later researchers has already been partly settled because the data themselves are already published, so the researcher only faces the decision about making syntax files, code lists or similar of their own evaluations available.

The situation is different when fresh data material is collected for a particular project: in this case it is necessary to clarify at an early stage what types of data will emerge (e.g. →transcripts of qualitative interviews, observation records, documents from the field, survey data, →vignette data, video material, audio documents, social media feeds, physiological measurements, psychometric test procedures), how the material is to be collected and processed, and what data protection and research ethics requirements ensue.

² The terms primary and secondary data are sometimes applied differently in social, behavioural and economic science disciplines. In this guide, primary data are understood as data that are used in the context in which they were collected and secondary data as data that are re-used outside their original context.

³ You can find an overview of RatSWD-accredited RDCs on the KonsortSWD website: <https://www.konsortswd.de/en/services/research/all-datacentres/>

⁴ <https://www.nfdi.de/?lang=en>

3.1.2 Diversity of data

The diversity of both research designs and the data collected or used for them requires a number of decisions necessary when planning a research project. These decisions have a considerable influence on how the study progresses, but also on the re-use potential of the data generated for the study. Even if not all questions have found final answers when the research begins, it is necessary and useful to address them from the very start. The point of planning is generally not the plan itself but the planning process, that is, an active dimension of continually reflecting on the process. Based on well-considered planning, new demands that arise later in the process, e.g., variations in sampling, the inclusion of additional case domains, focusing on partial aspects of the originally intended study or adding further data types, can be justified and systematically incorporated.

To sum up, a range of issues and actual tasks relating to RDM already arise at the research planning stage.

3.1.3 Issues relating to the preservation and re-use of data

One issue must be clarified for every study: how the empirical data gained for the project are to be preserved in the long term, i.e. how they are to be archived, and whether and if so how these data can and should be made available, partially or completely, not only for quality checks in accordance with the rules of good scientific practice but also for re-use by third parties (e.g. for follow-up research or as teaching material). Research funding organisations such as the DFG now require that funding applicants consider this question and if they reject archiving of their data to enable their re-use, that they give well-founded reasons for this decision (cf. Open Science Future, n.d.). In other words, researchers are not obliged to make their data available for re-use, but they must give sound reasons why re-use is either possible and intended or excluded or significantly restricted.⁵ At the same time, research funders ensure that well-founded exclusions or agreements to data re-use options do not play a role in grant decisions.



The following issues must be clarified relating to preserving and re-using data

- Does the informed consent clearly cover the possibility of passing on the (anonymised) research data beyond the current project to other entities (RDCs or specific or not yet defined follow-up users)?
- Is there a risk that the intention to pass on data stated in the informed consent may make rapport with and access to the research field more difficult or at worst, may even make research impossible? (Primacy of primary research applies here).
- What measures are necessary to enable the data to be passed on? For example, must the material be partially or completely anonymised or pseudonymised beyond the normal requirements of the research project? Are the time and staff resources for the necessary measures planned as part of the project and is the effort required reasonable?
- Is the effort of preparing and transferring the data reasonable in relation to the anticipated added value for research and teaching?
- Is it likely that, when prepared for an ethically and legally acceptable transfer, the material will still be useful for further research (e.g. blurring of video data and photo material)?
- Which →repository or RDC (if any) can host the dataset after project completion?
- What support in preparing material and making it available can each RDC provide and on what conditions?

⁵ See also Stellungnahme des DFG-Fachkollegiums 111 „Sozialwissenschaften“ (DFG - Fachkollegium 111 „Sozialwissenschaften“, 2020) (in German).

3.1.4 Data management plan (DMP)

The DMP is the document in which all these considerations are systematically documented. It is a relatively new instrument and not yet firmly established in many fields of empirical social and economic sciences research, particularly for small-scale projects. However, nowadays research funders (e.g. Horizon Europe) increasingly require the submission of a DMP as part of a project proposal. The DFG does not demand a DMP in their funding information but does require information on data handling: "In principle, all funding programmes that describe a concrete work programme for a research project must include a description of how the research data will be handled." (DFG, 2022b).

Even apart from these regulations, working with a DMP is extremely conducive to competent, efficient project implementation. The DMP has a range of functions:



The DMP

- promotes systematic reflection about the generation, use and inclusion of empirical research data in the research process.
- promotes transparency about data generation and use.
- encourages communication between different institutional actors in each research context (in particular: research funding institutions, project-implementing bodies, researchers and RDCs).
- provides a reason to address at an early stage the issues of long-term secure storage and possible forms of re-use of the data collected or used during the research process.
- constitutes at best a →"living document" that is a flexible part of the data use processes, continuously documenting changes, but does not produce fixed commitments that may hinder practical research.

What does a DMP contain?

There are no general standards defining precisely how a DMP is designed; the specifications vary depending on the funding institution.⁶ Most importantly, however, a DMP should contain the following information:



- a. *Specifications of the data material* to be used, collected and/or generated in the project. The following issues should be clarified:
 1. What type(s) of data will be used and by which methods will they be collected?
 2. Does the project team intend to use existing data and what conditions of use apply?
 3. In what ways will existing data be made accessible?
 4. Will different datasets from a range of sources be linked to each other?
 5. Which quality assurance measures will be applied to data generation?
- b. Information about handling research data *during the project*:
 1. How will the data be described to ensure traceability (metadata, documentation standards)?
 2. Are there any legal obligations and other conditions to be complied with relating to the data that could restrict accessibility or influence later publication options?
 3. How will the data, particularly sensitive data, be stored and secured while the project is running? How is access organised?

⁶ The Open Science Magazine of the ZBW – Leibniz Information Centre for Economics (Open Science Future, n.d.) gives an overview of research funding institutions' different requirements and guidelines.

4. Who is responsible for the correct handling of the data during and after the project?
5. Which resources are necessary for data preparation and archiving?
- c. Information about dealing with research data *after project completion*:
 1. Is there a plan to provide options for re-using the data collected?
 2. If so, how will the data be made accessible?
 - i. In a minimum version, i.e. depositing the dataset with minimal documentation and contextualisation in an (e.g. local) repository or another institution of research data infrastructure.
 - ii. As a systematically prepared and archived dataset including comprehensive documentation making it serviceable for practical re-use, e.g., in replications, accessible either from the data producers or via an RDC.
 - iii. As a dataset prepared specifically for potential secondary analyses and equipped with metadata, to be passed to an RDC.

Source: cf. RatSWD (2016) and Science Europe (2021)

Depending on the type of data and the circumstances, preparing data for archiving and particularly for possible re-use can require considerable effort and expense that must be considered when applying for resources. The DFG is one institution that permits researchers to make a separate application for the additional staff and material expenses required for this purpose. However, making an appropriate estimate of the necessary costs can be difficult, particularly in small projects or those where researchers are not so familiar with RDM. In this context it can be helpful to make early contact in the planning phase with an RDC that may receive the data later, asking for support in calculating the expense. The RDCs' experts for curating and preparing research data have a good overview of the tasks entailed over the whole lifecycle of archived data and can estimate the expense by making comparisons with other projects.

3.1.5 RDM and the contingency of research processes

The precise definition and rounding off of the body of data often takes place gradually during the project and depends on initial findings (e.g. in the course of →"theoretical sampling"). This applies in particular to a range of designs used in qualitative social research (e.g. Grounded Theory, documentary analysis, ethnographic field research or discourse analysis). Therefore, it is important a) to define the DMP presented with the project proposal as provisional and b) to update the DMP continuously during the project – a measure that also corresponds to the idea of a DMP being a "living document".



Useful links and support for **setting up a data management plan**

- Adapt your Data Management Plan. A list of Data Management Questions based on the Expert Tour Guide on Data Management (CESSDA, 2019).
- Der Datenmanagementplan. Eine Wegbeschreibung für Daten (forschungsdaten.info, 2022b). (German)
- Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen (Gollwitzer et al., 2020). (German)
- Elements of a Data Management Plan (ICPSR, n.d.-a).
- Practical Guide to the International Alignment of Research Data Management – Extended Edition (Science Europe, 2021).
- Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung (VerbundFDB, 2015). (German)



Useful tools for setting up a data management plan

For project proposals in *Germany*:

- Standardisierter Datenmanagementplan für die Bildungsforschung (Stamp) (VerbundFDB, 2022b). (German)
- Research Data Management Organiser (RDMO Arbeitsgemeinschaft, n.d.).

For project proposals for *Horizon Europe*:

- Argos – Plan and follow your data (Argos, n.d.).
- DMPonline – Plan to make data work for you (DMPonline, n.d.).
- Data Stewardship Wizard (Pergl et al., 2019).

For collaborative projects with researchers from the *USA*:

- DMPTool – Build your Data Management Plan (DMPTool, n.d.).

3.1.6 Research ethics: aspects to consider when planning a project

Right at the start of a research project it is necessary to determine whether an →ethics committee approval is required for applying or implementing the project, or for the publication of its findings.



Explanation: universities, research funding institutions and increasingly, also professional journals expect ethics committee approval for research on sensitive topics and/or with vulnerable groups. This approval should competently estimate what possible unintended side effects the planned study could entail for participants or the groups/milieus they represent, whether these risks seem acceptable in view of the anticipated usefulness of the research, and what measures are envisaged to avoid or mitigate negative effects particularly for vulnerable groups. Because the infrastructure of competent professional ethics committees is currently rather rudimentary in the social sciences, it is important for the project to clarify whether approval is required, and which ethics committee would be suitable in each case.

While the significance of ethical considerations and their consistent implementation in research activities is indisputable, there is disagreement about whether ethics approval or establishing blanket coverage with ethics committees is justified.⁷ Nevertheless, institutional pressure from research funders and publishers increasingly may make it necessary to gain the relevant ethics committee approval.

There is no general answer to the question of which ethics committee researchers can turn to when they either doubt the ethical acceptability of their project or face the above-mentioned institutional expectations. This is because the ethics committee infrastructure varies widely depending on the discipline, research institution and funding structure (e.g. for interdisciplinary projects). To give a professional assessment of possible ethical problems, the committee must have both the methodological and specialist skills in the relevant discipline and sufficient knowledge of the ethical and legal conditions. That is not always the case, particularly in the case of projects using highly specialised research methods. Therefore, it is advisable first to seek a more-or-less relevant ethics committee in the project's own research institution and if no such committee exists, then to contact the relevant professional association. It is less advisable to fall back on a committee that is not closely related to the relevant field, e.g., the committees that exist in every medical faculty, due to the risk of inappropriate assessments by a committee that lacks the expertise to assess the methodology with sufficient sensitivity.

⁷ On the importance of reflecting on research ethics, see e.g. Friedrichs (2014) and Hopf (2016) and for a critical view e.g. Hammersley (2009).



Useful links and publications on the topic **research ethics**

- FAQ on ethics committee approval: Humanities and Social Sciences (DFG, 2022a).
- Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften (RatSWD, 2017). (German)
- Best Practices for Research Ethics portal for Social and Economic Sciences (RatSWD, n.d.–a).



3.2 Data collection phase

Before collecting the data, most research projects plan a preparation phase that should be used to set up an efficient, sustainable data management process. This section covers in more detail some important fields in which organisational measures must be arranged and research strategy decisions taken in advance of the actual project work.

