RTG 2636

Form-meaning mismatches

Data Policy

PREAMBLE

The PhD students, postdocs, and other researchers of the DFG Research Training Group (RTG) 2636 "Form-meaning mismatches" are collecting/compiling extensive datasets, leading to various publications and other instances of knowledge dissemination. In addition, the RTG offers opportunities for collaborative results and publications across projects. The RTG – designed as a long-term project of international visibility and cooperation – therefore requires careful data documentation, quality, storage, and accessibility, an endeavor guided by the FAIR data principles and striving to go conform with the developing NFDI initiatives. These aspects are the focus of this data policy.

§1 Coverage

a) This policy applies to all members of the RTG 2636: Participating Researchers (PRs), doctoral and postdoctoral researchers (irrespective of individual funding source), and (student) research assistants, in as far as research activities are carried out within the RTG and data that were generated within the RTG will be used.

§2 Data Management and Publication Committee

a) A Data Management and Publication Committee (DMPC) composed of the Governing Board (GB) members and the Data Management Representative (DMR) sets the data standards with respect to permissible file formats, and level and format of documentation (metadata). The DMPC also serves to adjudicate possible disputes relating to this policy.

§3 Rights and responsibilities

a) Each member is responsible for entering his or her data into the central RTG "Form-meaning mismatches" database (short "database").
b) The DMR, assisted by the coordination office, oversees the operational management of this database and compliance with this policy. The DMR decides on acceptable file formats and appropriate documentation of provided data. Decisions will be informed by the CLARIN infrastructure and the evolving NFDI initiatives

§4 Documentation of datasets

a) Each dataset must be documented with meta-information. Meta-information shall be compatible with community standards as recommended by the DMPC and provide a level of detail guided by the FAIR principles. All datasets will be stored in the database at least with information on who collected them, which version of the data it contains, the date of the latest update, as well as any potential use restrictions.
b) The coordination office will provide templates to facilitate meta-information and data entry.

§5 Access to data

a) Each member of the RTG has unrestricted access to the meta-information in the database.
b) Each member should usually have access to his or her own datasets. After receiving permission from the researchers who originally collected (obtained) the data and the PRs of the respective project, data can be shared with other members. All data sharing is moderated by the coordination office. Members making use of data supplied by other researchers of the RTG for scientific publication must acknowledge their use of the data appropriately. Potential co-authorship has to follow the rules of good scientific

practice. Before anything is submitted for publication, the PRs involved in the respective projects must give their permission.

c) Members pledge that they will not pass on any data or passwords that enable access to the database to third parties. Until datasets are made publicly available (see point e of this paragraph), release of data to non-members of the RTG 2636 will be an exception for which explicit permission must be sought from the original data suppliers and the PRs involved in the respective projects after consultation with the DMPC. Original data suppliers are the members of the RTG who were responsible for data collection/compilation.

d) All (non-confidential and non-licensed) data will be made available internally within the RTG 2636 two years after the person that collected (obtained) the data has left the program or completed his or her dissertation.

e) All (non-confidential and non-licensed) data will be made publicly available if applicable in thematic repositories as, for instance, recommended by CLARIN-DE or in the repository of the University of Göttingen (https://data.goettingen-research-online.de) at the latest five years after the person that collected (obtained) the data has left the RTG or completed his or her dissertation. All data will be referenceable with an identifier, preferably a DOI. Open licenses such as CC-BY will be applied.

f) If there are legitimate reasons, the PRs of the project in which particular data were collected can apply to postpone the procedures described in points d and e of this paragraph. Decisions on such matters will be made by the DMPC.

§6 Use of data

a) Data use must always be based on an agreement between an original data supplier and the data user. The following procedure will be implemented:
- The data user discusses the intended use with the data supplier.
- The data supplier communicates his or her agreement, in which the involvement in the use of the data is regulated. Agreements on co-authorship of publications must follow the rules of good scientific practice. Publications must acknowledge that the data were compiled in the RTG "Form-meaning mismatches", funded by the German Research Foundation (DFG).
- The DMR, assisted by the coordination office, provides the data of the data supplier to the data user upon receipt of this agreement. This ensures that the exact data version used can be documented.
- The DMR, assisted by the coordination office, keeps the records of these agreements.

b) Data accessed by a member should only be used for purposes necessary to carry out his or her own research within the RTG 2636 as specified in the agreement. Data accessed may only be used for scientific purposes, i.e., commercial use of data is not allowed. It is prohibited to distribute other researchers' data to a third party without the explicit consent of the original data supplier and the respective PRs.

§7 Delivery of data and quality control

a) The project and meta-information must be made available by the respective researcher no later than three months after receiving the request from the coordination office. Every member is responsible for the supply of his or her own data within this period.

b) Complete data must be deposited as quickly as possible, normally within one year of collection. An updated version of the data is to be deposited in parallel with handing in the dissertation (in the case of doctoral researchers), finalization of other research output, or expiration of the employment contract.

c) Exceptions to rules a and b can be made with the permission of the DMPC.

d) Full data comprises the elementary data, meta-information, and all details required to understand the elementary data (i.e., sampling frame, description of data structure, methods used, documentation of the data-generating process, experimental design, questionnaire, etc.).

e) Compliance is controlled by careful review of the various components submitted to the database. The DMR, assisted by the coordination office, determines whether the submitted components are acceptable.

f) All data must be stored in accordance with DFG and university rules.

§8 General provisions

a) The Governing Board of the RTG (which is elected by the members) should ratify this policy.
b) In all cases for which this policy does not give an explicit rule, the DMPC is responsible for a decision.
c) This policy is in full compliance with the DFG and university policies on (linguistic) data management. [1,2]
d) This policy is subject to change, e.g., if DFG or university policies change or if for the General Assembly of the RTG decides to implement changes, as long as 8b remains valid.

---

[1]http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf
[2]https://www.uni-goettingen.de/en/research-data-guideline-of-the-g%C3%B6ttingen-university-incl-umg/509047.html