Guidelines for Ensuring the Reproducibility of Research at the LMU Munich School of Management

Version as of: December 15, 2025

Authors: Susanne Adler, Ralf Elsas, Marko Sarstedt, Thorsten Sellhorn, Ingo Weller

---

Outline

Preamble

High-quality research is only as strong as its ability to be understood and trusted by others. In many fields and journals, open science and reproducibility practices have already become the norm. Reproducibility is not an administrative constraint but a strategic asset for every researcher. It strengthens credibility, increases visibility and impact, and protects researchers when findings are scrutinized. Reproducible research travels better, survives longer, and is more likely to be published, cited, and built upon.

These guidelines are therefore not meant to slow researchers down, but recommendations to help us all meet external expectations reliably, to reduce the risk of later compliance problems, and to standardize good habits early in the research process. The goal is to make the LMU SOM's research easier to understand, easier to build on, and easier to defend - to the benefit of each individual researcher and of the LMU SOM as a whole.

1 Introduction

In alignment with the guidelines of the **Deutsche Forschungsgemeinschaft (DFG)**[1] and LMU's Code of Conduct[2], research outputs comprising empirical analyses, simulations, numerical computations, or experimental work produced by academic staff members (including doctoral students, post-doctoral researchers, and professors) of the LMU SOM should adhere to the **FAIR principles** (Findable, Accessible, Interoperable, and Reusable).

These "Guidelines for Ensuring the Reproducibility of Research at the LMU Munich School of Management" serve as the LMU SOM's recommended research standards. They take into account current best practices in different research areas of the LMU SOM, including appropriate distinctions between qualitative and quantitative research approaches. They shall be adapted regularly to evolving best practices.

These guidelines apply to:

- All *final* research outputs by academic members of the LMU SOM (such as journal articles, conference papers that count as publication, dissertations, and monographs), on and after the date of submission. Conversely, these guidelines do not apply to non-final work in progress, such as working papers (whether made publicly available or not) and presentation drafts.
- All academic members of the LMU SOM.

It is recommended that all research outputs follow these guidelines. Responsibility for adhering to best research practices in the sense of the guidelines remains with the authors of the respective research outputs.

2 Core Requirements for Reproducibility

2.1 Quantitative Studies

1. All research outputs (as defined in Section 1) should be reproducible. LMU SOM researchers should ensure that a reproducibility package is available for their research output and that it complies with these guidelines, regardless of whether it is prepared by them or by co-authors.
2. Data and code should be shared transparently in a repository, ensuring that all elements contributing to the results are accessible and reproducible. The repository should be established in a timely manner, ideally immediately when the research output reaches the stage of "final output", such as at the time of journal acceptance.
3. For doctoral dissertations,

    - Repositories are ideally established and made available with the thesis submission. At the latest, they need to be available before the end of the dissertation review period; delays must be coordinated with the respective reviewers/supervisors of the thesis.

[1] Guidelines of the DFG: https://www.dfg.de/resource/blob/174052/1a235cb138c77e353789263b8730b1df/kodex-gwp-en-data.pdf.
[2] Ordnung der Ludwig-Maximilians-Universität München zur Sicherung guter wissenschaftlicher Praxis vom 17. November 2023.

- Generally, such repositories need to be "frozen", such that all results remain reproducible in the originally submitted form. If the research is updated for subsequent journal submission or similar reasons, the repository can be amended, but without obstructing reproducibility of any of the results of the original submission of the thesis. Alternatively, an additional repository can be generated.
- If parts or single papers of the dissertation are under submission when the dissertation is submitted, the repository can be made non-public until the paper is accepted for publication.

4. Content and style recommendations are:

   - To make research outputs reproducible, code and data should be prepared such that an external researcher with adequate proficiency in handling data and software can reproduce the research output's results, tables, and figures with acceptable effort.
   - All code used to generate the analysis data from raw data should be included, even if the raw data cannot be provided.
   - Code that produces computational results such as estimates, simulations, model solutions, and visualizations should be included. Ideally, these programs reproduce all the computational exhibits in the paper and approved online appendices.
   - The code may be provided in any format compatible with commonly used statistical packages, software, or programming languages (e.g., Matlab, Python, R, Stata). Should the code require unusual or costly software, the LMU SOM Open Science Committee can help in finding a viable solution.
   - For research outputs that use primary data collected with surveys or experiments, the materials should include the survey instruments or the experimental instructions, code for survey or experiment collection mechanisms, and original instructions and details on the selection of human participants, unless this information is already provided as part of the work (whether in the main text or in an appendix).
   - If the data are from a commercial vendor or publicly available but restricted with regard to redistribution, the generation of pseudo data is recommended. Pseudo data are fabricated data which mimic the *structure* and key *statistical properties* of the original (restricted or confidential) data so that others can run the code, test the workflow, and reproduce results mechanically, without accessing the real data. Alternatively, detailed instructions on how to obtain the data can be posted in lieu of posting the raw data.
   - Where data are subject to contractual or legal restrictions (e.g., NDAs, data-use agreements, or health-data regulations), alternative disclosure methods must not breach contractual or data protection obligations.
   - The code and data should be accompanied by a README file that clearly indicates any omission of the required recommended parts of the reproducibility package due to legal requirements, limitations, or other approved agreements.
   - Technical style recommendations are provided in Appendix 1 to these guidelines.