3.2.1 Complying with data protection regulations

In the social, behavioural and economic sciences, the research object is usually people, meaning that many projects in these fields entail the processing of personal data. How to handle personal data is covered by the EU's General Data Protection Regulation (GDPR), defined in more detail in Germany by the supplementary regulations of the Federal Data Protection Act (BDSG), the data protection legislation of the *Länder* (LDSG), and specialised legislation (e.g. legislation for schools or statistics). Researchers who deal with personal data must comply with these fundamental legal requirements (Art. 2 (1) GDPR).

What are personal data?

“Personal data” means “any information relating to an identified or identifiable natural person” (Art. 4 (1) GDPR). For a person to be *identifiable*, the decisive factor is whether the processed data can be linked to the person. Correlation with identification numbers such as a social insurance number, location data such as an address, or online identifiers such as an IP address, for example, are sufficient to count as personal data. Another example would be if a person is identifiable through the combination of several factors that are “specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (Art. 4 (1) GDPR).



Basically, whenever third parties could discover with a certain (reasonable) amount of effort the people to whom the data refer, these data are covered by the data protection regulation. However, these regulations do not apply to anonymous or completely anonymised data which cannot (or no longer) be linked to individual people even with a great deal of effort.

What are special categories of personal data?

In addition, there are special categories of personal data that are also known as sensitive data. They are considered worthy of special protection, so handling such data requires complying with a higher level of technological, organisational and legislative demands.



These special categories include data containing information on the following topics (Art. 9 (1) and Art. 10 GDPR):

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Genetic and biometric data
- Health-related data
- Data concerning a natural person's sex life or sexual orientation
- Data relating to criminal offences, convictions and related security measures

What measures are necessary when working with personal data?

How extensive the measures must be is dependent on the individual research object and design. Here we present some *fundamental measures* that all researchers working with personal data should take in most cases. These explanations are expressly not to be understood as legally binding; they are simply recommendations for handling personal data.

(1) Determine the controller(s)

The controller, the person(s) or body responsible for handling personal data, should be determined at the latest by the start of a research project. According to the GDPR, the controller is the person who “determines the purposes and means of the processing of personal data” (Art. 4 (7) GDPR). In the case of an individual project, this is usually the project management. In a collaborative project, the project management almost always bears joint responsibility. In this case, an *arrangement regarding joint responsibility* is required that sets down in a transparent manner their respective responsibilities for compliance with their data protection obligations (Art. 26 (1) GDPR). The essence of this arrangement shall be made available to the data subject (Art. 26 (2) GDPR). The academic institution almost always bears responsibility for representing the project to external data protection authorities, however. The project management is thus to be seen as a responsible (sub) body within the academic institution that is responsible for implementing the stipulated data protection measures (Meyermann & Porzelt, 2019, p. 32). For this reason, all documents relevant to data protection should always be agreed with the institution's data protection officer and passed on for the officer's information.

(2) Creating a record of processing activities

Whenever personal data are processed in research projects, an additional *→record of processing activities* (Art. 30 (1) GDPR) is to be created for documentation purposes at the start of the project. This gives information on the type of personal data that may be processed by which individuals (or groups) and for what purpose. It also reveals the maximum storage time, who has access to the data, and what measures have been implemented to protect the data. This written record fulfils the GDPR accountability obligation to comply with the protection regulations (Art. 5 (2) GDPR). If a state supervisory authority makes enquiries on reasonable grounds, the researchers can present this record.

(3) Obtaining consent

Processing personal data is only lawful when it has a *legal basis* (Art. 6 GDPR). The most commonly applied legal basis for scientific projects is informed consent (Art. 6 (1a) GDPR). This means that the data subjects have been given sufficient information about the data processing as the basis on which they give their consent to the collection, storing and making available of their data, and for the data's (long-term) preservation and preparation for (academic) re-use, if planned. This consent must be *freely given*, for a specific case, and *unambiguously expressed* (Art. 4 (11) GDPR). The data subject has the right to withdraw consent at any time without explanation (Art. 7 (3) GDPR). In this case the data are to be deleted immediately. The data processing activities up to the point when consent was withdrawn are not affected by the withdrawal.

A voluntary decision presupposes a *capacity to consent* (indirectly derived from Art. 8 GDPR). This means that the person must be able to understand the information given and to assess the consequences of their consent. This capacity is assumed in persons of full age with legal capacity, so individual examinations must only take place in justified cases. For children under the age of 16, however, consent must generally be obtained from the holder of parental responsibility over the child *in addition to or instead of* the child's consent (Art. 8 (2) GDPR; Meyermann & Porzelt, 2019, p. 15).

For consent to come into effect, its *form and content* must also comply with certain requirements. Although consent can be given orally or in other forms, it is recommended to opt for consent *in writing*, unless that would (considerably) detract from the research objective. The fact is that the GDPR places the burden of proof on researchers. In case of doubt, the researchers must be able to demonstrate clearly that the study participants were informed and have consented to the processing of their data (Art. 7 (1) GDPR). A signature under a declaration of consent is the safest method of proof. If the research objective does not permit written consent because this could negatively influence the interaction process in the field, another method must be considered by which consent can be documented in the form of a clear affirmative action (e.g. by detailed field records, audio or video recordings; RatSWD, 2017, p. 23).



Formal condition that (written) consent and the related information section shall fulfil:

- clear, plain language appropriate to the target group (Art. 7 (2) GDPR)
- an intelligible and easily accessible form (Art. 7 (2) GDPR)
- comprehensive information presented in a form that is clearly separate or distinguishable from the declaration of consent (Art. 7 (2) GDPR)



Information to be provided to the data subjects in the *declaration of consent* in accordance with Art. 13 GDPR:

- the name and contact details of the controller and, where applicable, of the controller's representative and of the data protection officer
- the purpose(s) of the data processing
- the legal basis for the data processing (consent in accordance with Art. 6 (1a) GDPR); in case of sensitive data, Art. 9 (2a) GDPR is also relevant)
- information about (or criteria for) the period for which the data will be stored (Art. 13 (2a) GDPR)
- information on the data subject's main rights in relation to their data: right of access (Art. 15 GDPR), right to rectification (Art. 16 GDPR), right to erasure (Art. 17 GDPR), right to restriction of processing (Art. 18 GDPR), right to object (Art. 21 GDPR)
- the voluntary nature of the data subject's participation (Art. 4 (11) GDPR) where the provision of personal data is not a statutory or contractual requirement (Art 13 (2e) GDPR)
- the right to deny or withdraw consent at any time without giving reasons (Art. 7 (3) GDPR), without incurring any disadvantages; data processing up to the time of withdrawal of consent remains lawful
- information about the right to lodge a complaint with a supervisory authority (Art. 77 GDPR)
- the planned transfer of data to countries outside the EU, if applicable
- the recipients or categories of recipients of personal data, if applicable



Further information to be included in the declaration of consent if it applies to the project:

- If the project plans to archive the data outside the direct project context, or if their availability for (academic) re-use is planned, the data subject shall be informed.
- If the personal data have not been directly obtained from the data subject but come from other sources, the declaration of consent must also give information about the origin and categories of personal data to be processed in accordance with Art. 14 GDPR.
- If special categories of personal data in accordance with Art. 9 (1) GDPR are to be processed these must also be expressly listed in the declaration of consent (in accordance with Art. 9 (2a) GDPR).

Not all information must always be precisely described. For example, if true information about the research objective would negatively influence data collection (e.g. due to response bias by respondents in a survey), more general formulations are permissible that do not conflict with the study's scientific aims. In addition, differing ethical standards have developed to some extent in different disciplines in relation to how far deception in describing the research objective is permissible for the benefit of a significant knowledge gain (RatSWD, 2017, pp. 21–22). On study completion, respondents must be informed about any deceptions and must have the right to withdraw their data.

If the data processing objective changes or expands during the project and the data subjects have not already been informed about this and given their consent, additional informed consent must be obtained (Art. 13 (3) GDPR and Art. 14 (4) GDPR).

Declarations of consent may be independently prepared for a particular project. However, it is also possible to use existing templates and adapt them for the relevant project. Some examples follow:



Templates and support for preparing informed consent forms

- Consent Form Wizard (CFW) - Ein Werkzeug zur Erstellung von Einwilligungserklärungen (DARIAH-EU ELDAH, 2021).
- *Ethisches Handeln in der psychologischen Forschung. Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* (DGPs, 2018). (German)
- Website with templates for informed consent forms (FDZ Qualiservice, n.d.).
- Formulierungsbeispiele für „informierte Einwilligungen“ (VerbundFDB, 2018a). (German)
- Formulierungsbeispiele für „informierte Einwilligungen“ in leichter Sprache (VerbundFDB, 2018b). (German)
- Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen (VerbundFDB, 2019). (German)

(4) Technical and organisational measures

Personal data must be appropriately secured by *technical and organisational measures (TOM)*. Which measures are considered appropriate depends on “the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the varying likelihood and severity of risks” (Art. 32 (1) GDPR and Art. 24 (1) GDPR). In addition to pseudonymisation and anonymisation, explained below, the following TOMs may be implemented, for example:

- encryption of files and storage systems
- use of secure passwords or pass phrases
- specification of storage duration and deletion of data that are no longer needed
- limitations to and checks on staff access
- storage of personal data within the EU
- confidentiality agreements and agreements to comply with data protection

Other possible TOMs are listed in section 3.2.2 under data security. The measures should be regularly tested, assessed and evaluated to ensure that they guarantee secure data processing (Art. 32 (1d) GDPR).

(5) Pseudonymising and anonymising research data

According to the principle of data minimisation (Art. 5 (1c) GDPR), only as many personal data as necessary should be processed to achieve the research objectives. Two important methods of achieving the most economical use of personal data possible are pseudonymising and anonymising. Accordingly, data protection regulations also require that research data be anonymised or pseudonymised if and as soon the research objective permits (Art. 89 (1) sentences 3 and 4 GDPR; Section 27 (3) sentences 1 and 2 Federal Data Protection Act (BDSG).

Pseudonymisation (in accordance with Art. 4 (5) GDPR) entails substituting personal details one-to-one, e.g. using an alias, a code or by giving comparable details (the pseudonym). However, a reference list, allocation table or record exists that enables the pseudonyms to be reconnected to the original data. Even if the allocation document is secured and stored separately from the data (e.g. with a trustworthy →data trustees), the data can thus theoretically be re-identified. Therefore, pseudonymised data are still personal data to which data protection regulations apply. However, the personal references in these data are much reduced, so less restrictive access controls are permitted during the work process. Generally, data can be pseudonymised at the latest once data collection is complete. In some cases it is possible to work with pseudonyms as early as the field phase (e.g. personal identifiers on panel studies).

Anonymisation is not defined in the GDPR. From Recital 26 GDPR it can be deduced that statements are based on effective anonymisation. Hence data are considered to be anonymised if unreasonable effort would be entailed in relating them to a natural person (RatSWD 2020, p.18). According to this concept, re-identifying the person is not completely impossible, but so laborious in practice that it is considered extremely unlikely. Therefore, data protection regulations no longer apply to these data. Most of the techniques used to remove the relationship to the person from the data either reduce or alter the information content.



