

Reporting Research Data Use for DFG Project Proposals in the Business and Economic Sciences

These recommendations build upon the general guidelines for reporting research data usage in DFG project proposals and extend them for applications in the realm of business and economic sciences. They are, therefore, complementary to the already existing guidelines and are only relevant when conducting research in the respective research fields. This version was last updated on October 23, 2023.

Introduction

Given the modern methodological advances in business and economic research, the complexity of scientific studies has reached new heights. Many of these studies rely on specific datasets developed by the authors for their research projects. This interplay of high complexity and custom datasets, however, often makes it difficult for others to fully comprehend the underlying studies. As a general comprehension would accelerate knowledge processes and research efficiency, ensuring the reproducibility of research has, therefore, become good research practice.

That is why in proposals submitted to the German Research Foundation (DFG), we also recommend applicants to make their research reproducible. Applicants are strongly encouraged to commit to the so-called “FAIR principles”, so that the submitted data for the funded research projects ultimately become Findable, Accessible, Interoperable, and Reusable for all humans and machines. This allows to stimulate new research projects in the future as many datasets can be used to answer multiple research questions. To easily and successfully follow these principles, we propose to use the [BERD@NFDI platform](#). BERD is a research platform for **B**usiness, **E**conomics, and **R**elated **D**ata which provides various functionalities to make research reproducible and contains many [field-specific research data management trainings](#).

The following contents provide field-specific recommendations for those sections that should be included in the proposal according to the [general DFG guidelines](#), and illustrate how the BERD platform can support applicants to follow these recommendations. The DFG guidelines recommend inputs on the (A) data description, (B) documentation and data quality, the (C) storage and technical backup during the project, (D) legal obligations and framework, (E) data exchange and permanent accessibility of data, and (F) the responsibility and resources when describing the handling of research data. Please refer to the general DFG guidelines first before reading this document.

A) Data Description

Applicants are advised to report all primary and secondary data sources that they (intend to) use in the scope of their research project, including the variables they would like to measure and how they will measure them. While primary datasets are all those explicitly newly created for the research project, secondary datasets have already existed in the past and are reused for the respective project. Potential data sources could, for instance, be surveys, qualitative interviews, experiments, simulated data, and data from administrative, corporate, or social media sources. The previously recommended research platform BERD allows users to store files and information about all these structured and, in particular, also unstructured data sources.

Besides datasets and program code, it depends on the data source which information is recommended to be made accessible. In the case of qualitative analysis, we suggest providing information on the respondent selection process, theoretical assumptions, used instruments, and protocols. In the case of quantitative analysis, information about the data collection process (e.g., concerning who was involved, for how long it was performed, and which tools have been used) could be provided. Further, in the case of experiments, original instructions, information regarding the eligibility and selection of subjects, and any program or script used to conduct the experiment is recommended. If multiple datasets are later used and merged for the final analysis, all raw datasets are advised to be submitted. The merging steps could be thoroughly explained in the coding files for a better understanding. Once available, the raw data should be stored securely (see section C for recommendations) in a pseudoanonymized (or better: anonymized) format, and shared, if it is not legally protected from doing so (see section E). We recommend including programming code for raw data pre-processing and analysis steps only as a single file in the submission. By providing the raw data with the code, the reproducibility of the author's results is ensured.

B) Documentation and Data Quality

We suggest applicants to sufficiently document each dataset used for the research project with metadata. Metadata contains several pieces of information about the dataset at hand, e.g., the title of the dataset, the creators' names, its size, a short description, or the date of publication. Applicants are strongly encouraged to report the metadata of each of their datasets by an established, standard metadata scheme, such as Dublin Core or DataCite, thereby fostering the interoperability of research data. The availability of metadata also ensures that the employed datasets are later findable and reusable by others, as part of the FAIR dimensions. Thorough data documentations also contain information about the content (e.g., keywords or JEL codes), unit of observation (e.g., countries, companies, or individuals),

data basis (e.g., complete survey or random sample), methodology (e.g., questionnaire or sales statistics), and scope (e.g., time period, region, and number of observations or variables). The description, measuring unit, and relations among all variables of each dataset also need to be clear (e.g., by providing a README file or a codebook for tabular data). In addition to raw datasets, newly created datasets could also be described, which would include information about the merging of datasets, narrowing down of observations, recoding or generation of variables, and further changes to raw data.

