

Interim Report

PW-51672-14

Coptic SCRIPTORIUM: Digitizing a Corpus for Interdisciplinary Research in Ancient Egyptian

Project Directors:

Caroline T. Schroeder, University of the Pacific

Amir Zeldes, Georgetown University

Institution: University of the Pacific

November 29, 2014

Coptic SCRIPTORIUM has met all the projected milestones for the grant period thus far. This report outlines the activities and accomplishments originally planned for each of the two major phases of work described in the original work plan for the grant. Additional activities are described at the end of the report. In addition, co-Director Dr. Amir Zeldes has taken a position at Georgetown University; the grant activities now take place primarily at the University of the Pacific in Stockton, California, and Georgetown University in Washington, D.C.

Phase 1: May 2014– Fall 2014:

Transcription, digitization, and encoding of expanded corpus by participants (Directors and Editors/Annotators/Digitization Contributors). Corpus: Sayings of the Desert Fathers, Writings of Shenoute

In October and November 2014, we published an expanded corpus on our website, copticSCRIPTORIUM.org, in the PAULA XML, TEI XML and reIANNIS formats. It includes texts digitized by the three Editors/Annotators/Digitization Contributors. The publication includes more Sayings of the Desert Fathers, as well as Coptic New Testament material which is now available; the entire corpus was also edited to account for the changes to the data models and standards discussed in May. More digitized Shenoute documents are in the pipeline for publication this winter or Spring along with other texts.

Another project participant, an M.A. student and graduate of the University of the Pacific, has been updating the TEI XML files for this version of the corpora. His work is funded by an internal University of the Pacific grant, and those files will be published in early December.

Consultation with contractor/consultant about uniform references and data curation

Bridget Almas, senior programmer for the Perseus Digital Library, began developing models for canonical URNs for Coptic literature using CTS and CITE architectures

(<http://www.homermultitext.org/hmt-docs/specifications/ctsum/>,

<http://www.homermultitext.org/hmt-doc/cite/>). We coordinated through Skype meetings during the summer and early fall with project directors.

Workshop at the University of the Pacific (Stockton or San Francisco campus) for all staff followed by additional working session for Project Directors to review and modify digitization processes; finalize decisions on corpus architecture; discuss data curation and models; address linguistic and transcription issues that have arisen; plan for remaining grant period based on the outcomes from the Workshop; review progress of technology for processing the digitized text; discuss and make note of needs and requirements for an online platform for users to add text and/or annotations.

Two of the project's Editors/Annotators/Digitization Contributors participated in a two-day workshop on digitizing the corpus May 15-16, 2014. (The third was on parental leave.) We met one day at the Berkeley Center for Tebtunis Papyri, where we discussed our work with scholars at the University of California working either in Digital Humanities or in Coptic. We met a second day at the University of the Pacific San Francisco campus. Both days we

trained the Contributors/Editors in digitization methods, encoding, and project standards. May 12-14 was the project Directors' working session, where we mapped out changes to the data model and corpus architecture and planned for the digitization of corpora during the grant period.

Phase 2: Fall 2014-early Spring 2015:

Continued transcription and digitization of expanded corpus (Shenoute, Sayings).

Complete formatting of the select biblical text and begin pilot import, tokenizing, and tagging of papyri from Papyri.info.

Editors/Annotators/Digitization Contributors continue to digitize and annotate Biblical texts, Sayings of the Desert Fathers, and Shenoute material. We are on target to have two books of the Bible (the Gospel of Mark and 1 Corinthians) digitized by the end of this phase. The project has also begun work on select Coptic papyri and ostraca from Papyri.info as test cases for digitizing documentary material.

Update metadata based on data curation models and other guidance provided by consultant.

Almas completed her modeling of data curation standards during this phase and has written a white paper to guide our project's implementation of the canonical citation methods. Under her leadership, we consulted with Hugh Cayless (of the Linked Ancient World Data project), and Christopher Blackwell and Neel Smith (both of the Homer Multitext project). We have engaged another contractor who has worked on the Homer Multitext Project in the past, to implement her plan in stages. The implementation will be funded by our NEH ODH grant HD-51907.

Refinement of methods and documentation.

Additional documentation, including a YouTube video (<https://www.youtube.com/watch?v=JZW6oWHbwQU>) describing the data model and searching the corpus, was added to the website (www.copticscriptorium.org). Documentation on the data models and processes for encoding and digitizing have been drafted and will be added to the GitHub repository and website.

Meeting of Advisory Board in Fall 2014

The Advisory Board met virtually in an email consultation about the project in November 2014. Board members indicated that our outreach to other scholars in our subject field and in Digital Humanities was effective and that we are on track with the expansion of our corpora. They also indicated that our data curation efforts are on track. For future activities, one board member recommended providing additional, more detailed documentation for searching the digital corpus. Another recommended additional features to language of origin tagging (providing Greek lemmas for Greek loanwords) and reconsidering decisions about tagging Greco-latin, Greco-hebrew, etc., as Latin and Hebrew instead of Greek.

Additional Activities

In addition to the originally planned goals and milestones, we have several further achievements:

- Another Shenoute scholar has given the project his transcriptions of manuscripts of Shenoute's texts in Microsoft Word format. We will begin converting and processing those texts in 2015.
- We have received permission from the Society of Biblical Literature to publish their XML Greek New Testament in alignment with our Coptic New Testament texts. We will pilot at least one chapter of this alignment in 2015.
- Schroeder has presented conference papers about the project at the North American Patristics Society annual meeting in Chicago (May 2014), DH 2014 in Lausanne (July 2014), and the UCLA-St. Shenouda Society annual Coptic Studies conference (July 2014).
- Zeldes used the corpora to teach a Coptic course at Humboldt University. students used the projects' technologies to prepare Coptic text for publication with Coptic SCRIPTORIUM as their final projects.
- Schroeder and Zeldes have submitted two co-authored articles, for special issues of the *Journal of Digital Humanities* and the *Journal of Digital Scholarship* in the Humanities
- Zeldes has submitted a paper to the North American Conference on Afroasiatic Linguistics in 2015.
- Schroeder met with the staff at the Perseus Digital Library on site at Tufts University in September, 2014, to talk about their workflow and management processes and received some good advice and suggestions.
- Schroeder is attending the Berkeley Prosopography Services Digital Humanities workshop series at the University of California (another NEH-funded project) in Fall 2014.