Important *anonymisation techniques* are:

- deleting identifying features (including the key list, allocation table or anonymisation records)
- aggregation of information (e.g. grouping values of a feature in more general categories; reducing detailed personal information by replacing it with more abstract information; rounding; generalising values)
- masking (i.e. original values are replaced by random but technically valid new values; it is important that the new values have a “comparable meaning” so that e.g. the meaning of a text in qualitative research or the information content of a distribution in quantitative research are maintained^{d)})

Generally, data are anonymised using a combination of techniques. While automatised regulation can sometimes be applied to the whole dataset in the case of quantitative data, the manual effort is much greater with qualitative data, where in general all the data must be viewed, the risk of possible re-identification estimated and then anonymised. This can be a huge challenge with dense, detailed qualitative data. Currently, important steps are being taken to develop powerful (semi-)automated processes to anonymise qualitative data (e.g. the anonymising tool QualiAnon, cf. Nicolai et al., 2021). In addition, when handling qualitative data, the issue arises of whether the anonymisation process leads to so great a loss of information that possibilities for subsequent analysis by third parties are too greatly reduced. Every researcher must answer this question individually for their own study.



Useful tips and recommendations for **anonymising research data** (German)

- Empfehlungen zur Anonymisierung quantitativer Daten (Ebel, 2015).
- Hinweise zur Anonymisierung von quantitativen Daten (Ebel & Meyermann, 2015).
- Hinweise zur Anonymisierung qualitativer Daten (Meyermann & Porzelt, 2019).



Useful tool for **pseudonymising or anonymising research data**

- QualiAnon - Qualiservice tool for anonymizing text data (Nicolai et al., 2021). For information on registration/activation see: <https://www.qualiservice.org/en/the-helpdesk/tools.html>

3.2.2 Storing, backing up and securing research data

The data that are collected, processed and analysed and the relevant documentation are the core element of every research project. Therefore, when the project starts, at the latest, it is essential to develop a concept by which digital data and documentation can be securely saved during the project phase and data loss and unauthorised access can be prevented.

Selecting a storage location

Selecting a suitable storage location and if necessary, rights management depends heavily on the individual requirements of each project. In the following we list some aspects that offer orientation when selecting a storage solution.



Important aspects to consider when selecting a suitable storage solution

Storage capacity: how much data will accumulate during the project? Which storage systems could provide the necessary capacity?

Data access: who needs access to the data and how will access be provided? If people from another academic institution are collaborating on the project, a storage system that enables external access may be necessary.

Teamwork: will it be necessary to work simultaneously on the data? If so, a storage system with a Sync&Share function may be useful.

Storage location: where are the data? Are they on the project institution's own servers, on servers in Europe or in non-EU countries (e.g. in the USA)? This is an important issue in relation to personal data because the data protection regulations of the country where the data server is located apply. The GDPR only permits data storage outside the EU if the data subjects have given their consent or if the country provides an equivalent level of data protection. It is advisable to use a local storage option in the project's own institution which is subject to internal tests and generally complies with European data protection regulations.

Access restrictions: if the project uses personal data, the storage system should provide options to limit access to individual files. In these cases it may also be necessary to store the data in a specially protected location that only permits access to very few people, and only to transfer them to a joint storage system later in the research process once they are anonymised.

Backup: does the storage system offer regular automatic backup, which helps prevent data loss? In which intervals is the backup carried out and how long is the storage duration for each backup?

Storage duration: how long will the storage be needed? Many storage media have very short lifespans. Some of them (e.g. external hard drives, recorded DVDs or CDs) do not even fulfil the demand of good scientific practice to guarantee secure storage for at least ten years.

Budget: how much money is available for storage and backup and how much do the different storage solutions including backup cost?

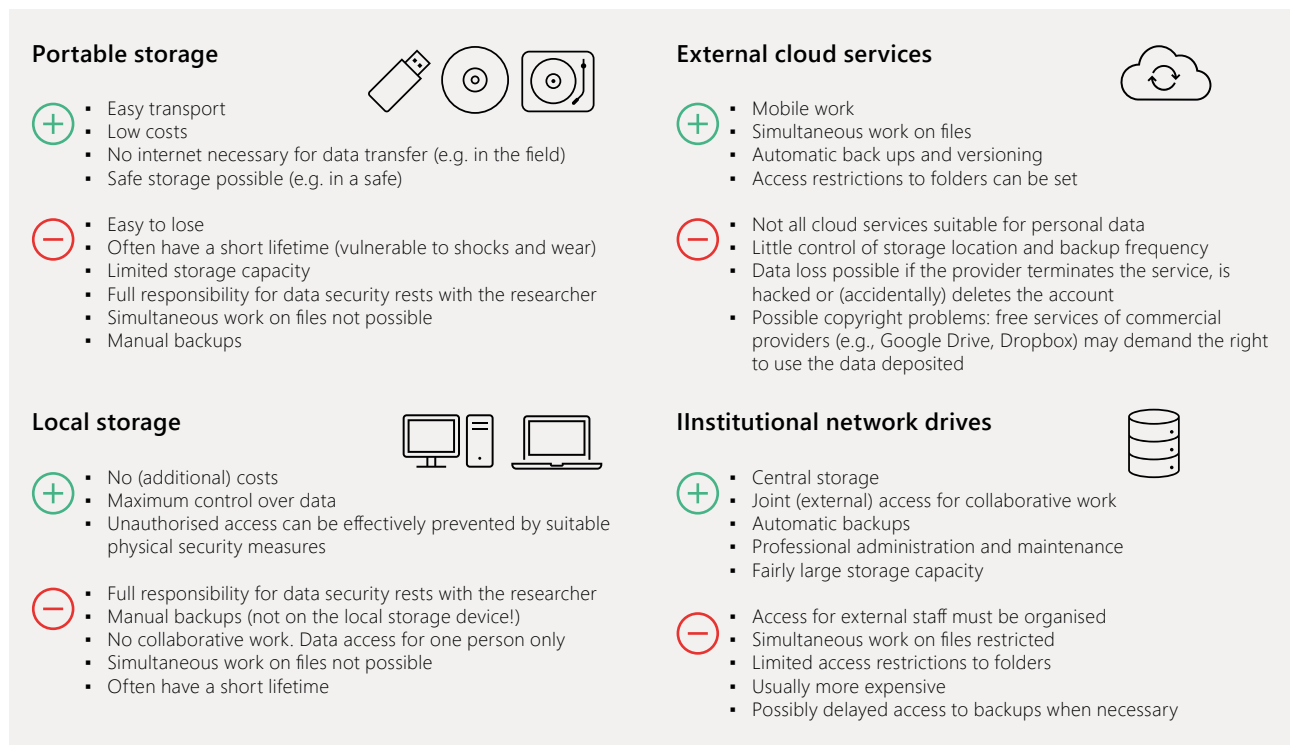
Which storage options are available and what are their advantages and disadvantages?

There is a wide range of storage options. One or more of the following storage devices are often selected:

- portable storage devices (e.g., USB sticks, external hard drive, CD/DVD)
- local storage in one's own working space (mainly desktop PC, notebook)
- cloud services (including commercial or non-commercial providers, or external or local clouds hosted at the project institution)
- institutional network drives

The following illustration, ▶ Figure 2, gives a brief overview of the main advantages and disadvantages of these four storage options.

Figure 2: Advantages and disadvantages of frequently used storage options



Source: authors' presentation based on CESSDA Training Team (2020) and Eisentraut & Hutflasz (2021)

Designing a backup strategy

It is essential to create backups of all files accumulated during the research process. Regular backups protect against data loss that may occur due to e.g. faulty hardware or software, viruses or hacker attacks as well as individual application errors (e.g. overwriting or accidentally deleting data) or natural disasters (e.g. fire, flood, earthquake). If any of the above events occurs, backups save a great deal of time and money.

What is a good backup?

In the best case, backups are carried out automatically and regularly, so that a current backup copy is always available. A good backup system not only saves the newest version of the data; it also retains older versions that may be fallen back on if necessary: for example, if a processing error in the data is only noticed at a later point or when individual work steps must be reconstructed while searching for errors in the data. The backup should include not only data, but also save (non-proprietary) software applications necessary for the reproducibility of the data and research findings at regular intervals whenever possible.



Tips for developing a good backup strategy

Define responsibilities: a contact person responsible for the topic should be named and defined. Project members' roles in the backup process (e.g. carrying out manual backups, checking backups) should also be specified.

Apply the 3-2-1 rule: at least three data copies should be installed on at least two types of storage system. At least one of the storage systems should be at an external location. If storage amounts are very large or the data very sensitive, it may in some cases be necessary to work with fewer than three copies, but these should then be particularly well secured against data loss.

Regular, automatic backup systems are preferable: in the best case, a backup is carried out automatically at regular, fairly short intervals, so that a current backup copy is always available and older versions are retained that can be used for reconstruction processes in case of errors discovered at a late stage. This makes researchers' work easier and prevents data losses. If there are good reasons for selecting a storage system that requires manual backups, routines should be established to specify at what intervals and in what form (i.e. backup variants) backup copies are to be created.

Use services of the researchers' own institution: academic institutions usually offer storage services that include integrated, automatic, regular backups. These services have usually been checked by the research institution to ensure that they comply with data protection regulations.

Protect personal data: backup data that include personal data must be protected in the same way as original data. Therefore, the same protection measures should be implemented (e.g. encrypting or restricting access to the data).

Monitor backups: if the backup system permits, regular tests should be carried out to check that the data can be restored, to ensure data integrity.

Define the maximum storage duration for backups: backups are part of the deletion concept for research data. Right at the start, the maximum storage duration for backups should be defined. In particular, personal data should be reliably and finally erased from all backup storage repositories at the latest on project completion, (e.g. using deletion software such as dban, Eraser or BCWipe (cf. Art. 17 (1a) GDPR).

Improving data security

All researchers should take measures to protect their digital data. If they do not, unauthorised persons may gain access to the data and (deliberately or accidentally) manipulate them, delete them, use them illegally or pass them on to third parties. Research projects that work with personal data have an even greater responsibility in this regard, particularly when dealing with sensitive personal information. The data protection regulations for this context stipulate that researchers must protect personal data with specific *technical and organisational measures (TOM)*.

**Tips for increasing data security on the project****Technological measures**

- Use secure passwords for computer and storage systems
- Encrypt data and hard drives
- Use anti-virus software and activate firewalls
- Carry out regular (security) updates for software and operating systems
- Do not use open unencrypted WLAN connections
- Avoid sending data via email
- Delete data that are no longer needed using a secure method (e.g. erasure software such as dban, Eraser or BCWipe, physically destroying portable storage systems or shredding paper documents)

Organisational methods**Physical security methods**

- Store portable storage media in secure, lockable locations (e.g. safes, document boxes)
- Set up automatic desktop blockers
- Lock doors and windows when not present

Access restrictions/ controls

- Introduce a rights management system
- Reduce numbers of people with access rights to a minimum

Sensitising, rules and communication in the project team

- Develop checklists of technological measures
- Raise awareness and active communication on these topics within the project team
- Staff who work with personal or confidential research data should have an obligation to comply with special rules of behaviour when dealing with data. This may be regulated in the work contract

3.2.3 Organising data

At the start of every project, researchers should consider how to organise the data in the most efficient way for day-to-day project work. At the same time, they should ensure that the data organisation is as sustainable as possible so that if archiving and (restricted) access to the research data in a repository or an RDC are planned, the additional effort required at the end of the project is kept to a minimum.

Designing a folder structure

A good folder structure means that a folder index including files can be passed on to other researchers who can independently infer what was done in the study and identify the individual research steps. Once this condition is met, the folder structure should also ensure that the primary researchers can maintain a good overview of their files and find them easily in their day-to-day work.



Tips for creating a folder structure

If Best Practice Rules already exist in the relevant discipline, these should be used.

