

6. Data Management Plan

Responsibilities

Project Directors Caroline T. Schroeder and Heike Behlmer will oversee the data management plan, ensuring that all processes are implemented.

Expected Data, Collection Methods, Data Formats, and Data Dissemination

Digitized Text from Shared Corpora: The digitized text will be raw data in the form of text files, XML files, Microsoft Excel files, PAULA XML files in English, German, Greek Unicode and Coptic Unicode (using the Antinoou font created by the International Association of Coptic Studies.) We anticipate no changes to the Unicode standards for the Coptic character set that would affect our work.

The digitized text files, Excel files, html files, and draft XML files will be stored on version-controlled servers on Göttingen, Münster, and Georgetown Universities using Git or Subversion, as well as the Coptic SCRIPTORIUM GitHub site at <http://www.github.com/CopticScriptorium>. The INTF text files are already stored and versioned using a Git server. Some of the project text data will be drawn from ancient and medieval manuscripts. Under intellectual property law in Germany and the United States, the text from the manuscripts is in the public domain; editorial work can be under copyright. The project will not be publishing online editorial work under existing copyright. Other text data, such as the *TLA* lexicon, will be drawn from digital-native editorial work, for which the *TLA* will give project partners permission to distribute under an open source, Creative Commons attribution license (CC-BY). (See Dr. Frank Feder's letter in Appendix 10.) Prepublication files will be stored on these servers and GitHub. Digital publication of shared data occur on the Göttingen Coptic Old Testament project site, Coptic SCRIPTORIUM's site, and the INTF Virtual Manuscript Room.

Image files from the NT.VMR sample corpus of manuscripts of John's Apocalypse (the book of Revelation) manuscripts are free to view by written permission of the holding institutions and the transcriptions are made available under a Creative Commons license (BY-SA). These image files are available as imaged pages alongside with their electronic transcriptions.

Tools and Technologies: The digital tools to annotate and format the text files will be written in Java, Python, Perl, JavaScript or other scripting languages. The NT.VMR suite of tools and Coptic SCRIPTORIUM tools are already open-source, and the development of future tools during the project period will be open-source, as well. We will distribute the tools as free public downloads under open-source licenses, such as the Apache 2.0 license, or a GNU license. The software will be distributed on GitHub, with links on project partners' sites.

Documentation: The project will provide documentation of the tools and technologies developed on partner project websites. Documentation will be labeled with date and version information and disseminated under open-source licenses, such the Apache license, GNU Free Documentation License, and the Creative Commons Attribution (CC-BY) License.

Detailed documentation on data curation, methods, and standards for digital Coptic Studies will also be distributed in this manner. But in addition, project participants will present papers at the appropriate conferences (such as the International Association of Coptic Studies in 2016 and 2020) and will publish standards in relevant journals.

Period of Data Retention

The project participants embrace the principles of timely, rapid, and open-source data distribution. Tools, methodologies, and documentation will be released as they are created. The tools and standard developed during the project period will be stored on the Göttingen Coptic Old Testament servers and site (due to their long-term funding) and in the Georgetown University Institutional Repository (<http://www.library.georgetown.edu/ir/policies>).

Data Formats and Dissemination

The project participants embrace the principles of timely, rapid, and open-source data distribution. Tools, methodologies, and documentation will be released as they are created. The tools and standard developed during the project period will be stored on the Göttingen Coptic Old Testament servers and site (due to their long-term funding) and in the Georgetown University Institutional Repository (<http://www.library.georgetown.edu/ir/policies>). Georgetown University's Institutional Repository can hold files in .pdf format and datasets. Repository Staff will assist Faculty if the format of the data or publications require conversion.

Project participants will also disseminate their results in journal articles and at professional conferences and symposia. The most important of the latter events are the *Society of Biblical Literature* annual meetings, the quadrennial international congress for the International Association of Coptic Studies, and the annual international Digital Humanities conference.