Next to documenting their data, we encourage applicants to also provide their documented code. The program code (i.e., analysis documentation) is the core of any empirical analysis and crucial for the traceability and replicability of research projects. It needs to be functional, understandable, and accessible. Code documentation, for instance, can be accomplished by adding comments directly in the code. Here, it is essential to explain each of the performed coding steps from analysis dataset creation, recoding, and generation, to the evaluation of the results. Having a clear and readable code is particularly essential in the case of unstructured data to allow reproducibility of the pre-processing and results.

Finally, applicants are also encouraged to demonstrate how they intend to ensure high data quality for their datasets. Here, they could, for instance, define accurate data quality measures for each of their datasets, explain how they will assure the quality of the data, and, once the data is available, readily assess how far data quality impinges on the results of the research project (e.g., if they know about any sources of errors or redundancies). The latter is important as poor data quality can introduce errors, bias, and uncertainty into the analysis, potentially leading to incorrect conclusions and affecting the overall reliability and validity of the applicant's research.

C) Storage and Technical Backup During the Project

In their proposal, the DFG requires applicants to outline how they intend to store data. This includes how the researchers ensure that only individuals who need to and may access the data for research purposes can do so, in particular for sensitive, personal data that is often collected in this research field. While there are multiple options to store datasets (e.g., on local or cloud drives), we recommend storing the datasets on BERD. BERD offers vast private storage capacities for research users. It allows to store datasets and codes jointly with metadata on the platform and can also be shared with other collaborators. A link can be created and shared in the DFG project proposal and subsequent reports to prove faithful and secure storage of one's research data. Storing data in such a way ensures its accessibility during the project and paves the way for its accessibility in the long-run, as required in the FAIR dimensions.

D) Legal Obligations and Framework

In business and economic sciences, research is often based on the behavior of individuals or firms. Researchers in these fields frequently handle sensitive and confidential data, such as personal information, individual survey responses, or internal corporate data. Safeguarding privacy and complying with data protection regulations such as Germany's General Data Protection Regulation (DSGVO) is, therefore, critical. Applicants for a DFG project funding are advised to provide details on how they will handle and protect these types of data during and after the research project and whether they will make use of, e.g., data anonymization and/or encryption techniques. The latter techniques could allow the sharing of the data and make it more favorable for reuse and future new research insights (see section E). [BERD's interactive Virtual Assistant \(iVA\)](#) can help to guide researchers in this realm by showing them if the EU's General Data Protection Regulation (GDPR) applies to their data. Applicants are encouraged to also describe the processes they might use to anonymize qualitative data. We also recommend that the data ownership and copyrights for the data are clearly specified. This also increases interoperability of the data as one of the FAIR dimensions. If applicants seek to work with commercial data providers for their research project, we propose them to ensure that they comply with the terms and conditions specified by the institution. Additionally, when confronted with potential ethical considerations in the research process (e.g., when working with data generated from experiments involving minors), we suggest providing a comprehensive evaluation of the magnitude of these effects and how they will overcome them.

E) Data Exchange and Permanent Accessibility of Data

After the research project has been terminated, it is strongly encouraged to permanently store and share data publicly (in compliance with DSGVO and other legal obligations). This allows one to make the data findable, accessible, and reusable. We further advise applicants to outline to which extent this is possible and intended for their datasets and code. In the realm of business and economics, we strongly recommend the applicant to use the BERD platform for these purposes. BERD allows users to specify whether they want to share their data publicly or in restricted access. Restricted access can either mean that the data is openly available, yet usable only under specific conditions described in a standard license (e.g., Creative Commons, GNU General Public, or MIT Licenses), or that the data remains inaccessible unless access is granted through an application and approval process. In some instances, data may not be publicly shared for sound legal or ethical reasons. However, it could be turned open under specific restrictions, e.g., data and code files which are only accessible under supervision on a local computer hosted at a research institution for reviewing and replication purposes. Applicants could clearly outline such cases and the related reasoning in their proposal.

F) Responsibility and Resources

In their proposal, individuals are advised to specify who is responsible for the adequate management of research data during and after the research project. If no information on this topic is provided, it will be assumed that the first author oversees the appropriate research data use and is responsible for it as well as the infrastructure needed to process and store data. In addition, we also suggest adding if any costs associated with providing additional and not already commonly used infrastructure for data collection and documentation, analysis, storage and long-term archiving of the data are incurred because of the research project. Other than this, there are no further field-specific recommendations and the general DFG guidelines apply.

Acknowledgments

These recommendations were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of BERD@NFDI (www.berd-nfdi.de), Grant Number 460037581.