Basically, folder structures should:

1. be *hierarchical* in design,
2. include a *maximum of 3 layers of sub-folders*,
3. have a *clear, consistent* structure.

If qualitative research is being carried out, it is advisable to organise the data so that complete *traceability* of the findings is guaranteed.

A *readme file* containing all the rules for the folder structure should be set up.

For project teams, it is worth establishing *responsibilities* in advance: i.e. who sets up new folder structures in joint directories when required and who regulates access to the folders.

Naming conventions for files and folders

Clearly established naming conventions do not currently exist in the social, behavioural and economic sciences. However, there are a few general hints that may be helpful when developing rules for naming folders and files.



Tips for naming files and folders

Naming conventions should be applied *consistently*. Unified naming rules for folders and files over the whole research process make it easier to find things and maintain an overview. "Consistent" means that the same set elements are always used in the same order in file names (e.g. date, key words on content, name abbreviation of person working on the file, version: YYYYMMDD_sampleX_ag_v001).

File names must be *machine-readable*. Therefore, avoid spaces and special characters (e.g. \$, @, %, #, &, *, (), !, /, ?) because not all computer systems can read them correctly. Instead, folder and file names should consist of the letters *a-z*, the numbers *0-9* and underscore (*_*). Capital letters and hyphens (*-*) may be used as long as the software can distinguish between capitals and lowercase, or between hyphen and underscore. Similarly to programming languages, *four variations* are often recommended for file names: →snake_case, →kebab-case, →camelCase und →PascalCase.

The name should describe the *content*. The better the name reflects the file or folder content, the easier it is for human users to understand it. This makes it easier to find files within the structure and increases efficiency. In addition, all metadata that are already contained in the file name do not need to be explained elsewhere.

The names selected should be *as short as possible but as long as necessary* to enable the reader to understand the content of the file or folder. A rough guide value for the maximum length of a file name is 32 characters. Folder names should be much shorter. For orientation, most operating systems restrict the length of a file name including folder path to no more than 255 characters.

File names should be selected so that they *automatically* lead to a *sensible arrangement of files in folders*. Because the order of elements in file names determines how the files are arranged, the features according to which the files are to be sorted should always be placed first in the name. Numbers are Therefore, often placed first in file names to reflect the chronological order, making it easy to find the most recent file. This is usually the date the file was created: YYYYMMDD (e.g. 20220420).

The naming conventions should match the researchers' *own work style*. Researchers work daily with the data and best know the special features of the project and data. Therefore, they should be the ones to decide on the names they find most useful.

The *naming rules* should be *documented* in a text file (readme file). All team members should have access to the naming rules and apply them consistently.

3.2.4 Continuously documenting and contextualising data

Good *documentation or contextualisation* is an essential condition for the (re-)use of research data. This means that researchers present and explain the collection and processing of data in such a way that other researchers can follow the process and evaluate the data if necessary. In other words, the documentation should help secondary data users to “understand and apply the data as well as possible, i.e. to “put themselves into the primary researcher’s shoes’, although they do not share the immediate experience of actually collecting the data” (Steinhardt et al., 2020, p. 18, author’s translation). In qualitative research this is also referred to as “contextualising”, because the interpretative approach of qualitative research always aims at understanding the sense, and the sense emerges in relation to the context(s). Therefore, data must be supplemented with the relevant contextual information. For both qualitative and quantitative data, good research documentation is generally a core quality criterion on which the transparency of research findings and re-usability of data depends. Therefore, researchers should pay particular attention to this point.

The *advantages* of careful, detailed research documentation are obvious, but in practice, documentation is often neglected in favour of other apparently more important tasks. However, it is not merely an element of good scientific practice in that it creates transparency and traceability for third parties. Good research documentation also contributes to visibility of research data – and thus to their use and citation. It also offers opportunities to improve the position of the often-neglected field of data collection for the academic reputation and to increase appreciation of performance in this context. Finally, good research documentation is helpful for the data-generating project itself, in that information on the research process is recorded and can therefore be passed on within the project and systematically used for evaluations.

Research documentation *procedure* has hardly been touched on in social sciences methodology teaching. It may vary considerably depending on the discipline and the data and may take place or be required on different levels, with a range of procedures and at different levels of detail. A conceptual distinction is often drawn between documentation at the level of the whole project (*study level*) and of individual processes of data collection, e.g. an interview (*data level*). In addition, a superordinate level to the research project may be considered that includes the contemporary cultural, regional, etc., context, which may become less familiar with increasing temporal and physical distance from the project.

The *main aim* of documentation or contextualisation is always to make the generation of the data transparent (in the long term). Various presentation options exist, of which the three most widespread are explained in the following paragraphs: using metadata, additional research or context materials, or a report. The extent of documentation may vary depending on the data: “Thus the calculation code (e.g. do-file, sps-file) may be sufficient to show a later user how a dataset was created on the basis of freely available and well-documented primary data” (ZBW et al., 2014, p. 21, author’s translation).

In any case, good access to the core elements of the documentation (e.g. metadata, report) is essential, because this helps potential re-users to decide whether the data are relevant for them.

The most important *recommendation* for documentation is that it should be carried out during the research process, i.e. integrated into the process as far as possible. Compared to creating documentation shortly before project completion, this not only reduces the total effort involved; it also often improves the quality because in this way, the research process does not need to be laboriously reconstructed from memory, so avoiding the risk of presenting the process in the light of the findings. In addition, we point to aids and recommendations on documentation that can be adapted to suit the requirements of a particular research project (e.g. Heuer et al., 2020; ZBW et al., 2014; VerbundFDB, 2019).

Documentation using metadata

A presentation of metadata is essential for both quantitative and qualitative research data. Metadata are “data about data”, giving information on both scientific content and formal and technical aspects. Unlike other forms of documentation, metadata primarily serve to inform external third parties and are usually only recorded when the data are archived. However, their role for the searchability and findability of research data – and thus for (potential) re-use – can hardly be overestimated. In the case of quantitative data, metadata alone may give exhaustive information about the data and their suitability for re-use – for example, if they are available in detail not just on dataset level but also give details of individual data (e.g. variables).

One important recommendation is to describe research data using widely recognised *→metadata standards*, if possible. This considerably improves compliance with the FAIR criteria of interoperability and findability. There are general standards for metadata such as Dublin Core that enables the description of documents and objects using 15 core items. Numerous discipline-related standards also exist that permit detailed descriptions. For example, the Data Documentation Initiative (DDI)'s metadata standard is widely used in the social, behavioural and economic sciences, as is the da|ra metadata scheme that supplements DataCite's metadata standard with specific social and economic science metadata.

Unified terminology should be used for the information contained in metadata as far as possible. This not only improves the searchability and findability of the data, but also enables users to compare and link different datasets. When filling in metadata fields, therefore, it is important to use a *→"controlled vocabulary"* if possible. This term refers to guidelines or definitions of terms and rules that are collected in word lists or structured thesauruses. For example, the European Language Social Science Thesaurus (ELSST; CESSDA and Service Providers, 2022) created by the Consortium of European Social Science Data Archives (CESSDA) is relevant for the social sciences. Controlled vocabularies also include *→standardised data*, such as the Integrated Authority File (GND; German National Library, 2022) often used by academic institutions in the German-speaking region, as well as national and international classifications, e.g., for branches of economics, professions, school types or educational levels.

Unfortunately, both metadata standards and controlled vocabularies are often designed to fulfil the demands of quantitative research or data. An adequate description for qualitative research is desirable, but so far only isolated solutions exist for this purpose. Generally, when facing questions about using metadata we recommend contacting FDM staff or research data infrastructures that can give information about relevant provision and requirements.

Documentation using (context) materials

While metadata can give a good (initial) impression of research data, context materials offer more detailed insights and information about the data. Context materials are documents or materials that – independently of a possible intention to archive the data – are created in the various phases of a research project: from the project proposal through interim report to final report; from the ethics commission vote through design of informed consent to the anonymisation concept; from initial considerations about case selection and methodology of survey instruments (e.g., questionnaires, interview guidelines, field and methodology reports) and preparation procedures (e.g., transcription rules, anonymisation measures) to evaluation aids (e.g., memos and case study descriptions, category systems, software commands and code books).

One aspect which all these materials and information have in common is that due to their origin in the research process, they may be important for understanding and re-using research data. Therefore, researchers should consider which materials can be usefully prepared in this way. As with research data, ethical and legal aspects should always be taken into account when considering this aspect: e.g., sensitive personal data may need to be anonymised. Regarding access to this material, RDCs can ensure that context materials are given the same protection as the research data themselves – particularly since the line between these categories is often flexible. In the interests of re-usability, researchers should ensure that *"open"* formats and standards are preferred to closed, proprietary versions. The core recommendation is thus to consider which materials could be important for understanding the research data and to make these available as far as possible.

Documentation using a (data or study) report

A third form of research documentation is central for transparency and re-usability of research data: a stand-alone text that describes and explains the data. A text on data development may take a variety of forms depending on the data and discipline: e.g., as a simple document, a working paper or an academic paper. RDCs in particular often require this type of report – referred to e.g. as *"study report"*, *"data and methodology report"* or *"context paper"* – because it makes the context information required by secondary users available in one package. A report of this type usually includes information about the research project (e.g. institutions and individuals; research question, preliminary work and concept), how data collection was prepared and implemented (selection of methods, participants, etc.) and all further steps in processing and analysis that resulted in the final dataset (e.g., transcription, anonymisation,

evaluation procedures). It may also include information about research findings, papers that emerged in the project context, additional context materials and the data's re-use potential.

While the context materials mentioned above emerge almost naturally during the research process, a stand-alone report demands additional work by the researchers that should be considered during project planning. Experience has shown, however, that the amount of work is reduced when the report is written alongside project work. Report writers can also often make use of preliminary work (DMP, other context materials). The report also offers the opportunity to present implicit and informal contexts that other context materials do not reveal.

Finally, a report on research documentation is an important *methodology paper* of the research project that gives the opportunity to present the research process in detail. RDCs not only often require such a report; they also often offer the opportunity to publish this report, provided with a persistent identifier.



Useful tools und links on **data protection**

- Interactive Virtual Assistant (iVA) on data protection regulations GDPR (DSGVO), BDSG, LDStG; BERD@BW, Projektlaufzeit 2019-2022).
- Open source PIA Software to prepare a data protection impact assessment (CNIL, 2021).
- Die Datenschutz-Folgenabschätzung nach Art. 35 DSGVO: Ein Handbuch für die Praxis (Martin et al., 2020). (German)
- Data Protection Guide: 2nd fully revised edition (RatSWD, 2020).
- Webpage on data protection in the social and economic sciences (RatSWD, n.d.–b).
- Webseite zu datenschutzrechtlichen Aspekten und Handreichungen (VerbundFDB, 2022a). (German)



Useful publications & links on **storing, backup, securing and organising research data**

- Open Science Training Handbook (Bezjak et al., 2018).
- CESSDA Data Management Expert Guide (CESSDA Training Team, 2020)
- During Data Collection – Overview (DataWiz Knowledge Base, n.d.–a)
- E-Learning Plattform on Open Science (FOSTER, n.d.).
- Beitrag zur Dateiorganisation in empirischen Forschungsprojekten (Recker & Brislinger, 2019). (German)
- Learning Hub on Research Data Management (UK Data Service, n.d.–a).
- Managing and Sharing Data. Best Practice for Researchers (van den Eynden et al., 2011).
- Webseite zum Forschungsdatenmanagement (VerbundFDB, n.d.). (German)



Useful publications on **contextualising and documenting research data** (all in German)

- Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: eine Handreichung für Forschende zur Erstellung eines Studienreports (Heuer et al., 2020)
- Leitfaden zur Kontextualisierung von qualitativen Befragungsdaten (VerbundFDB, 2021).
- Auffinden, zitieren, dokumentieren. Forschungsdaten in den Sozial- und Wirtschaftswissenschaften (ZBW et al., 2014)

3.3 Processing and analysis phase

It is important to document all data processing and analysis steps as fully and transparently as possible. This enables other researchers to reproduce the findings of analyses and increases the transparency of one's own research. Numerous studies have shown that it is difficult to reproduce published results even when the analysis files are available (e.g., Hardwicke et al., 2021; Stodden et al., 2018). Principles of data processing and analysis are discussed in the following. However, we must point out that the explanations mainly relate to quantitative studies. The core steps of data processing and analysis may differ for other study methodologies (e.g., qualitative research).

