

5-9-2019

Of Scribes and Scripts: Citizen Science and the Cairo Genizah

Laura Newman Eckstein

University of Pennsylvania, lauraeck@upenn.edu

Of Scribes and Scripts: Citizen Science and the Cairo Genizah

Abstract

In August 2017, the University of Pennsylvania Libraries, in partnership with the Princeton Geniza Project, the Library of the Jewish Theological Seminary, the Genizah Research Unit at Cambridge University, and the Zooniverse, began the first phase of a larger project to attempt to sort and transcribe Cairo Geniza fragments, entitled “Scribes of the Cairo Geniza.” This article describes the first phase of the projects and its results.

Keywords

Cairo Genizah, paleography, Jewish manuscript studies, digital humanities, crowd-sourcing

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 3, NUMBER 1

(Spring 2018)

Manuscript Studies (ISSN 2381-5329) is published semiannually
by the University of Pennsylvania Press



The Schoenberg Institute
for Manuscript Studies

UNIVERSITY of PENNSYLVANIA LIBRARIES

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 3, NUMBER 1

Special Issue:

The Syriac Galen Palimpsest

Editors: William Noel and Ralph M. Rosen

Articles

- The Syriac Galen Palimpsest Project: An Introduction
WILLIAM NOEL AND RALPH M. ROSEN 1
- Pulling It All Together: Managing the Syriac Galen
Palimpsest Project
MICHAEL B. TOTH 9
- The Codicology and Conservation of the Syriac
Galen Palimpsest
ABIGAIL B. QUANDT AND RENÉE C. WOLCOTT 33
- Spectral Imaging Methods Applied to the Syriac
Galen Palimpsest
ROGER L. EASTON, JR., KEITH T. KNOX, WILLIAM A. CHRISTENS-BARRY,
AND KEN BOYDSTON 69
- The Galen Palimpsest and the Modest Ambitions of the
Digital Data Set
DOUG EMERY 83
- The Syriac Galen Palimpsest: A Tale of Two Texts
NAIMA AFIF, SIAM BHAYRO, GRIGORY KESSEL, PETER E. PORMANN,
WILLIAM I. SELLERS, AND NATALIA SMELOVA 110

Analyzing Images, Editing Texts: The Manchester Project
NAIMA AFIF, SIAM BHAYRO, PETER E. PORMANN, WILLIAM I. SELLERS,
AND NATALIA SMELOVA 155

The Textual Interest of the Syriac Versions of Galen's *Simples*
IRENE CALÀ, JIMMY DACCACHE, AND ROBERT HAWLEY 186

Annotations

Of Scribes and Scripts: Citizen Science and the Cairo Geniza
LAURA NEWMAN ECKSTEIN 208

Preserving Endangered Archives in Jerba, Tunisia:
The al-Bāsī Family Library Pilot Project
ALI BOUJDIDI AND PAUL M. LOVE 215

The Intricacies of Capturing the Holdings of a Mosque
Library in Yemen: The Library of the Shrine of Imām al-
Hādī, Ṣa' da
SABINE SCHMIDTKE 220

Compilation, Collation and Correction in the Time
of Encyclopedism: The Case of UPenn LJS 55
NATHALIE LACARRIÈRE 238

Mapping Manuscript Migrations: Digging into Data for the
History and Provenance of Medieval and Renaissance
Manuscripts
TOBY BURROWS, EERO HYVÖNEN, LYNN RANSOM, AND HANNO WIJSMAN 249

Reviews

Aḥmad 'Abd al-Bāsiṭ. *Catalogue of the Private Collections of
Manuscripts in the Egyptian National Library*
ELIAS G. SABA 253

David T. Gura. *A Descriptive Catalogue of the Medieval and Renaissance Manuscripts of the University of Notre Dame and Saint Mary's College*

LISA FAGIN DAVIS

256

Christopher De Hamel. *Meetings with Remarkable Manuscripts*

DANIEL TRAISTER

260

Of Scribes and Scripts

Citizen Science and the Cairo Geniza

LAURA NEWMAN ECKSTEIN
University of Pennsylvania

THE ATTIC OF BEN Ezra Synagogue, in what is today known as Old Cairo, once held hundreds of thousands of manuscripts: the corpus of the Cairo Geniza. Traditionally, the term *Geniza* is understood as a *temporary* storage chamber for unusable Hebrew religious texts. For unknown reasons, however, the Jews of Fustat never removed their texts from the Geniza. Not only did they deposit religious texts in the Geniza, they also left their letters, court records, receipts, and other conventional texts that tell us about daily life during the tenth to the thirteenth centuries. In the late nineteenth century, the Geniza was emptied and the fragments were dispersed to various scholarly institutions. Yet, over a century later, less than half of the approximately 350,000 fragments have been cataloged, and those that have been cataloged have varying levels of detail.¹ This is largely due to the complexity of the task and the particular skill sets required—many languages, hands, and scripts need to be deciphered by just a handful of experts.

