Reviews in Digital Humanities • Vol. 2, No. 1

Review: Scribes of the Cairo Geniza

Caitlin Haynes¹

¹Smithsonian Transcription Center

Published on: Jan 04, 2021

DOI: 10.21428/3e88f64f.52b54957

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Project

Scribes of the Cairo Geniza

Project Team

[Laurie Allen](#), Library of Congress (formerly Assistant Director for Digital Scholarship, University of Pennsylvania)

[Samatha Blickhan](#), Zooniverse Humanities Lead, Adler Planetarium

[Laura Newman Eckstein](#), University of Pennsylvania (formerly Judaica Digital Humanities Coordinator, University of Pennsylvania)

[Emily Esten](#), Judaica Digital Humanities Project Coordinator, University of Pennsylvania

[Arthur Kiron](#), Schottenstein-Jesselson Curator of Judaica Collections, University of Pennsylvania

[Marina Rustow](#), Khedouri A. Zilkha Professor of Near East, Professor of Near Eastern Studies and History, Princeton University

Project URL

<https://www.scribesofthecairogeniza.org/>

Project Reviewer

[Caitlin Haynes](#), Smithsonian Transcription Center

Project Overview**Emily Esten**

Scribes of the Cairo Geniza engages people from across the globe in the sorting and transcribing of multilingual medieval and pre-modern manuscript fragments from the Cairo Geniza. This corpus of more than 350,000 documents, the majority of which date from the tenth to thirteenth centuries CE, accumulated over the course of a thousand years in a storeroom (or “geniza”) of the Ben Ezra synagogue in Fustat (Old Cairo). Geniza fragments, exhumed for study at the end of the 19th century, serve as a time capsule of Jewish history throughout the Mediterranean world during periods when ninety percent of Jews lived under Islamic rule. Though originally stored in one place, these hundreds of thousands of fragments are now dispersed among dozens of collections around the world, including the seven libraries directly involved in this project.

Launched in 2017, Scribes of the Cairo Geniza has three primary goals: 1) provide our community of citizen scientists opportunities to view and decipher Cairo Geniza fragments; 2) contribute to the classification of fragments by script-type and content; 3) produce classifications and transcriptions of the material to be available as open-source datasets for historians, linguists, and other scholars to reuse, republish, and communicate research findings back to the crowdsourcing community. It is one of the first custom projects on [Zooniverse](#), the world's largest platform for online crowdsourced research, to address right-to-left languages, incorporating Hebrew and Arabic translations within the project interface and data output. However, no language proficiency in either Hebrew or Arabic is required to participate.

Scribes of the Cairo Geniza is an international partnership headed by the University of Pennsylvania Libraries and Zooniverse. It has been supported by a National Leadership Grant from the Institute of Museum and Library Services. Zooniverse led the technical development of the project, designing, implementing, and testing functionality for the custom transcription interface and classification workflows. Penn Libraries provides overarching project management, moderates day-to-day engagement on the site, and exports and processes the crowdsourced data for regular review by scholars and for future public use. Researchers from the Princeton Geniza Lab at Princeton University and the e-Lijah Lab and the Centre for Interdisciplinary Research of the Cairo Genizah at the University of Haifa contributed to the content development and translations for the project and continue to engage communities while reviewing crowdsourced data for accuracy.

With the crowdsourced classification and transcription data produced through this project, Scribes of the Cairo Geniza is helping to rewrite the history of the pre-modern and medieval Middle East, the Mediterranean and Indian Ocean trade, and the Jewish diaspora within an Islamic environment. Among its successes, the project has engaged over 9,500 users, who have made almost 300,000 classifications and transcriptions. Early results of the sorting workflow are already available as [open data](#) via University of Pennsylvania's open-access repository. Researchers have already incorporated early transcription data outputs into ongoing work to map the entire corpus of non-literary Geniza documents and apply text recognition software to medieval Hebrew-character handwriting.

Project Review

Caitlin Haynes

Scholars of pre-modern Mediterranean history are undoubtedly familiar with the historical texts of the Cairo Geniza, a collection of over 300,000 primary source textual fragments documenting the Jewish and Islamic world in the 10th-13th centuries CE. Originally discovered in the attic of the Ben Ezra synagogue in Fustat, these rich materials are now held in repositories around the world. Dispersed, damaged, and written in a variety of languages and scripts, the texts have long been challenging for researchers to easily access and analyze. The Scribes of Cairo Geniza crowdsourcing project and their team of “volunteer humanists and historians,” however, are changing that.

A collaboration between the University of Pennsylvania Libraries, the Princeton Geniza Lab, the e-Lijah Lab and the Centre for Interdisciplinary Research of the Cairo Genizah at the University of Haifa, the Library of the Jewish Theological Seminary, the Genizah Research Unit at Cambridge University Library, the University of Manchester Library, and the Bodleian Libraries at University of Oxford, the Geniza project was launched on Zooniverse in 2017. The project seeks to make the Geniza documents — of which close to 85,000 have been cataloged and digitized at their respective libraries and archives — text-searchable and indexed through public transcription and identification. Zooniverse volunteers working on the Scribes of Cairo Geniza are invited to transcribe, classify (including searching for identifiable phrases), and sort “difficult” and “easy” Arabic and Hebrew fragments, deciphering and organizing the texts into manageable categories for future project phases. Content on the site is available in English, Hebrew, and Arabic.

As the world’s largest “people-powered” crowdsourcing site, Zooniverse provides a stable and trusted platform for this large-scale, cooperative project, and the Geniza’s team has done an impressive job leveraging its features and technical functionality to improve access to the Cairo Geniza texts. The strength of this project lies in its multi-phased, step-by-step process that allows users at all skill levels to meaningfully contribute. Clear instructions, tutorials, and workflows — developed through Zooniverse’s standardized and adaptable [Project Builder](#) — guide volunteers through discrete tasks that simplify the challenging aspects of the content and present the work in a fun, engaging fashion. New and returning users can easily interact with the Geniza team, partnering organizations, and each other through Zooniverse’s “Talk” message board. Beyond this, users can explore the history of the project and the impact of transcription and classification via the project’s social media platforms, newsletters, and in-person and virtual events with partner institutions.

Perhaps most impressive are the custom-created “keyboards” for Hebrew and Arabic letters identified in the Geniza fragments developed by the Geniza project team and Zooniverse staff, which make tasks even less daunting and ensure further accuracy in transcriptions. Users simply match characters in the original fragments with those in the related project keyboards. The creation of this linguistic tool, as well as the Geniza project team’s active work to openly share completed transcriptions and data, provides a wealth of resources for digital humanities and historical scholars alike.

There is still much to be learned about the research impact of the Scribes of Cairo Geniza project — especially as additional phases of work and digitized texts continue to be launched on the site. Three years in, however, it is clear that completed work is already enhancing historical understandings of the pre-modern Islamic world and the Jewish diaspora. Over 9,500 volunteers have contributed to project tasks, and together more than 280,000 sorting classifications of the fragments have been completed. According to recent posts on the [Judaica Digital Humanities blog](#), researchers in Israel, for instance, are currently using the first completed Geniza project data set of over 40,000 fragments to create a technical script that compares the transcriptions to texts available on [Sefaria: A Living Library of Jewish Texts](#).

With a proven scholarly impact, successful technical development and maintenance, and support and participation from worldwide institutions and volunteers, the Scribes of Cairo Geniza project is an impressive example of how digital crowdsourcing projects can provide effective and engaging solutions to the issue of disparate and inaccessible archival collections.