We can differentiate between raw data and processed data. Raw data files contain the first record of raw data (e.g., questionnaires, audio recordings of interviews) in digital format. The raw data files are transformed into analysis files during data processing. Typical steps in data processing are, e.g., testing and transforming variables, developing scale values, dealing with outliers and missing data, and coding open answers. A data processing workflow is recommended to carry out these steps, made up of three components as posited by Brislinger & Moschner (2019, p. 108):



Three components for implementing Brislinger and Moschner's data processing workflow: (2019, p. 108):

1. A hierarchical folder structure that enables structured storage of data and work steps. It is advisable to set up a readme file that describes the folder structure (see section 3.2.3 for how to set up a folder structure).
2. Raw data files, processed data files and analysis files. Different versions of these are extant that make the data collection and processing steps visible.
3. Programme code (syntax files in R, Stata, SPSS etc.), that carries out the steps for coding variables, data transformation and error search.

The most important rule of data processing is never to make alterations to the raw data files of original data, but always to save the transformed variables in new data files. Therefore, it is advisable to store the raw data files write-protected in a separate folder right from the start and to make no changes to this folder. In addition, every processing step should be carried out using programme code (syntax file) and not "by hand" (i.e., by using drop-down menus or point-and-click in the graphic user interface). This has the advantage that all transformations of raw data are documented. It also means that errors in data processing (e.g., wrong coding of a variable) that may only be detected at a later stage (e.g. data analysis or publication of findings) can still be easily corrected later. Because the raw data are still available in a completely unprocessed form, all processing steps can be reproduced using the adapted syntax files to produce a new dataset that includes the corrected values.

In order to document the data processing and analysis steps, data and syntax files should always be given new names when alterations are made. The file name may identify e.g. the contents and date it was created (see section 3.2.3 on naming conventions). This can result in a huge number of files in the course of a project, particularly if several people are working simultaneously on the data processing and a record is kept of who made which alterations to the data or syntax files over time. It is therefore often advisable to work with a version control system in which alterations to files are documented giving the time and person responsible. Version control software (e.g., git/GitHub, Bazaar) that can be used to manage the different versions of files is suitable for this purpose. Version control not only enables efficient documentation and reproduction of each work step; it also makes collaborative working easier because each alteration is recorded and can therefore be clearly attributed to one person.



Further tips for an efficient and reproducible workflow (Christensen et al., 2019; Long, 2009):

1. Syntax files for data processing and analysis should include detailed commentaries that explain individual work steps.
2. Data processing and data analysis should be separated and undertaken in separate syntax files.
3. Variable names should be as informative as possible (e.g., "male" instead of "gender" if 1 = male and 0 = female).
4. The version of statistical software used for analyses should always be documented (including any supplementary packages required).

Depending on the research question and design, it may be advisable to document in advance both the hypotheses and the statistical or other analyses planned to test them (preregistration of planned analyses; Christensen et al., 2019). This step prevents the hypotheses being altered in retrospect to fit the analysis results or the statistical analyses being varied (e.g., by using different specifications of the models) until the desired result is obtained. In addition, preregistration enables a clear distinction between confirmatory research (hypothesis testing) and exploratory research. Preregistration is less useful in designs for qualitative social research where it is explicitly intended that data collection will be successively adapted depending on the initial findings (e.g., theoretical sampling).

It is nowadays regarded as good scientific practice in quantitative social research that for an individual publication (e.g., an article in a journal), the data and programme code for reproducing the findings are made available. Many journals (e.g., the American Journal of Political Science) expect that a "replication package" will be submitted along with the manuscript that enables the findings to be verified (Christensen et al., 2019). A replication package usually includes (1) a readme file describing the package structure, (2) the data files for the statistical analyses including a description of the variables (codebook), (3) the syntax files for the statistical analyses and (4) information about how the manuscript contents (e.g., tables, illustrations) in the data files can be reproduced.

Individual data in the social sciences often include private or sensitive information about the data subjects (see section 3.2.1). Therefore, it will not always be possible to provide the original data files in a replication package for privacy reasons. In some cases, additional protection can be provided for data subjects by data access restrictions (see section 4). If this is also impossible, at least the programme codes for the statistical analyses should be made available.



Useful tools and links on research data processing and analysis

- BITSS Resource Library with resources for research transparency and reproducibility (BITSS, n.d.).
- Website with information on version control (DataWiz Knowledge Base, n.d.–b)
- Toolbox for fostering Open Science practices (OSC - LMU, n.d.)
- Website on integrating transparency and reproducibility in research training (Project TIER, n.d.)
- A handbook for reproducible, ethical and collaborative research (The Turing Way Community, 2021).
- Website on research data management (VerbundFDB, n.d.).



3.4 Publishing / archiving phase

Even small-scale research projects should give early consideration to the question of whether and if so, where and how they could archive the research data generated by their project and make it available for scientific re-use after project completion. These considerations should be *part of project planning before making an application*; this is the only way to ensure that the necessary activities and resources are included in the project planning (e.g., as part of a DMP) and that, if necessary, the costs can be included in a project proposal.

The widely used term →“data sharing” refers to opening, sharing, making available, releasing or publishing research data. All these terms refer to similar practices: research projects, as data depositors, turn to an infrastructure institution for research data – e.g., an RDC or an institutional or general repository – through which they can make their data (and supplementary material) available for re-use. It is important to note that for legal and ethical reasons, sensitive and personal data often cannot be made freely available but require certain protective measures. Data sharing in an RDC accredited by the RatSWD has the advantage that these providers guarantee compliance with the RatSWD’s quality criteria and orientation on the FAIR principles and can at the same time enable data protection – e.g., by guidelines for use, phased access paths or embargos.

3.4.1. Vorteile und Anforderungen beim Datenteilen

Research funding institutions, professional associations and scientific publishers or journals increasingly expect researchers to preserve their data and make them available, or to explain why data sharing is not possible or useful. The *advantages* of data sharing for the scientific community lie in the greater transparency and traceability of research results, as well as in better (re-)use of research data that have often been collected with a great deal of effort, resulting in an increased amount of available data. Section 2 above pointed to other advantages for data depositors and users.

On the other hand, it is possible that the research data cannot be made available, or only to a limited extent, for ethical, legal or professional reasons. The open nature of the research process in qualitative research and the heterogeneity and complexity of the data products make processing and documenting the data particularly challenging. However, even for these non-standardised research data – that are often particularly rich in information and therefore interesting for re-use – there is increasing RDM provision in the form of information resources and tools as well as advisory services and workshops that small-scale research projects in particular should make use of. In this way, researchers can gain an overview of the possibilities and limits of data sharing and the requirements it entails, enabling them to make a carefully considered decision based on this background information.

3.4.2 Finding suitable data infrastructure

Finding suitable data infrastructure for sharing research data is often closely linked to the question of whether and how the data are to be shared and should therefore be considered as early as possible – i.e. in the planning phase before a project proposal.⁸ There is now a vast number of infrastructures for archiving and sharing research data. The Registry of Research Data Repositories (re3data) website gives a comprehensive overview and enables users to filter the range of providers according to numerous criteria. To find a suitable institution for their research data in this way, researchers should consider the standards in their subject and the requirements of their own project. These include issues of the protective level the data require, and the access paths offered, searchability and findability, and whether the infrastructure is suitable for particular data types. Other issues that may play a role include considerations about the need for advice and support or the options provided for data curation (e.g. technological and other quality testing, legal testing and if necessary, supplementary processing).

As a rule of thumb, general or openly available repositories (e.g. Zenodo) are usually (only) suitable for data where no access restrictions are required (e.g. for ethical or legal reasons); otherwise, specialist providers should be selected. Sometimes a research data infrastructure that specialises in a particular

⁸ However, this decision cannot always be made so early in the project process because it is not always clear whether collecting primary data may be impeded by the intention to share the data. It may therefore be necessary to revise a decision already made or to postpone the decision about making data (partially) available to a later date.

type of data or discipline is explicitly recommended, while the institution's own store or general repositories are only listed as second or third choice (OpenAIRE, 2022). These recommendations are based on the idea that research data often require particular technological and organisational protective measures that openly accessible repositories do not or cannot provide. Specialist research data infrastructures usually also use metadata standards that enable better findability of data and can contribute to better data quality through advice and data curation. The latter is also an advantage of specialist research data infrastructures over institutions' own repositories – in case of the latter, researchers should also check whether they are even able to or intend to provide long-term archiving or availability.

In the social, behavioural and economic sciences the model of decentralised research data infrastructure in RDCs that specialise in particular topics, disciplines or data types has proved its worth. The RatSWD has currently accredited 42 RDCs (as of October 2023) that also guarantee compliance with particular quality criteria and orientation on the FAIR principles. However, it should be noted that not all the RDCs listed there are able or designed to take in external research data. On this issue, too, the websites of KonsortSWD and individual RDCs, or direct contact to an RDC, can provide further helpful information.

3.4.3 Data sharing, a collaborative task

Archiving and sharing research data is ideally a collaborative process between researchers and the RDC. This is why it is highly advisable for researchers to contact an RDC directly at an early stage. In this way, the RDC can inform the user about any requirements they may have regarding the data and documentation; at the same time, many RDCs also offer advice and support with planning the necessary tasks. Early contact with the RDC also enables any expenses for archiving, curating and sharing the data to be included in the project proposal. Depending on the extent of data processing required to prepare for archiving, it may be advisable for researchers to include the relevant tasks as a separate work package in a project proposal and to apply for the necessary resources. Archiving BASIS, a free service offered by GESIS,⁹ is a key service provided for the social sciences.

Researchers should preferably prepare and document the data while the project is running, i.e. this should be integrated in the research process. Experience has shown that plans to archive and share data should be considered in various processes, particularly when seeking informed consent, anonymising or pseudonymising data and for research documentation. If questions arise, project staff should always approach the RDM staff of their own institution or their RDC, to ensure that the data can be archived and shared with minimal friction.

Before transferring data to an RDC, it is common to sign a *data transfer agreement* that regulates the rights and obligations of both researchers and RDC, in which researchers can often stipulate conditions or requirements for sharing/re-use of their data (e.g. types of use, access paths, embargos). Depending on the agreement and on what services the RDC offers, the data are checked after transfer for certain – technological, legal or content-related – aspects and the (long-term) archiving is prepared. This process may reveal whether any clarification of content and possibly post-processing are required. The archiving process is usually completed when the metadata and if applicable, additional freely accessible materials, are published, enabling secondary users to gain information about the data. The datasets themselves should be given a persistent identifier (e.g., DOI) so that they are visible and citable (cf. ZBW et al., 2014, p. 17).