In August 2017, the University of Pennsylvania Libraries, in partnership with the Princeton Geniza Project, the Library of the Jewish Theological Seminary, the Genizah Research Unit at Cambridge University, and the

1 “About Scribes of The Cairo Geniza,” Zooniverse.org, <https://www.zooniverse.org/projects/judaicadh/scribes-of-the-cairo-geniza/about/research>, accessed 27 September 2017.

Zooniverse, began the first phase of a larger project to attempt to sort and transcribe Cairo Geniza fragments, entitled “Scribes of the Cairo Geniza.” The Zooniverse is the world’s largest and most popular platform for people-powered research. Hundreds of thousands of people around the world come together on this platform to assist professional researchers. Their goal is to enable research that would not be possible, or practical, otherwise.² Instead of relying on the expertise of a few scholars, we will be relying on the interest of the global population, and the remarkable potential of crowdsourcing technologies. We provide basic guides that give citizens the information they need to make useful distinctions that can contribute to deciphering the Geniza.

In the first phase of the project, we are asking citizen-scientists to sort Cairo Geniza fragments in preparation for our later transcription phase. By asking a series of tailored questions, we will be able to gather more metadata about each individual fragment. To devise a successful workflow, in which citizens are asked questions that generate useful data, and in ways that engage the widest possible audience, we worked from the following principles:

1. No expert knowledge is required for participation.
2. Reward for work is almost immediate.
3. The work undertaken adds to the sum of knowledge about the Geniza.

The third principle is particularly important. Mia Ridge, in her volume *Crowdsourcing our Cultural Heritage*, writes, “Think of crowdsourcing in cultural heritage as a coalescence around a set of principles, particularly the value placed on meaningful participation and contributions by the public.”³ Here are two examples of good questions we decided not to ask because they would not result in meaningful data that we did not intend to acquire through other means.

2 “About the Zooniverse,” Zooniverse.org, <https://www.zooniverse.org/about>, accessed 27 September 2017.

3 Mia Ridge, *Crowdsourcing Our Cultural Heritage* (London: Routledge, 2017), 4.

- *Do you see the word “W”?* While this question can be extremely helpful in determining the genre of a particular fragment, individual words will be deciphered in the later transcription phase. There is no need to ask citizen-scientists to duplicate later work. If it seemed that we could answer a question through the transcription process, or would be too confusing for citizen-scientists and hence result in data that would not be helpful for sorting purposes, the question was nixed from the workflow.
- *Is there a drawing on the fragment?* This category encompasses such phenomena as embellished letters, doodles, and ink smudges, and thus might result in data that is too diverse to be useful. (Drawings undoubtedly remain interesting, and users can record observing them using hashtags on talk boards, as discussed below).

The questions we decided to pose are as follows:

- *In what script is this text written?* This will tell us whether the fragment is in Hebrew, Arabic, or both Hebrew and Arabic scripts. We provide users with easy tutorials that allow them to recognize these scripts by sight. We need this information for sorting before we begin Phase II: we want those with language skills to be able to transcribe difficult texts in familiar scripts in Phase II (figure 1).
- *Is the script written in a formal or informal style?* Again, we provide visual guides for users in this task. This information tells us whether the fragment will be easier or more difficult to transcribe. Given the fact that formal scripts are more uniform in their lettering styles, with fairly even lines in comparison, they should be much simpler to transcribe than informal scripts (figure 2).
- *Identify (using the point tool) a few different visual characteristics depending on whether the fragment is in Hebrew, Arabic, or both types of script.*
 - Diagonal text and/or perpendicular text in the margin of fragments in Hebrew script tells us that it is a documentary fragment rather than a religious one, probably a letter.
 - A horizontal line above a word in a Hebrew script tells us that the fragment contains a literary text.

- Colons indicate the end of a verse or a line of a literary text.
- Seals, frequently found on Arabic fragments, indicate a state document (figure 3).
- In addition to answering these questions, the citizen-scientist may also participate through the talk boards. When a user has answered all the questions for a given fragment, they can click either “done” or “done & talk.” The latter brings users to a window with a picture of the fragment they just worked on, together with a comment box in which they can write. They can add further observations about their fragment in this comment box; they are encouraged to use hashtags, which are searchable. Some of these hashtags we prepared in advance, such as “#redink” or “#drawings,” knowing that they have the potential to be useful. The tags are also useful if you have a question. All you do is use the tag #askanexpert. Each day one person from our content team, an expert in the Geniza, is assigned to the talk boards (see figure 6). Below I’ve included a screenshot of all of the talk boards as well as an example talk discussion (screennames omitted for privacy) to show the type and wealth of interactions that the talk board facilitates.

