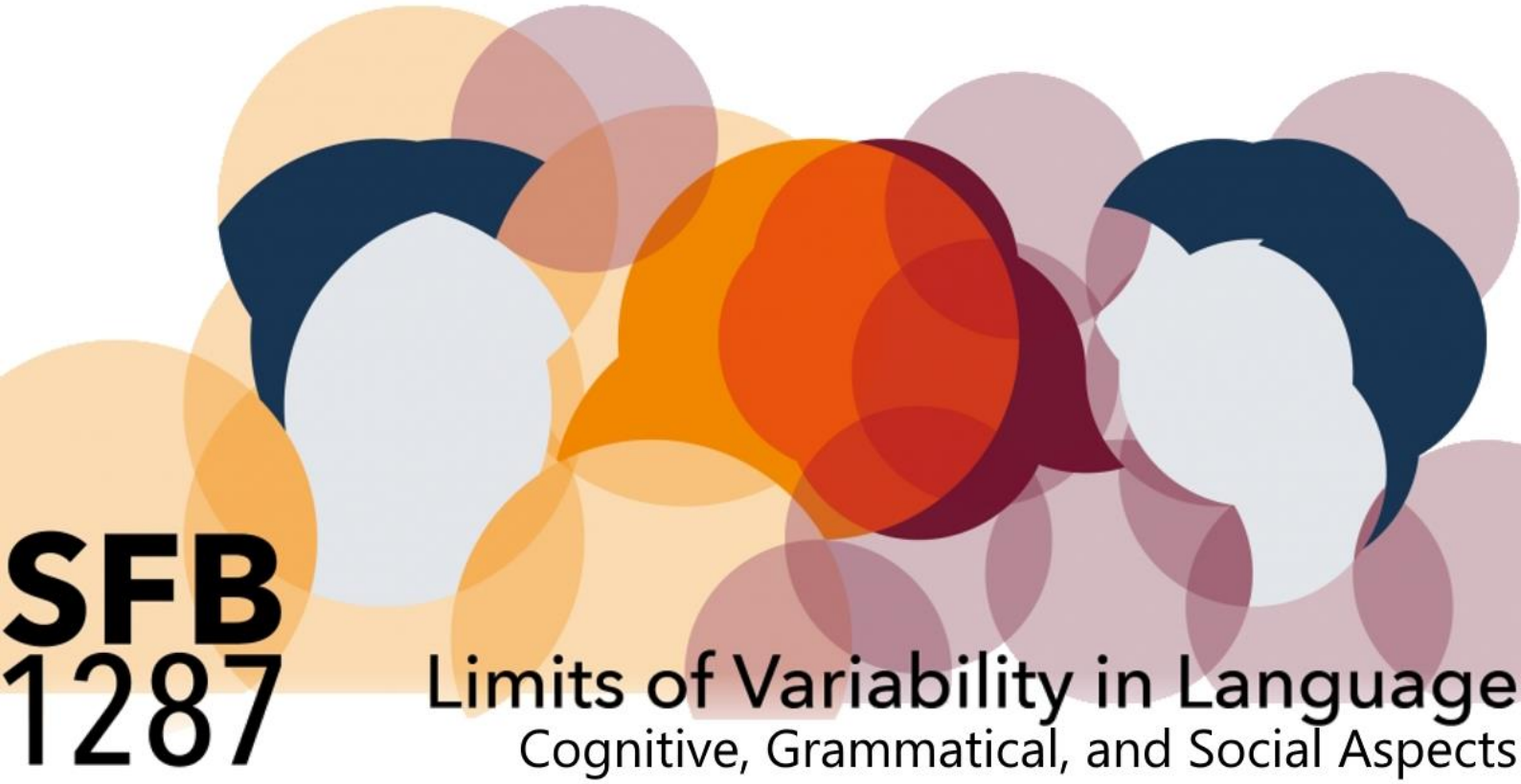
DATA MANAGEMENT PLAN

SFB 1287 / 2021 / Phase1 / Project B05



SFB 1287

Limits of Variability in Language

Cognitive, Grammatical, and Social Aspects

TABLE OF CONTENTS

GENERAL INFORMATION

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

- yes

(a)If yes, what are the requirements?

- DFG Guidelines on the Handling of Research Data
- Data Management in Psychological Science

(2) What kind of dataset is it?

