

## Data Management Plan

The bilateral MEDEA project will create data as test sets for the data models developed in the meetings. This data will be created by human experts in the fields of scholarly editing and economic and social history, who will transcribe from already existing images of original sources. Thus the leading questions, individual problems, and suggestions for solutions must be documented extensively during the data creation process. The project will store these kinds of comments on the transcription process, annotation and structuring of the data together with the created data, preferably in headers and as comments directly in the data.

Data and code created during the project will be managed locally under the responsibility of the individual contributor. These files will contain enough metadata to identify the sources from which the data is drawn (archival references) and the responsibilities of persons creating and modifying the data. The metadata for the test sets will preferably be organised following the guidelines for the TEI header section. During the publication process the project will create a representation of the metadata in the Dublin Core elements set.

Data created in other projects and reused in the project will be managed at the originating institutions. Suggestions for modifications will either be submitted to the originating institutions or added to the documentation of the project with the consent of the originating institution in accordance with their intellectual and other property rights.

The major format to store the data in the project will be XML. CSV may be used for tabular data considered on a case by case basis. RDF files can be stored as N3/Turtle triples. In cases of complex objects integrating multiple files, the project will rely on METS. The complete documentation of the project will be published at the <http://encodinghfrs.org/> site maintained by Kathryn Tomasek and in the Humanities Asset Management System at Graz University (GAMS, <http://gams.uni-graz.at>, see Sustainability Plan). The identifiers created in the GAMS will be used as permanent identifiers.

The publication process will be managed by the principal investigators of the project Kathryn Tomasek, Mark Spoerer, and Georg Vogeler who will evaluate the scholarly quality of the data created in the project and submitted by external parties. For collaborative production of code and data the project will make use of git technologies and use github (<http://github.com>) as a repository where the information will be made public after quality assessment by the Project Directors. The Project Directors will be supported in their assessment of data and reports by the Advisory Board.