Depending on the level of protection the data require, RDCs often offer a range of access paths. They differentiate between generally accessible – but, if necessary, very strictly anonymised – public use files (PUF) or campus files (CF) for academic teaching and less easily accessible – but also (if appropriate) less strictly anonymised – scientific use files (SUF) for academic research. It is also possible to place research data in a specially secured “safe room” where they are only re-usable on-site at guest workstations; this is often the case with data where anonymisation is impossible or pointless (or possible only to a limited extent) either technologically or due to content-related issues.

Data depositors specify the conditions under which research data can be shared and re-used. Many RDCs offer data depositors the option of specifying these conditions in detail in order, among other things, to do justice to the diversity and particular sensitivity of (personal) data. These specifications are often set

⁹ <https://data.gesis.org/sharing/>

down in a data transfer agreement between the RDC and the data depositors, stipulating e.g., embargos, access paths and use options. Accordingly, secondary users of the data must sign a user agreement in which they agree to comply with these conditions for data use (cf. also the model agreement for data use developed by KonsortSWD: Schallböck et al., 2022).

→Licences that stipulate use conditions in standardised form are an alternative to the above agreements. Standardisation has the advantage that it is easier for scientists who re-use the data to bring together data from different sources, but they offer fewer configuration options for data depositors. Although some licensing models have been specially developed for research data (e.g., Open Data Commons, ODC), Creative Commons (CC) licences are most often used. The latter offer a modular system that provides four licence elements to build six licences that mainly differ in whether the further processed data may be shared and whether commercial use is permitted. If no data use restrictions are necessary, a recommended option is the CC BY open licence that simply provides for mentioning the name/attribution or the Public Domain Dedication (CC0) that does without re-use conditions altogether (for more details, cf. Brettschneider et al., 2021). Generally, it is important to ensure that not only use agreements but also licences are legally binding and thus stipulate what the data may or may not be used for.

Finally, it is important to be aware that the services offered by individual RDCs may differ considerably, for example in relation to consultation, data curation and the forms of access (and also in relation to the fees thus entailed). To sum up, making early contact with an RDC is and always will be essential for researchers.

3.5 Re-using data

Creating data pools for the re-use of data according to the FAIR principles can also be seen as a *public good*, whose function depends on the dissemination of trustee-like user behaviour and institutionalisation of the function of a trustee (Ostrom, 1990). In order for research data from a range of institutions such as universities, non-university institutions and commercial organisations to be transferred to data trustees, recognition through citation is necessary and trust is needed that these data will then be available to the scientific public, that they will be used, and that misuse is impossible. In addition, archiving research data also requires different organisational principles from the library catalogue method of listing by author and title. To ensure that archived data are findable, metadata on the dataset and its origin are required, as are indicators of the often diverse topics that are dealt with in datasets in social and economic research. These guidelines were already followed as early as 1960 when the Central Archive for Empirical Social Research (ZA) was founded, Europe's first archive for social and economic science (survey) data – now part of GESIS, the Leibniz Institute for the Social Sciences (Mayer & Dalhoff, 1998). Since then, the concept of a research-related institution that archives data as a trustee has become widespread on many levels.

For *data users*, the continually growing store of data for secondary analysis is of great interest, particularly in view of diminishing response rates for large-scale standardised surveys (Birke & Mayer-Ahuja, 2017; Medjedović, 2014; Mochmann, 2014). Research subjects have limited time and their numbers do not increase at the same rate as the growing demand from social and economic sciences. For this reason alone, secondary analyses should be considered for content-related and particularly for methodological research questions. Large-scale research instruments such as ALLBUS, NEPS and SOEP can be used for research questions over and above a narrow research purpose. In addition, long-term developments can be analysed ex post on the basis of existing data, if not in longitudinal section, at least in multiple cross-sections. However, secondary analyses require careful research into the content and methodology of potential datasets to assess how well the data fit the new research question and the re-use it entails.

The GESIS *data archive* now contains more than 6,500 studies that are currently available for secondary analysis. These datasets can be searched online¹⁰ and questionnaires, lists of variables, code books, etc., can be viewed. The data service makes the data available, depending on how the data depositors have classified their data (see ► Fig. 3). Sensitive data that require particular protection are only passed on under special contractual obligations (off-site access) or are in principle only available at a guest workstation in a “safe room” (on-site access). Many RDCs offer a range of access paths. Similarly, many RDCs provide support for data depositors in preparing their research data for re-use purposes. These RDCs help with



¹⁰ <https://www.gesis.org/en/services/finding-and-accessing-data>

anonymisation, documentation and classification and ensure long-term availability of the data in their archives.

Figure 3: User regulations of the GESIS data archive

| | |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Category 0 | Data and documents are available to all. |
| Category A | Data and documents are available for academic research and teaching. |
| Category B | Data and documents are available for academic research and teaching if the findings are not to be published. If publication or further processing of findings are planned, approval must be gained via the data archive. |
| Category C | Data and documents are only available for academic research and teaching with written permission by the data depositor. The data archive obtains this permission, giving information about the user and the purpose of the data processing. |

Source: GESIS data archive (2007)

In view of the growing diversity of specific research fields and methods, centralisation proved to be an inefficient organisational principle in the long term, particularly since technologies became available that enable users to search decentralised storage locations. The Open Science Framework (OSF) in particular recommends using these instruments, not only to make texts and findings available but in particular also for survey materials, laboratory records, etc., as well as datasets (cf. inter alia Max Planck Society (MPG), 2003). In principle, the combination of repositories, licences (see section 3.4.3) and search engines (see section 3.4.3) enables decentralised provision of and research in datasets on the internet (Federal Ministry of Education and Research (BMBF), 2021). In addition, repositories can be certified regarding long-term usability and citability (cf. DIN 31644:2012-04 and associated seals such as Nestor or CoreTrustSeal). Unlike publications or data from natural or technical sciences, whose availability ideally merely implies a relationship between the data's originators and users, in the case of the social, behavioural and economic science research data the data subjects are usually also involved and therefore privacy issues must be considered. This involves rights and obligations – from information to revocation – that require consultation and support, tasks that work against the powerful drive towards rationalisation. This means that institutional publication servers are only useful to a limited extent for the dissemination of social and economic science research data.

These disciplines have reacted to the increasing thematic and methodological diversity in recent decades with *differentiation* of RDCs. The current RDC landscape comprises specialised institutions for certain datasets, disciplines, research objects or research methodologies. However, the decentralised archiving structure is (increasingly) intransparent for the data user. A search in the above-mentioned GESIS or KonsortSWD portals includes all linked RDCs and delivers a citable source, link or access path to the dataset, plus its description, based on the persistent indicators it bears (e.g., a DOI via da|ra in collaboration with DataCite).

Within Europe, CESSDA European Research Infrastructure Consortium (ERIC) provides a metaportal to data archives of European member countries. Research results, however, fall short of the standard of findings in national archives or portals. Yet, links to members' data archives, for example CESSDA-France/Réseau Quetelet (PROGEDO), UK Data Service and the Swiss Centre of Expertise in the Social Sciences (FORS), are stored in CESSDA as are links to non-members such as UniData in Italy, Irish Social Science Data Archive (ISSDA) and the Spanish database Centro de Investigaciones Sociológicas (CIS).

In the USA, the inter-university Consortium for Political and Social Research (ICPSR) is recommended where not only inter-university databases of social and political science studies are located, but also information about other data archives and their possible uses. A global overview – of both data archives and other statistics and data sources – is found at the Social Science Information System (SOCIOSITE) maintained by the University of Amsterdam.

The situation is different regarding data production. Since RDCs are specialised, they can only give tailored support for relevant datasets. Whereas a generalised survey project is well-supported at e.g. GESIS, a generalist institution, a mixed-methods project in organisational research will be appropriately supported by the Research Data Center for Business and Organizational Data (FDZ-BO). If qualitative social research methods are mainly applied, the Qualiservice RDC offers targeted support.

The debate about Open Access in science demands, and technological developments enable, that additional storage options are continuously being set up that also work alongside the research process. In recent years, increasing numbers of Research Data Services have been set up at larger universities that support RDM up to the point of data publication and re-use. Repositories at higher education institutions, where they exist, are therefore worth considering as an alternative to non-university RDCs. This makes not only the upstream RDM tasks easier. For smaller HEIs as well as for established communities, the attached RDCs with their disciplinary specialisation in research objects and methodologies are still the key contact for archiving.



Useful links on re-using research data

National:

- DSS (Data Services for the Social Sciences) (GESIS, n.d.).

International:

- DataCite (International Data Citation Initiative) (DataCite, n.d.).
- ICPSR (Inter-university Consortium for Political and Social Research) (ICPSR, n.d.-b).
- e3data.org (Registry of Research Data Repositories) (re3data.org, n.d.).
- UK Data Service (UK Data Service, n.d.-b).

4 Glossary

| | |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| → Anonymisation | Altering personal data so that either re-identification is impossible (<i>complete</i> anonymisation, extremely difficult and rare) or is only possible with an unreasonable amount of effort (effective anonymisation, the usual method for survey data). |
| → camelCase | The words are written without spaces or punctuation. The first word is written without a capital letter, the first letter of all the following words are capitalised. |
| → Controlled vocabulary | Vocabulary that can be clearly assigned. |
| → Curating data /studies | Organising, checking and processing or overseeing data /studies. |
| → Data trustee | An independent intermediary between data depositors and data users. Data trustees should protect the interests of all actors and in particular, guarantee data protection and data usability. |
| → Data sharing | Here the term refers to sharing research data with third parties who are not/were not involved with the primary research. |
| → Ethics approval | The evaluation of a research proposal by an ethics commission results in written ethics approval. |
| → Hosting data /studies | Preservation of research data or entire studies, making them available if appropriate; the core task of research data centres. |
| → Informed Consent | Informed Consent refers to the declaration of consent between researcher and data subject in an empirical study which precisely describes what participants agree to, how the data gained from them will be processed and what will happen to the data. In this document in the form of a contract, researchers guarantee compliance with relevant data protection or research ethics requirements when dealing with the knowledge gained in the research process and the related data (e.g., strict anonymisation, maximum retention period, conditional passing on to third parties if appropriate). |
| → Kebab-case | The words are separated by hyphens. |
| → Licence | Licences are standard contracts that permit certain actions to all re-users of research data. Data producers can use licences to lay down conditions of re-use and to present them in a comprehensible way. |
| → Living document | Documents that are continually added to, thus keeping them up to date. |
| → Metadata standards | Metadata are information about data. Metadata define which structured information about data is required. |
| → PascalCase | All word are capitalised and written without spaces or punctuation. |
| → Pixeling/blurring | Partial masking of digital photos or video frames, usually to prevent identification of people or locations, for the purpose of anonymisation. |

| | |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| → Pseudonymisation | Pseudonymising in accordance with Art. 4 (5) GDPR means “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”. |
| → Record of processing activities | A record of processing activities is always required according to Art. 30 (1) GDPR when personal data are processed. In particular, it contains information about which categories of personal data were processed by which persons/institutions for which purpose, and which data protection measures were taken. In Germany, academic institutions or the data protection officer of the Länder usually hold templates that can be used to document data protection measures. |
| → Repository | Managed location for preserving documents; here, managed storage location for research data. |
| → Research data centre (RDC) | Institutions that archive and process research data from internal and external sources, making the data available for re-use in some cases. RDCs for the social and economic sciences are accredited in Germany by the RatSWD for quality control. |
| → Research, qualitative | This refers to all research approaches that take an interpretative and/or reconstructive approach to qualitative data, regardless of which specific research approach is selected. |
| → Secondary data | Data gained in other empirical studies, processed and made available for re-use. Secondary data also differ from primary data in that they were not collected in relation to the concrete research question of a current project. |
| → Snake_case | The words are separated by an underscore. |
| → Standardised data | Standardised data are datasets that are clearly attributable to entities. The attribution can take place regardless of the writing format. |
| → Theoretical sampling | Theoretical sampling is a strategy of the successive development of case selection in qualitative empirical studies, suggested in relation to Grounded Theory. The sampling stages and criteria are designed according to the requirements of an object-related theory developed step-by-step in the course of research. |
| → Transcripts | Verbatim transcription of audio or video documents such as interview recordings, videos of group discussions or of interactions in public places. |
| → Vignette data | Data gained in the course of vignette analysis (factorial surveys). In vignette analysis, respondents are usually presented with several short fictional case studies in which individual text blocks (dimensions) are experimentally varied and the participants’ response to these is noted. |