The following questions collect information on the data that is produced or used in the project. They also help to estimate the value of the data in terms of potential re-use and long-term preservation. In the case of personal data, the principle of data minimization (Art. 5 EU General Data Protection Regulation) allows the collection of personal data only when there are no other reasonable means to clarify the research question (re-use of existing data would be such a reasonable means). Also, there shall be no more information collected than necessary. The information regarding the data collected, produced, or used in the project is gathered along the datasets. The definition of these datasets is an important conceptual decision that has to be done individually and carefully for each project.

Name of Experiment / Acronym / Number:	Experiment 02 (Speeded picture word interference task)
PI or responsible person (head of the study):	Audrey Burki
Other persons involved:	Sylvain Madec, Thea Villinger, Shereen Elbuy, Hyein Jeong, Pamela Fuhrmeister
Subject area:	Language production
Method / Type of data:	Behavioral data collected with computerized and paper tasks, with recording of vocal responses to isolated words and measures of performance on several non-linguistic cognitive tasks
Anonymizable data:	yes
Participants:	48 young adult (18-30 years old). Right handed.
Short description (of the study):	In the main task of this study participants perform a picture word interference task (i.e., pictures are presented one by one on the screen with superimposed written words, participants are instructed to name the pictures) in two conditions. In the first (speeded condition), filler trials are displayed for a shorter duration than test trials. In the second (slow condition), filler and test trials are displayed for a similar duration. In addition, participants perform a series of secondary tasks: A recall task where participants are presented with the written words of the picture-word interference task and have to indicate whether they have seen them in the experiment or not; and reading aloud task where they name all the written words aloud,

Short description (of the study):	a delayed naming task where they name the pictures of the main task after a delay, span tasks (operation span task, symetry span task, rotation span task), inhibition tasks (Flanker task, Simon task, stop signal task), attentional tasks (continuous temporal expectancy task, conjunctive continuous performance task), Edinburgh Handedness Inventory, the mehrfachwahl-Wortschatztest (MWT), and the digit symbol test (WAIS).
Other comments:	
Time for data collection (approximate):	01/09/2019 - 30/03/2020
Time for data analysis (approximate):	01/04/2020 - 30/10/2020
Related publications:	Fuhrmeister, P., Madec., S., & Bürki, A. (in preparation). The role of task difficulty in assessing individual differences in word production (data: https://osf.io/4ftex/files/).
Keywords (used in publications):	Language production, attention, inhibition, working memory, variability
Link to preregistration:	
Funding reference:	Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 317633480 – SFB 1287

(3) 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?

It is important to specify whether the data will be permitted for reuse. But legal impediments, such as privacy, and copyright must be taken into account.

TECHNICAL INFORMATION

(4) Where is the dataset stored during the project?

Please delete all project-files from source that is not part of the University of Potsdam.

for raw data:	SFB1287 - File Server
for analysis data:	Researcher´s Computer SFB1287 - File Server Personal Dropbox
for further documentation, related code, or software:	Researcher´s Computer SFB1287 - File Server

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

- The university cloud and SFB-File Server is backed up regularly.

(5) Which file formats are used?

When choosing a data format, one should consider the consequences for collaborative use, long-term preservation as well as reuse. It is advisable to use formats that are standardised, open, non-proprietary, and well-established in the respective scholarly community. A table with recommended file formats can be found in found in Kristin Briney, Data Management for Researchers, Pelargic, 2015, pages 133-134.

Raw data:

- PWI task: wav files recorded during the experiment for each trial/participant; log file generated by the software Presentation for each participant
- recall task: log file generated by the software Presentation for each participant
- delayed naming task: wav file for each trial participant
- reading task: wav file for each trial participant
- operation span task: a csv file for each participant
- symmetry span task: a csv file for each participant
- rotation span task: a csv file for each participant
- Flanker task: a csv file for each participant
- Simon task: a csv file for each participant
- stop signal task: a csv file for each participant
- continuous temporal expectancy task: a csv file for each participant
- conjunctive continuous performance task a csv file for each participant
- Edinburgh Handedness Inventory, mehrfachwahl-Wortschatztest (MWT), and digit symbol test (WAIS): we enter the results on paper, later copied in csv files.

