

PROJECT “S”

DATA, DESIGN AND SUSTAINABILITY



Job Schepens
Statistical Analysis
job.schepens@uni-koeln.de



Luke Günther
Data Processing
luke.guenther@uni-koeln.de

Hey, and welcome to the CRC 1252 "Prominence in Language"!

We are project S, your one-stop-shop for data management and statistical analysis.

Not only will we help you figure out how to access the CRC's technical infrastructure but also guide you through the myriad of research tools out there, wrangle with your data until it's readable and teach you how to do a proper statistical analysis. Together, we'll learn all about sustainable and reproducible research and how to practice open science.

– Job and Luke

What we do

1. Research Data Management

- Teach you the skills to master the **research data life cycle**: from experimental design to data collection/organization, curation/analysis, publication/archiving and, finally, re-use.
- Guide you through making your research data and results **open access** to facilitate knowledge transfer and collaboration and to ensure resource re-usability and transparency.
- Create the **long-term archives** for all projects and collect metadata about your files. This way your work is preserved for the future and properly credited.

2. Data Analysis Support

- Offer **workshops and one-on-one support** for conducting statistical analysis using the R programming language. Don't worry, no prior knowledge required.
- Train you in **Mixed/Bayesian modelin** to answer complex questions. Also not as scary as it sounds.

3. Technical Support and Annotation

- Pre- and post-process your data. Don't know how to convert your files from format X to format Y? Can't figure out what to do with that weird looking export once you're (finally) done annotating? We can help.

- Search for tools, corpora and other resources to find those that fit your needs and show you how to use them.
- Manage the technical infrastructure, including Sciebo and various annotation tools (more information below). If you need assistance setting up, just let us know.

Where to find us

At our beloved *HoP* (*“House of Prominence”*), *the CRC’s office headquarters* (Luxemburger Str. 299):

Go up to the second floor and you’ll find us in the third office to the right. It has our names on it but our door is usually open anyway. Say hello to our office neighbour Timo, too, while you’re there.

If we are nowhere to be found, you can always reach us via e-mail (see above) and we are happy to help you right away and/or schedule in-person or virtual meetings with you.

Our tools

Store your research data on [Sciebo](#), a *secure cloud storage provider for universities* in NRW, used by all researchers at the CRC. This is where you can place your corpora, notes and scripts and access your project’s files as well as our resources and administrative documents. It also offers an integrated suite of office applications (think Microsoft Word but in your browser/Google Docs but private).

Annotate your text documents with [INCEpTION](#). Mark relations between tokens, provide labels and link reference chains to add semantic context.

Transcribe *video and audio recordings*, create hierarchical annotations for multiple participants, *gloss your corpus*, export dictionaries and more with [ELAN](#).

Create *transcriptions and annotations for audio recordings* and do phonetic analysis using [Praat](#).

Get an overview of your corpus, select examples and *analyze language in use* with [SketchEngine/NoSketchEngine](#).

Further reading

[What is Research Data Management \(RDM\)?](#) – Universität Tübingen

[What is Research Data?](#) – C³RDM

[Open Science Stories](#) – previous projects from our CRC!