



**SFB
1287**

Limits of Variability in Language
Cognitive, Computational, and Grammatical Aspects

DATA MANAGEMENT PLAN
SFB 1287 / 2025 / PHASE 2

PROJECT Q

Table of contents

General Information	1
Overview	1
Data Management Requirements	2
Financial Support	2
Dataset Information	3
Data Origin	3
Data Collection	3
Data Handling	3
Data Analysis	4
Data Reuse	4
Legal and Ethics	5
Personal Data	5
Property Rights	5
Publication	6
Publishing or Sharing Data	6
Storage and Long-Term Preservation	7
Archive	7

General Information

Overview

Project number

Q

Name of Experiment / Acronym / Number

Reproducibility

Involved persons

Anna Laurinavichyute, Shravan Vasishth, Himanshu Yadav

PI or responsible person (head of the study)

Shravan Vasishth

Subject area

metascience, meta-research

Method / Type of data

We work with data made available by the authors of published papers and try to reproduce the results reported in the papers.

Participants (of the study)

-

Short description (of the study)

In 2019 the Journal of Memory and Language instituted an open data and code policy; this policy requires that, as a rule, code and data be released at the latest upon publication. We compared 59 papers published before, and 59 papers published after, the policy took effect. After the policy was in place, the rate of data sharing increased by more than 50%. We further looked at whether papers published under the open data policy were reproducible, in the sense that the published results should be possible to regenerate given the data, and given the code, when code was provided.

Comments (optional)

-

Data Management Requirements

Are there requirements regarding the data management from your scholarly / scientific community?

yes

If yes, what are the requirements?

- *DFG Guidelines on the Handling of Research Data*
- *Data Management in Psychological Science*
- *„Handlungsempfehlungen zum Umgang mit Forschungsdaten“ University of Potsdam*
- *„Technische und organisatorische Maßnahmen“ (TOM) gemäß Art. 32 Abs. 1 DSGVO.*

Financial Support

Who is funding the project?

DFG - Deutsche Forschungsgemeinschaft e.V. (German Research Foundation) - <https://www.dfg.de/en/>

In which funding line and / or which funding program is the project funded?

Collaborative Research Centre 1287 - Project number 317633480

Dataset Information

Data Origin

Is the dataset being created or re-used?

Re-used

If re-used, who created the dataset and under which address, PID or URL is the data set available?

The 59 datasets used in the project were created by different teams of scientists; they were made available together with the papers published in the Journal of Memory and Language. The list of the surveyed papers is published in our paper summarizing the findings of the project.

Data Collection

When does data collection start? (approximately / tentatively)

-

When does data collection end? (approximately / tentatively)

-

Data Handling

Where is the dataset stored during the project?

Personal computers + gitUP-Server

If data is stored on lab or personal computers, please describe the backup strategy.

regular use of gitUP + daily automatic backups to home server

Which file formats are used?

Plain text, excel, latex

Which measures of quality assurance are taken for this dataset?

four-eyes principle, code review

Data Analysis

When does data analysis start? (approximately / tentatively)

01.09.2020

When does data analysis end? (approximately / tentatively)

01.02.2022

Data Reuse

Which individuals, groups or institutions could be interested in re-using this dataset? What consequences does the reuse potential have for the provision of the data later?

The authors of the surveyed papers as well as the general readership of the Journal of Memory and Language; the scholars interested in metascientific research.

Legal and Ethics

Personal Data

Does this dataset contain personal data?

yes

Are these data anonymised?

yes

Property Rights

Does the project use and/or produce data that is protected by intellectual or industrial property rights?

no

If yes, please explain which data protected by intellectual or industrial property rights?

-

Publication

Publishing or Sharing Data

Will this dataset be published or shared?

yes

If yes, the principal investigator of the study ensured that the consent form / subject information sheets support publishing of the data?

yes

If yes, under which terms of use or license will the dataset be published or shared?

Public domain (CC0)

If yes, when will the data be published?

Published under: <https://github.com/annlaurin/reproducibility>

In Publication:

<https://www.sciencedirect.com/science/article/pii/S0749596X22000195?via%3Dihub>

If no, please explain why not. Please differentiate between legal and contractual reasons and voluntary restrictions.

-

Storage and Long-Term Preservation

Archive

Does this dataset have to be preserved for long-term?

yes

How long does the data need to be stored?

The DFG expects primary data that is the basis of a publication to be stored in the researcher's own institution or an appropriate nationwide infrastructure long-term (for at least 10 years).

What are the reasons this dataset must be preserved for the long-term?

- *Use in a publication / Evidence of good scientific practice*
- *Reuse (if anonymizable data) in subsequent projects or by others*
- *Legal obligations*
- *Documentation because it is socially relevant*
- *Self-commitment*
- *Evidence of good scientific practice*
- *DFG requirements*

Where will the data (including metadata, documentation, and relevant code) be stored or archived after the end of the project?

- *SFB 1287 File-Server*
- *OSF*
- *Research Data Server from Project IN-FDM-BB (a.t.m. not available)*
- *GitHub*

coordinated by:



<https://www.uni-potsdam.de>

funded by:

DFG Deutsche
Forschungsgemeinschaft

<https://www.dfg.de/>

in cooperation with:

**RUHR
UNIVERSITÄT
BOCHUM** **RUB**

<https://www.ruhr-uni-bochum.de>