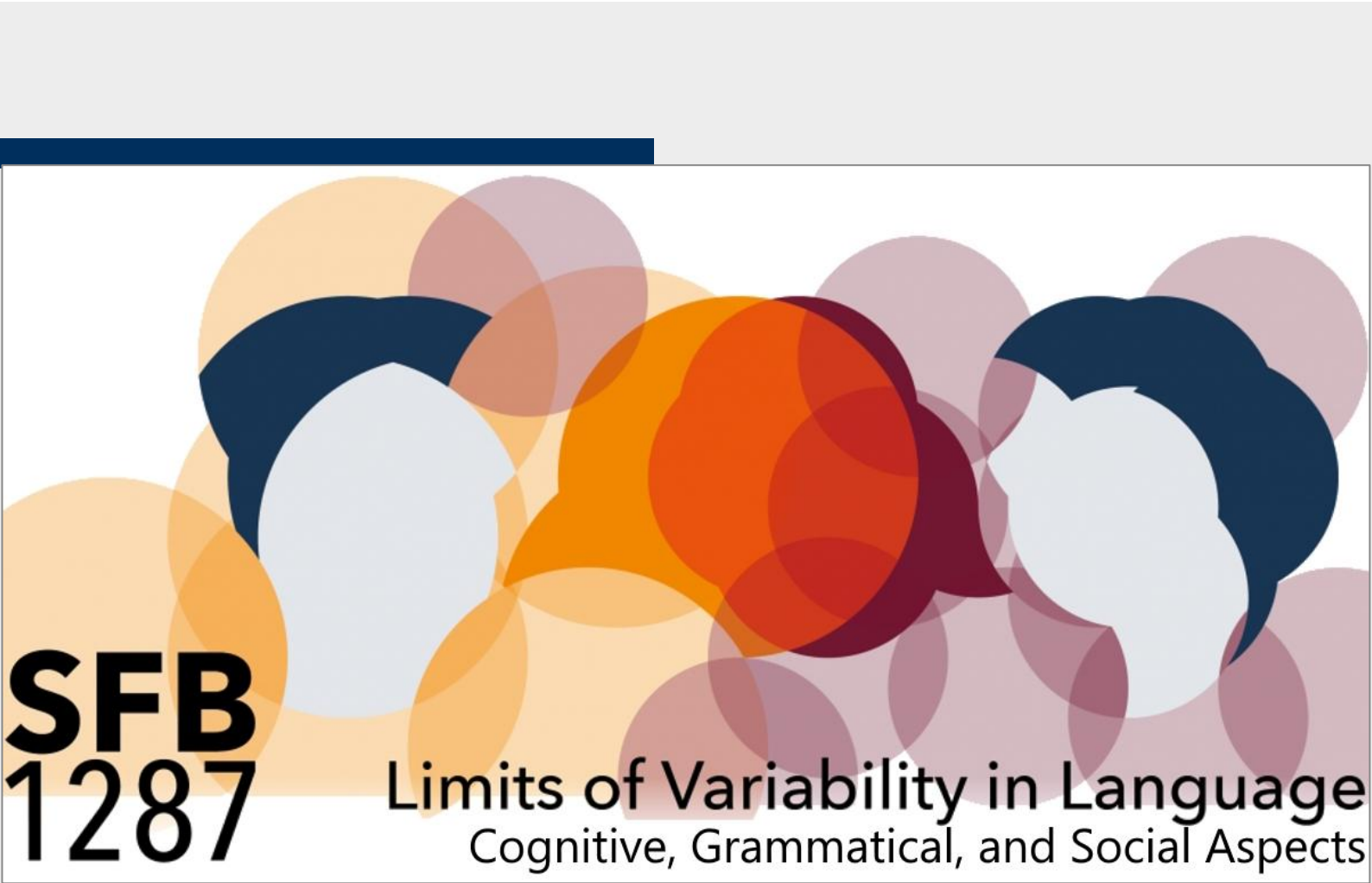




**SFB  
1287**

**Limits of Variability in Language**  
Cognitive, Grammatical, and Social Aspects

# DATA MANAGEMENT PLAN

**CRC 1287 | 2021 | Phase 1**  
**Project A03**

.....

## TABLE OF CONTENTS

|                                                 |           |
|-------------------------------------------------|-----------|
| <b>General Information .....</b>                | <b>3</b>  |
| Overview .....                                  | 3         |
| Data Management Requirements .....              | 4         |
| Financial Support .....                         | 4         |
| <b>Dataset Information .....</b>                | <b>6</b>  |
| Data Origin .....                               | 6         |
| Data Collection .....                           | 6         |
| Data Handling.....                              | 6         |
| Data Analysis .....                             | 7         |
| Data Reuse .....                                | 8         |
| <b>Legal and Ethics .....</b>                   | <b>9</b>  |
| Personal Data.....                              | 9         |
| Property Rights.....                            | 9         |
| <b>Publication .....</b>                        | <b>10</b> |
| Publishing or Sharing Data.....                 | 10        |
| <b>Storage and Long-Term Preservation .....</b> | <b>11</b> |
| Archive .....                                   | 11        |

.....

.....

## GENERAL INFORMATION

### Overview

#### **Project number**

A03

#### **Name of Experiment / Acronym / Number**

Twitter English Clefts

#### **Involved persons**

Luca Bevacqua (STF)

#### **PI or responsible person (head of the study)**

Tatjana Scheffler

#### **Subject area**

Linguistics

#### **Method / Type of data**

web corpus data

#### **Participants (of the study)**

online web data crawl

.....

### **Short description (of the study)**

Tweets including at least one cleft were collected on 25th to 29th March 2019 using TAGS Hawksey, 2016). The queries were url encoded to target specific tweets without excluding other factors such as emphasis.

### **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
- 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 317633515

.....

.....

## DATASET INFORMATION

### Data Origin

**Is the dataset being created or re-used?**

reused

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

Twitter

### 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?**

- CRC file server
- Box.UP university cloud
- researcher's computer
- computer in the laboratory

.....

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

- Sharing data through box UP
- Regular backups are taken to an external hard drive.
- The university cloud and CRC File Server is backed up regularly.

**Which file formats are used?**

- Corpus\_raw\_data: . csv
- Corpus\_analysis: R script
- Readme file: .txt explaining information in raw data

**Which measures of quality assurance are taken for this dataset?**

manual annotation with partial reannotation, adjudication

**Data Analysis**

**When does data analysis start? (approximately / tentatively)**

April 2019

**When does data analysis end? (approximately / tentatively)**

September 2019

.....

## **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?**

-

.....

.....

## LEGAL AND ETHICS

### Personal Data

**Does this dataset contain personal data?**

yes

**If 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?**

yes

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

Twitter

.....

.....

## PUBLICATION

### **Publishing or Sharing Data**

**Will this dataset be published or shared?**

no

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

-

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

-

**If yes, when will the data be published?**

-

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

Sharing is not allowed (licensed corpora, Twitter terms of use)

.....

## 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?**

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

.....

coordinated by:



Universität Potsdam

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

funded by:



Deutsche  
Forschungsgemeinschaft

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

in cooperation with:

RUHR  
UNIVERSITÄT  
BOCHUM

RUB

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