5 Bibliography

- Argos. (n.d.). *Plan and follow your data*. <https://argos.openaire.eu/splash/> (accessed on 03 February 2023)..
- BDSG (Federal Data Protection Act) of 30 June 2017 (BGBl. I p. 2097), last amended by Article 10 of the Act of 23 June 2021 (BGBl. I p. 1858; 2022 I 1045). (30 June 2017). http://www.gesetze-im-internet.de/bdsg_2018/
- BERD@BW (Business, Economic and Related Data Baden-Württemberg). (Project duration 2019-2022). *Interactive Virtual Assistant: Anwendungsbereich DSGVO*. <https://www.berd-nfdi.de/iva1/> (accessed on 03 February 2023).
- Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H. & Heller, L. (2018). *Open science training handbook*. <https://doi.org/10.5281/zenodo.1212496>
- Birke, P. & Mayer-Ahuja, N. (2017). Sekundäranalyse qualitativer Organisationsdaten. In S. Liebig, W. Matiaske & S. Rosenbohm (Hrsg.), *Springer Reference Wirtschaft. Handbuch Empirische Organisationsforschung* (pp. 105–126). Springer Gabler. https://doi.org/10.1007/978-3-658-08493-6_7
- BITSS (Berkeley Initiative for Transparency in the Social Sciences). (n.d.). *Resource library*. <https://www.bitss.org/resource-library/> (accessed on 03 February 2023).
- BMBF (Federal Ministry of Education and Research – Digitale Zukunft). (2021). *Lizenzvertrag und Offene Lizenzen*. <https://www.bildung-forschung.digital/digitalezukunft/de/wissen/urheberrecht/urheberrecht-in-der-wissenschaft/lizenzvertrag-und-offene-lizenzen/lizenzvertrag-und-offene-lizenzen.html> (accessed on 03 February 2023).
- Brettschneider, P., Axtmann, A., Böker, E. & Suchodoletz, D. von (2021). Offene Lizenzen für Forschungsdaten: Rechtliche Bewertung und Praxistauglichkeit verbreiteter Lizenzmodelle. *O-Bib. Das offene Bibliotheksjournal*, 8(3), 1–22. <https://doi.org/10.5282/O-BIB/5749>
- Brislinger, E. & Moschner, M. (2019). Datenaufbereitung und Dokumentation. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (pp. 97–114). Verlag Barbara Budrich. <https://doi.org/10.3224/84742233.07>
- CESSDA (Consortium of European Social Science Data Archives). (2019). *Adapt your data management plan: A list of data management questions based on the expert tour guide on data management*. https://static-archive.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.3.pdf
- CESSDA and Service Providers. (2022). *The European Language Social Science Thesaurus (ELSST)* [Data set]. <https://doi.org/10.5281/Zenodo.7097110>
- CESSDA Training Team. (2020). *CESSDA data management expert guide*. Bergen. CESSDA ERIC. <https://doi.org/10.5281/zenodo.3820473>
- Christensen, G., Freese, J. & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.
- CNIL (National Commission on Informatics and Liberty). (2021). *PIA* [Computer software]. <https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment> (accessed on 03 February 2023).
- DARIAH-EU ELDAH (Digital Research Infrastructure for the Arts and Humanities, Ethics and Legality in Digital Arts and Humanities). (2021). *Consent Form Wizard (CFW)*. <https://consent.dariah.eu/node/2> (accessed on 03 February 2023).
- DataCite (International Data Citation Initiative). (n.d.). *Welcome to DataCite: Connecting research, advancing knowledge* [Homepage]. <https://datacite.org/> (accessed on 03 February 2023).
- DataWiz Knowledge Base. (n.d.-a). *During data collection: Overview*. <https://datawizkb.leibniz-psychology.org/index.php/during-data-collection/> (accessed on 03 February 2023).
- DataWiz Knowledge Base. (n.d.-b). *Versioning*. <https://datawizkb.leibniz-psychology.org/index.php/during-data-collection/what-should-i-know-about-versioning/> (accessed on 03 February 2023).
- Deutsche Nationalbibliothek. (2022). *Gemeinsame Normdatei (GND)*. <https://www.dnb.de/gnd> (accessed on 03 February 2023).

- DFG - Fachkollegium 111 „Sozialwissenschaften“. (2020). *Stellungnahme des Fachkollegiums 111 Sozialwissenschaften zum Forschungsdatenmanagement in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft*. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/fachkollegium111_forschungsdatenmanagement.pdf
- DFG (Deutsche Forschungsgemeinschaft). (2021). *Umgang mit Forschungsdaten: Checkliste für Antragstellende zur Planung und zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben*. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf
- DFG (German Research Foundation). (2022a). *FAQ: Informationen aus den Geistes- und Sozialwissenschaften*. https://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/ (accessed on 03 February 2023).
- DFG (German Research Foundation). (2022b). *Forschungsdaten in den Förderprogrammen*. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/forschungsfoerderung/index.html (accessed on 03 February 2023).
- DGPs (German Psychological Society). (2018). *Ethisches Handeln in der psychologischen Forschung: Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* (1st edition). Hogrefe.
- DIN 31644:2012-04. *Information und Dokumentation: Kriterien für vertrauenswürdige digitale Langzeitarchiv*. Berlin. Beuth Verlag. <https://doi.org/10.31030/1854419>
- DMPonline. (n.d.). *Plan to make data work for you*. <https://dmponline.dcc.ac.uk/> (accessed on 03 February 2023).
- DMPTool. (n.d.). *Build your data management plan*. University of California. <https://dmptool.org/> (accessed on 03 February 2023).
- DSGVO/GDPR (European General Data Protection Regulation). (04 May 2016). Official gazette of the European Union (OJ L 119), 1–88. http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=OJ:JOL_2016_119_R_0001
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. GESIS - Leibniz-Institut für Sozialwissenschaften. https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Anonymisierung_quantitativer_Daten-0150512.pdf
- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten*. Version 1.2 (forschungsdaten bildung informiert Nr. 3). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:21970>
- Eisentraut, M. & Hutflesz, T. (2021). *Organisation und Speicherung von Forschungsdaten: Forschungsdatenmanagement in der empirischen Bildungsforschung* [Online seminar]. German Network of Educational Research Data (VerbundFDB).
- European Commission. (n.d.). *Data management*. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (accessed on 03 February 2023).
- FDZ Qualiservice (Forschungsdatenzentrum Qualiservice). (n.d.). Helpdesk: Downloads. <https://www.qualiservice.org/de/helpdesk.html#downloads> (accessed on 03 February 2023).
- forschungsdaten.info. (2022a). *Der Datenlebenszyklus: Stationen des Forschungsdatenmanagements*. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/> (accessed on 03 February 2023).
- forschungsdaten.info. (2022b). *Der Datenmanagementplan: Eine Wegbeschreibung für Daten*. <https://forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/> (accessed on 03 February 2023).
- FOSTER. (n.d.). *Resources: Open science*. <https://www.fosteropenscience.eu/resources> (accessed on 03 February 2023).
- Friedrichs, J. (2014). Forschungsethik. In N. Baur & J. Blasius (eds.), *Handbuch Methoden der empirischen Sozialforschung* (1st edition, pp. 81–91). Springer VS. https://doi.org/10.1007/978-3-531-18939-0_3
- GESIS (GESIS – Leibniz Institute for the Social Sciences). (n.d.). *Data Services for the Social Sciences (DSS)*. <https://www.gesis.org/institut/abteilungen/data-services-for-the-social-sciences> (accessed on 03 February 2023).

- GESIS Datenarchiv. (2007). *Benutzungsordnung: GESIS Abt. Datenarchiv und Datenanalyse*. https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Benutzungsordnung.pdf
- Gollwitzer, M., Adele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A., Schönbrodt, F. & Steinberg, U. (2020). *Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen*. <https://doi.org/10.31234/osf.io/hcxtm>
- Hammersley, M. (2009). Against the ethicists: On the evils of ethical regulation. *International Journal of Social Research Methodology*, 12(3), 211–225. <https://doi.org/10.1080/13645570802170288>
- Hardwicke, T., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M., Peloquin, B., deMayo, B., Long, B., Yoon, E. & Frank, M. (2021). Analytic reproducibility in articles receiving open data badges at the Journal Psychological Science: An observational study. *Royal Society Open Science*, 8(1). <https://doi.org/10.1098/rsos.201494>
- Heuer, J.-O., Kretzer, S., Mozygemba, K., Huber, E. & Hollstein, B. (2020). *Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: Eine Handreichung für Forschende zur Erstellung eines Studienreports* (Qualiservice Working Papers 1-2020). Bremen. <https://doi.org/10.26092/ELIB/166>
- Hopf, C. (2016). Forschungsethik und qualitative Forschung. In C. Hopf (ed.), *Schriften zu Methodologie und Methoden qualitativer Sozialforschung* (pp. 195–205). Springer VS. https://doi.org/10.1007/978-3-658-11482-4_9
- ICPSR (Inter-university Consortium for Political and Social Research). (n.d.-a). *Elements of a data management plan*. <https://www.icpsr.umich.edu/web/pages/datamanagement/dmp/elements.html> (accessed on 03 February 2023).
- ICPSR (Inter-university Consortium for Political and Social Research). (n.d.-b). *This is ICPSR* [Homepage]. <https://www.icpsr.umich.edu/web/pages/> (accessed on 03 February 2023).
- Long, J. (2009). *The workflow of data analysis using Stata* (1st edition). Stata Press.
- Martin, N., Schiering, I., Mester, B., Hallinan, D. & Jensen, M. (2020). *Die Datenschutz-Folgenabschätzung nach Art. 35 DSGVO: Ein Handbuch für die Praxis*. Fraunhofer Verlag. <https://doi.org/10.24406/publica-fhg-300193>
- Mayer, E. & Dalhoff D. (1998). Ein Gespräch mit Erwin K. Scheuch. *ZA-Information*(42), 13–26. https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-42.pdf
- Medjedović, I. (2014). *Qualitative Sekundäranalyse: Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung* (1st edition). Springer VS. <https://doi.org/10.1007/978-3-658-05488-5>
- Meyermann, A. & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung: Eine Handreichung*. Version 2.1 (forschungsdaten bildung informiert Nr. 6). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:21990>
- Mochmann, E. (2014). Quantitative Daten für die Sekundäranalyse. In N. Baur & J. Blasius (eds.), *Handbuch Methoden der empirischen Sozialforschung* (1st edition, pp. 233–244). Springer VS. https://doi.org/10.1007/978-3-531-18939-0_14
- MPG (Max-Planck-Gesellschaft). (2003). *Berlin declaration on open access to knowledge in the sciences and humanities*. <https://openaccess.mpg.de/Berlin-Declaration> (accessed on 03 February 2023).
- Netscher, S. (2020). *It's not such 'A Fair Way Off' to process open data: Facing requirements on open access and the FAIR data principles* (GESIS Blog: Growing knowledge in the social sciences). <https://doi.org/10.34879/GESISBLOG.2020.3>
- Nicolai, T., Mozygemba, K. & Hollstein, B. (2021). QualiAnon: Qualiservice tool for anonymizing text data [Computer software]. FDZ Qualiservice. Universität Bremen [for information on registration/access see: <https://www.qualiservice.org/de/helpdesk/webinar/tools.html>] (accessed on 03 February 2023).
- Open Science Future. (n.d.). *Was Forschungsförderer in Sachen Forschungsdatenmanagement fordern: Und was bei Drittmittelanträgen zu beachten ist*. Leibniz Information Centre for Economics (ZBW). <https://open-science-future.zbw.eu/was-forschungsfoerderer-in-forschungsdatenmanagement-fordern/> (accessed on 03 February 2023).
- OpenAIRE. (2022). *Guides for researchers: How to find a trustworthy repository for your data*. <https://www.openaire.eu/find-trustworthy-data-repository> (accessed on 03 February 2023).
- OSC - LMU (Open Science Center - Ludwig-Maximilians-Universität München). (n.d.). *Toolbox*. <https://www.osc.uni-muenchen.de/toolbox/index.html> (accessed on 03 February 2023).
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