Pre-processed data:

- a csv file with the results (accuracy, response times) for each trial of the main task, score on other tasks for participant
- a txt file describing the columns of the above csv file

PUBLICATION

(6) Will this dataset be published or shared?

anonymizable data

- *ID will be removed (and or code-list will be destroyed) [legally correct: code list will be destroyed as soon as possible without jeopardizing experiment; exception: follow-up study planned, if so, talk to UP data protection officer (Dr. David Kneis; mailto:datenschutz@uni-potsdam.de) on how to do this correctly]*
- *Publication of anonymized data and code on OSF or RADAR (University of Potsdam) (or as required by the Journal)*

non-anonymizable data

- *on RADAR (University of Potsdam) but not accessible from the outside world*

❖ yes

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

The options refer to the licenses of the Creative Commons family. If data is anonymised / pseudonymized, it's probably not legally required, but might be good in terms of research ethics to adjust consent forms / subject information sheets.

✓ **Principal investigator of the study assures that the consent form / subject information sheets support publishing of the data.**

for data: CC 0 (Public Domain) (recommended)

for scripts: none (shared only internally)

for software:

(b)If yes, when will the data be published?

Recommended procedure: Upload data and obtain digital identifier (e.g., DOI, OSF link) when submitting the first paper; thus, you can cite the data in the paper. If necessary, restrict public access (embargo) until last paper published (max. 2 years).

- ❖ when the paper is published

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

(7) Which measures of quality assurance are taken for this dataset?

- ❖ P.I. checked the pre-processed file and analyses

LEGAL AND ETHICS

(8) Does this dataset contain personal data?

The EU General Data Protection Regulation (GDPR) defines in Art. 4 personal data as "any information relating to an identified or identifiable natural person". An identifiable natural person is "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

For advice and/or trainings on how to comply with privacy regulations - including proper anonymization & pseudonymization - you can always contact UP's privacy officer, Dr David Kneis, at datenschutz@uni-potsdam.de. As the privacy officer, his perspective is focussed more on the legal side of things than the research ethics or technical aspects.

- yes

(9) Are the data anonymised?

Anonymised data: ID will be removed (and or code-list will be destroyed) [legally correct: code list will be destroyed as soon as possible without jeopardizing experiment; exception: follow-up study planned, if so, definitely talk to UP data protection officer on how to do this correctly]

for raw data:	No
for analysis data:	Yes
for published data:	Yes

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

Measurement data has no intellectual property, so usually, the answer here will be „no". Data or software can be subject to intellectual or industrial property rights. Applicable laws differ broadly even within EU. According to the German copyright law (UrhG) works of literature, scholarship and the arts that can be regarded as a "personal intellectual creation" are protected by copyright. Mere data, e.g., measured data or survey data, and metadata (except in some cases descriptive metadata) are not protected by copyright. § 2 of the UrhG lists the following kinds of protected works (list is not concluded):

- *linguistic works such as written works, speeches, and computer programs*
- *works or the fine arts including works of the applied arts as well as sketches of such works*
- *works of photography*
- *descriptions and illustrations of scholarly or technical nature such as drawings, plans, maps, sketches, tables, and three-dimensional representations*

According to § 3, copyright is also applicable to translations and other modifications or adaptions of work if they are individual intellectual creations of the editor. Finally, according to § 4 copyright also extents to collected editions and database works. Collected editions are: "collections of work, data or other independent elements that are individual intellectual creations based on the selection and arrangement of the elements".

Database works are defined as "collected editions, the elements of which are arranged in a systematic or methodical way and can be accessed individually by electronic means or in other ways".

❖ no

(a)If yes, please explain which!

STORAGE AND LONG-TERM PRESERVATION

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

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).

- At least 10 years after the end of the first funding period of the SFB1287.

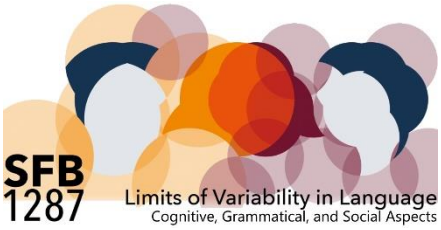
(12)What are the reasons this dataset has to be preserved for the long-term?

- Because of the DFG expectation.

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

- OSF
- SFB1287 - File Server

REFERENCES



SFB 1287
https://www.sfb1287.uni-potsdam.de

University of Potsdam
https://www.uni-potsdam.de

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Deutsche Forschungsgemeinschaft e.V.
https://www.dfg.de