Interim Report

PW5167214

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

May 31, 2016

This report covers activities from November 29, 2015, through May 31, 2016.

Objectives and Activities

Our Project's objectives by the end of the grant period (extended to August 2016) are to publish a digitized corpus of at least 25,000 tokens of Coptic text, transcribed and encoded according to standards we have been developing and publishing in the documentation for the project, and annotated using the digital tools developed under a simultaneous NEH-ODH grant (HD51907). We have met most of these objectives. We outline below each of our original grant application's objectives, followed by a description of the progress made during this period.

Continued transcription and digitization of expanded corpus (Shenoute, Sayings of the Desert Fathers) and completion of import, tokenizing, and tagging of papyri. Compile and publish online the pilot corpus.

We published an additional annotated fragment of Shenoute's *Acephalous Work 22* in multiple formats (in the ANNIS database, TEI-XML, PAULA XML). Additional sayings from the *Sayings of the Desert Fathers (Apophthegmata Patrum)*, chapters of the Sahidic version of 1 Corinthians, and fragments of Shenoute's work *I See Your Eagerness* have been digitally transcribed and annotated; they will be published this summer. We currently have a published corpus of over 54,000 tokens, machine-annotated with the tools developed in NEH grant HD51907 and then manually corrected by editors.

Discussion and planning regarding creating an online interface for researchers to contribute text and/or annotations to the corpus

Project directors and editors have made a list of the desirable features of such an online database. Implementation of this objective will occur under a bilateral NEH-ODH and DFG grant for the KELLIA project (NEH-DFG HG-229371).

Write and disseminate documentation online.

Using the data curation model for data identifiers developed in consultation with Bridget Almas (Perseus Digital Library), we wrote and disseminated online citation guidelines for researchers using our corpus. We also regularly updated our project wiki to reflect current editorial practices and data models.

Advisory Board meeting to assess the project progress, quality, and standards.

We held an advisory board meeting via Skype in January and have had individual consultations with advisory board members. Board members provided important guidance on annotation practices and citation guidelines.

External evaluation by invited, external users to test use of the corpus. Draft our white paper.

We will conduct these activities this summer.

Consultation with Bridget Almas (Perseus Digital Library) to create models for linked data.

We will consult this summer.

Additional Activities

Project directors Caroline T. Schroeder and Amir Zeldes met with editor Rebecca Krawiec in Washington, DC, (Georgetown University) in December 2015 to review transcriptions, annotations, and editorial practices.

We added to our data model methods for annotating for parallel textual witnesses in more than one manuscripts, as well as metadata to guide users who want to read fragmentary digitized texts in order.

Schroeder met with editors Rebecca Krawiec, Christine Luckritz Marquis, and Dana Robinson in Chicago in May 2016 (coinciding with the North American Patristics Society Meeting).

Activities on Related Grants

Under the NEH-ODH Digital Humanities Start-up grant (HD51907), we continued development of our annotation tools. We have also continued software development on our web interface for reading, accessing, and citing our digital corpus at http://data.copticscriptorium.org. We have an additional corpus of the entire Sahidic New Testament machine-annotated (not manually corrected).

Under the NEH-DFG HG-229371 grant for the bi-lateral KELLIA project, editor and Project Manager Elizabeth Platte digitized and annotated additional literary texts for publication later this year. German partner So Miyagawa (Göttingen University) transcribed, edited, and annotated a fragment of a letter of monastic author Besa, which we published this spring in our database, TEI-XML, and PAULA XML.

Conference Presentations and Papers

Almas, Bridget, and Caroline T. Schroeder. "Applying the Canonical Text Services Model to the Coptic SCRIPTORIUM." Under review.

Schroeder, Caroline T. and Amir Zeldes. "Raiders of the Lost Corpus", *Digital Humanities Quarterly*. At press.

Schroeder, Caroline T. "The Dark Side of Digitization: Ideology and Power in the Technological Utopianism of the "Archive." Paper presented at the North American Patristics Society Annual Meeting, Chicago, IL. May 27, 2016.