DDI Variable Documentation and data access using R

Claus-Peter Klas[*][1]

[1]GESIS - Leibniz-Institute for the Social Sciences – Germany

Abstract

Processing statistical files produced by STATA or SPSS are used for the documentation of research data. They contain information about the variables, explicitly variable name, and label, as well as the response scales. But foremost the statistical files contain the data. So, besides the general metadata documentation also frequency tables, summary statistics or cross tables are created for documentation purposes based on the files.

To enhance and automate the documentation process for archiving research data, I created an architecture based on R, exposed as generic REST API, partially based on the DDIwithR package. This demonstration will present the general architecture, how it is integrated into the DDI documentation process within the GESIS questionnaire editor. I will show the features consisting of

- Reading the statistical file to connect questions in the questionnaire editor with variable metadata and store it in DDI LC 3.2
- Generating statistical information such as summary statistics and store it as variable statistics in DDI.
- Furthermore, the package is technically able to give access to the actual data for direct usage in R, without downloading the SPSS/STATA file, either the complete dataset or singled out variables via PIDs.

---

[*]Speaker