- Pergl, R., Hooft, R., Suchánek, M., Knaisl, V. & Slifka, J. (2019). "Data Stewardship Wizard": A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal*, 18(1), Article 59, 1–8. <https://doi.org/10.5334/dsj-2019-059>
- Project TIER. (n.d.). *Teaching Integrity in Empirical Research (TIER)* [Homepage]. <https://www.projecttier.org> (accessed on 03 February 2023).
- RatSWD (German Data Forum). (n.d.-a). Best Practices for Research Ethics: Best practice collection on research ethics in the research process. <https://www.konsortswd.de/en/topics/best-practices-research-ethics/> (accessed on 03 February 2023).
- RatSWD (German Data Forum). (n.d.-b). Datenschutz: Information on data protection in research. <https://www.konsortswd.de/en/topics/best-practices-research-ethics/data-protection/> (accessed on 03 February 2023).
- RatSWD (German Data Forum). (2016). *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte* (RatSWD Output Series, 5. Appointment period No. 3). Berlin. <https://doi.org/10.17620/02671.7>
- RatSWD (German Data Forum). (2017). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften* (RatSWD Output Series, 5. Appointment period No. 9). Berlin. <https://doi.org/10.17620/02671.1>
- RatSWD (German Data Forum). (2020). *Data Protection Guide. 2nd fully revised edition*. Revised by Prof. Dr. Matthias Bäcker and Dr. Sebastian Golla with the RatSWD (RatSWD Output Series, 6. Appointment period No. 8). Berlin. <https://doi.org/10.17620/02671.57>
- RDMO Arbeitsgemeinschaft. (n.d.). *Research Data Management Organiser (RDMO)* [Computer software]. <https://rdmorganiser.github.io/>
- re3data.org (Registry of Research Data Repositories). (n.d.). *re3data.org: Registry of Research Data Repositories* [Homepage]. <https://www.re3data.org/> (accessed on 03 February 2023).
- Recker, J. & Brislinger, E. (2019). Dateiorganisation in empirischen Forschungsprojekten. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (pp. 81–95). Verlag Barbara Budrich. <https://doi.org/10.3224/84742233.06>
- Schallaböck, J., Hoffstätter, U., Buck, D. & Linne, M. (2022). *Mustervertrag Datennutzung KonsortSWD (1.0.0.)*. <https://doi.org/10.5281/zenodo.5828114>
- Science Europe. (2021). *Practical guide to the international alignment of research data management: Extended edition*. <https://doi.org/10.5281/Zenodo.4915862>
- Steinhardt, I., Fischer, C., Heimstädt, M., Hirsbrunner, S., Ikiz-Akinci, D., Kressin, L., Kretzer, S., Möllenkamp, A., Porzelt, M., Rahal, R.-M., Schimmler, S., Wilke, R. & Wünsche, H. (2020). *Das Öffnen und Teilen von Daten qualitativer Forschung: Eine Handreichung* (Weizenbaum Series Nr. 6). Berlin. Weizenbaum Institute for the Networked Society. <https://doi.org/10.34669/WI.WS/6>
- Stodden, V., Seiler, J. & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>
- Strübing, J. (2018). *Qualitative Sozialforschung: Eine komprimierte Einführung* (2nd edition). De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110529920>
- The Turing Way Community. (2021). *The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.1)*. <https://doi.org/10.5281/zenodo.5671094>
- UK Data Service. (n.d.-a). *Learning hub: Research data management*. <https://ukdataservice.ac.uk/learning-hub/research-data-management/> (accessed on 03 February 2023).
- UK Data Service. (n.d.-b). *Welcome to the UK Data Service* [Homepage]. <https://ukdataservice.ac.uk/> (accessed on 03 February 2023).
- van den Eynden, V., Corti, L., Woollard, M., Bishop, L. & Horton, L. (2011). *Managing and sharing data: Best practice for researchers*. Colchester. UK Data Archive, University of Essex. <https://dam.ukdataservice.ac.uk/media/622417/managingsharing.pdf>
- VerbundFDB (German Network of Educational Research Data). (n.d.). *Daten managen*. Leibniz Institute for Research and Information in Education (DIPF). <https://www.forschungsdaten-bildung.de/daten-managen> (accessed on 03 February 2023).

- VerbundFDB (German Network of Educational Research Data). (2015). *Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung*. Version 1.1 (forschungsdaten bildung informiert Nr. 2). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:22298>
- VerbundFDB (German Network of Educational Research Data). (2018a). *Formulierungsbeispiele für "informierte Einwilligungen"*. Version 2.1 (forschungsdaten bildung informiert No. 4). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:22301>
- VerbundFDB (German Network of Educational Research Data). (2018b). *Formulierungsbeispiele für "informierte Einwilligungen" in leichter Sprache*. Version 1.1 (forschungsdaten bildung informiert No. 5). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:22302>
- VerbundFDB (German Network of Educational Research Data). (2019). *Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen*. Version 2.0 (forschungsdaten bildung informiert No. 1). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:22297>
- VerbundFDB (German Network of Educational Research Data). (2021). *Leitfaden zur Kontextualisierung von qualitativen Befragungsdaten*. Version 1.2 (forschungsdaten bildung informiert No. 9). Leibniz Institute for Research and Information in Education (DIPF). <https://doi.org/10.25656/01:22386>
- VerbundFDB (German Network of Educational Research Data). (2022a). *Datenschutzrechtliche Aspekte*. Leibniz Institute for Research and Information in Education (DIPF). <https://www.forschungsdaten-bildung.de/datenschutzrechtliche-aspekte> (accessed on 03 February 2023).
- VerbundFDB (German Network of Educational Research Data). (2022b). *Stamp nutzen: Standardisierter Datenmanagementplan für die Bildungsforschung*. Leibniz Institute for Research and Information in Education (DIPF). <https://www.forschungsdaten-bildung.de/stamp-nutzen> (accessed on 03 February 2023).
- Wilkinson, M. D., Dumontier, M. & Aalbersberg, I. e. a. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018). <https://doi.org/10.1038/sdata.2016.18>
- ZBW (Leibniz Information Centre for Economics), GESIS (GESIS – Leibniz Institute for the Social Sciences) & RatSWD (German Data Forum). (2014). *Auffinden, zitieren, dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften*. Hamburg. <https://doi.org/10.4232/10.fisuzida2015.2>

Contributors

Members of the working group FDM (research data management) Services for Small Research Projects: Guide to Cultural Change

Prof. Dr. Hansjörg Dilger

Free University of Berlin

Prof. Dr. Michael Eid

Free University of Berlin, RatSWD

Dr. Anne Gresser

Julius Maximilian University of Würzburg

Prof. Dr. Christiane Gross, *co-chair*

Julius Maximilian University of Würzburg, RatSWD

Director of the working group on RDM for small projects

Dr. Jan-Ocko Heuer

University of Bremen

Prof. Dr. Oliver Lüdtke

Leibniz Institute for Science and Mathematics Education at Kiel University (IPN),

Christian-Albrecht University of Kiel, RatSWD

Prof. Dr. Wenzel Matiaske

Helmut Schmidt University/University of the Bundeswehr Hamburg

Prof. Dr. Laura Seelkopf

Ludwig-Maximilians-Universität München, RatSWD

Prof. Dr. Jörg Strübing, *co-chair*

Eberhard Karls Universität Tübingen, RatSWD

Director of the working group on RDM for small projects

RatSWD Office

Marie Eilers

Imprint

Editor:

German Data Forum (RatSWD)
Office
Am Friedrichshain 22
10407 Berlin
Germany
office@ratswd.de
<https://www.ratswd.de>

Translation:

Bridget Schäfer, B.Schaefer-Uebersetzerin@kabelmail.net

Layout:

Claudia Kreutz

Berlin, November 2023

RatSWD Output:

The RatSWD Output Series documents the German Data Forum's (RatSWD) activities during its 7th appointment period (2020-2023). It serves to publish its statements and recommendations and to make them available to a broad audience.

The RatSWD Office is funded by the German Research Foundation (DFG) as part of KonsortSWD in the NFDI framework – project number: 442494171.



This publication is licensed under the Creative Commons Licence (CC BY 4.0):

<https://creativecommons.org/licenses/by/4.0/>

Excluded from the aforementioned license are portions, illustrations, and other third-party material, unless otherwise indicated, as well as the images on pages 1 and 44. (Deemerwha Studio – stock.adobe.com).

DOI: 10.17620/02671.84

Suggested citation:

RatSWD [German Data Forum] (2023). *Research Data Management for Small Research Projects. A practical guide* (RatSWD Output Series, 7. Appointment period No. 3). <https://doi.org/10.17620/02671.84>

■ The **German Data Forum (RatSWD)** has been advising the Federal Government and the governments of the German states on questions of research data infrastructure for the empirical social, behavioural and economic sciences since 2004. In the RatSWD, ten representatives of the social, behavioural and economic disciplines, legitimised by election, work together with ten representatives of data production.

The RatSWD is part of the Consortium for the Social, Behavioural, Educational and Economic Sciences (KonsortSWD) in the National Research Data Infrastructure (NFDI). It sees itself as an institutionalised forum for dialogue between science and data producers and develops recommendations and statements. In doing so, it is committed to an infrastructure that provides science with broad, flexible and secure access to data. These data are provided by governmental, science-based and private-sector actors. Currently, the RatSWD has accredited 42 research data centres and promotes their cooperation.