2.2 Qualitative Studies

The current best practices for qualitative studies (as detailed by the Top 5 relevant journals in the field, e.g., Administrative Science Quarterly) do not mandate reproducibility. However, it is recommended to strive for full transparency, e.g., by generating transcripts of interviews and making them accessible in public repositories, clarifying data usage (and academic access for

such purposes) with interviewees as early as at the pre-interview stage, providing detailed study information, including details regarding design plan, data collection, analysis plan etc.

Recommended guidelines are:

1. The process by which data were collected, the steps taken to aggregate and analyze the data, and the way findings and the theory generated emerged from the data, are expected to be detailed and made public.
2. Researchers are supposed to state clearly in the methods section where the study took place, when the data were collected (i.e., months/years in which different sources of data were collected), how sampling took place, which types of data were collected, how and for what purposes data was collected, and who collected the data. Researchers should also reflect on their own role in the knowledge-building process.
3. Researchers should further communicate what population the findings are based on, as well as the reasoning for choosing a particular population, how cases were selected, and which stopping criteria terminated data collection.
4. Generated data should be made available as far as possible in an ethically responsible manner.
    a. If data collections involve interviews, an appendix with an interview protocol and an anonymized description or table of interviewees is needed. Similarly, aggregated data, coding schemes etc. should be made available.
    b. To provide data transparency even if confidentiality of the underlying raw data is required, researchers should apply de-identification, such as using confidential identifiers, pseudonyms, or remediation of sensitive content when presenting data gathered from different people, occupations, groups, or organizations.

If space limitations or other reasons prevent sufficient explanation and transparency in the research output itself, a special "reproducibility" package should be designed and put in a repository, similar to the recommendations outlined in Section 2.1.

In addition:

5. If analyses are based on statistical methods or use statistical software (e.g., for coding, for the analysis of coded items, or to illustrate or test sample characteristics), the code should be made available and put in a reproducibility package (see Section 2.1).

3 Copyright and Access to Reproduction Packages

1. It is desirable to make all data and code used in the research publicly available, without any restrictions. The authors' copyright is not affected.
2. It is recommended to use Creative Commons Attribution International 4.0 (CC-BY) for data and any documents, and the modified BSD license for software and code. However, researchers are free to choose other licenses if they allow third parties to use the data and code for reproduction purposes.
3. At their own expense, academic staff members of the LMU SOM may opt for protection whereby any person who downloads any of the files should certify that the files will be used solely for academic purposes.
4. Levels of Repository Accessibility: Depending on legal and privacy requirements, academic staff members of the LMU SOM can choose one of four approaches:
    a. *Public Repository*: Research data and code are deposited in publicly accessible repositories such as the ones from Open Science Framework (OSF), GESIS or ZBW.

b. *Public Repository with Anonymized Data***:** In the case of legal constraints, datasets can be anonymized and shared with code to ensure reproducibility.
c. *Protected Repository*: For sensitive data with strict legal and privacy restrictions, repositories such as the *EBDC at LMU's ifo-Institute* can be used, which allow controlled access for reproduction purposes.
d. *No Repository (only in exceptional cases)***:** If none of the repository options above are feasible, a concise written justification should explain why a repository could not be used.

---

4 Approved and Inadequate Repositories

4.1 Approved Repositories

Academic staff members of the LMU SOM are encouraged to use repositories that support the FAIR principles and offer unlimited lifetime, including (but not limited to):

- GESIS (Leibniz Institute for Social Sciences): Ideal for social science datasets.
- ZBW (Leibniz Information Centre for Economics): Specialized in economic datasets.
- EBDC (ifo Institute): Provides controlled access to protected data.
- Open Science Framework (OSF): General-purpose repository supporting transparency.
- Repositories maintained by journals where a research output is published.

4.2 Inadequate Repositories

The following repositories or methods are considered **inadequate**:

1. Personal Websites: Personal or institutional web pages are unreliable due to a lack of long-term accessibility.
2. Temporary File Sharing Services: Platforms like Dropbox, WeTransfer, or Google Drive are not suitable due to limited accessibility, poor metadata support, and potential file expiration.
3. Repositories without Persistent Identifiers: Repositories should provide DOIs or similar identifiers to ensure the data can be reliably cited.
4. Proprietary or Restricted Platforms: Repositories that require exclusive access or subscriptions without clear access policies for reproduction purposes should be avoided.

By adhering to these recommendations, academic staff members ensure that their work meets reproducibility standards while maintaining scientific integrity.

---

5 Steps to Ensure Compliance with the Guidelines

1. **Research Design and Data Management**
    - Reproducibility goals should be integrated into a research project as early as the planning stage.
    - Best practices in documenting all data sources, methods, and analyses should be followed.
2. **Privacy and Legal Compliance**

- Compliance with data protection laws (like GDPR) and agreements with data providers should be ensured and should be considered already at the design stage of research.
- Sensitive datasets can be anonymized if required; pseudo data can be created and used if full datasets cannot be shared.

6 LMU SOM Open Science Committee

The LMU SOM Open Science Committee ("the Committee") provides advice and recommendations on how to adhere to the guidelines and best practices.

The Committee can and will not test or validate reproducibility packages, except for special circumstances (e.g., in the context of §21 of LMU's Code of Conduct about scientific misconduct).

If LMU SOM researchers seek clarification or plan to deviate from these guidelines, the Committee can give its opinion and provide advice.

7 Resources and Support

The Master of Business Research (MBR) program at the LMU SOM has been extended to contain a mandatory course on the foundations of Open Science in Module A/II.

Other helpful resources:

- *LMU Open Science Center*[3]: Workshops and materials for implementing open science practices.
- *Research Data at LMU*[4]: Aimed at LMU members of all disciplines; provides comprehensive resources, tools and information to effectively manage research data and simplify its planning, organization, storage, analysis and publication.
- *EBDC (ifo-Institute)*[5]: A repository offering controlled access to sensitive datasets.
- *GESIS*[6] and *ZBW*[7]: Public repositories supporting the FAIR principles.
- *Center for Open Science* (OSF)[8]

These guidelines foster a culture of transparency and accountability while safeguarding ethical and legal standards. Academic staff members are encouraged to seek further assistance from their supervisors or the LMU Open Science Center.

[3] https://www.osc.uni-muenchen.de/index.html
[4] https://fdm.ub.lmu.de/
[5] https://www.ifo.de/ebdc
[6] GESIS Datenservices: Archivierung BASIS
[7] ZBW: https://openeconomics.zbw.eu/ ,
[8] OSF: https://www.cos.io/products/osf

Appendix 1: Preparing Reproducible Code: Example Instructions

To ensure reproducibility, academic staff members of the LMU SOM are encouraged to prepare their code according to accepted best practices (such as provided by the **American Economic Review (AER) Guidelines**).

Code Organization

- Clearly separate code into logical sections or scripts for:
  - Data preparation and cleaning
  - Main analysis
  - Output generation (e.g., tables, figures)
- Label files systematically to reflect their sequence and purpose (e.g., `01_DataPrep.R`, `02_Analysis.R`).

Documentation

- If researchers choose to include a **README file**, it should include:
  - An overview of the research output
  - Step-by-step instructions for executing the code
  - Required software and packages, including version numbers
  - Information about the hardware environment if results are hardware-dependent
- Use inline comments within scripts to explain key steps and logic

Modularity

- Ensure that the code is modular and avoids hardcoding file paths, variable names, or dataset specifics. Use relative paths and parameterize key settings for adaptability.
- A master script is strongly encouraged. When no master script is included, please provide sufficient and precise step-by-step instructions, allowing users to reproduce the generated outputs exactly and with the least amount of effort.
- When additional packages or libraries are required to run the code, provide a setup program, containing commands to download and install the necessary packages or libraries.

Input Data Files

- Provide sample or pseudo data if the original dataset cannot be shared due to legal or privacy concerns.
- Include scripts to simulate pseudo data with similar structure and characteristics.

Reproducibility Testing

- Conduct a dry run of all scripts on a clean system (ideally by someone other than the author) to ensure they execute correctly from start to finish.

Software and Dependencies

- Specify all software and dependencies, including:
  - Programming languages (e.g., R, Python, Stata) and their versions
  - Libraries and packages with installation instructions
  - External tools or platforms, if applicable (e.g., cloud services, proprietary software)

Error Handling

- Incorporate error messages and warnings to guide users in resolving issues if they occur during code execution.

Automation

- Use automation tools where feasible to simplify the execution process, such as `Make-files` or workflow managers (e.g., Snakemake).