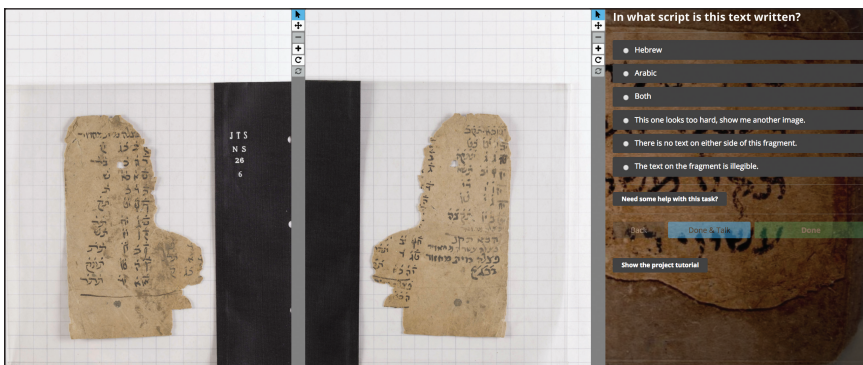


FIGURE 1. Questions distinguishing Hebrew and Arabic script types. Digital image. Scribes of the Cairo Geniza, 8 August 2017, <https://www.zooniverse.org/projects/judaicadh/scribes-of-the-cairo-geniza/classify>, accessed 27 September 2017.

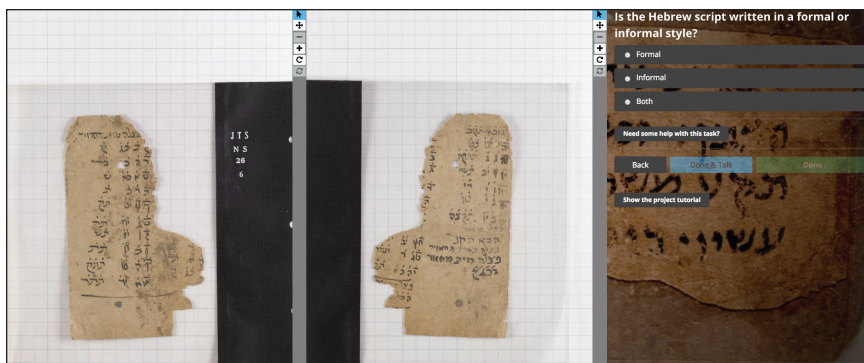


FIGURE 2. Question distinguishing formal and informal script types. Digital image. Scribes of the Cairo Geniza, 8 August 2017, <https://www.zooniverse.org/projects/judaicadh/scribes-of-the-cairo-geniza/classify>, accessed 27 September 2017.

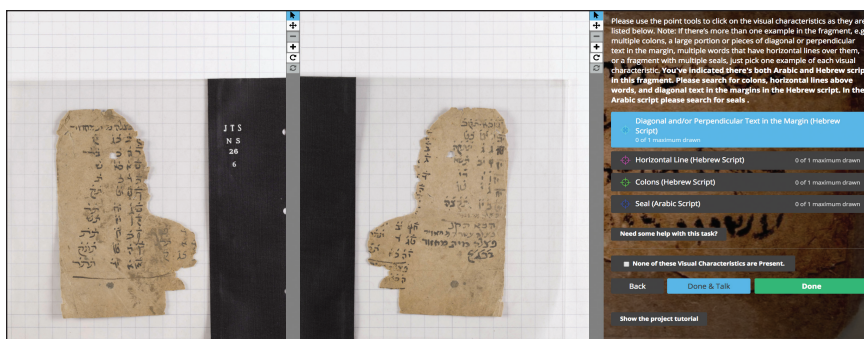


FIGURE 3. Question asking about the visual characteristics of a Cairo Geniza fragment using the point tool. Digital image. Scribes of the Cairo Geniza, 8 August 2017, <https://www.zooniverse.org/projects/judaicadh/scribes-of-the-cairo-geniza/classify>, accessed 27 September 2017.

Talk Board Example (see figures 4 and 5)

Person 1: #formalscript Practically intact. Some text seems to have been put in a box on the back side of the fragment.

Person 2: This may be a nineteenth-century piece. The Jewish community of Cairo was depositing things into the Geniza chamber practically until the day it was emptied in 1897.

Person 3: What language is this though?

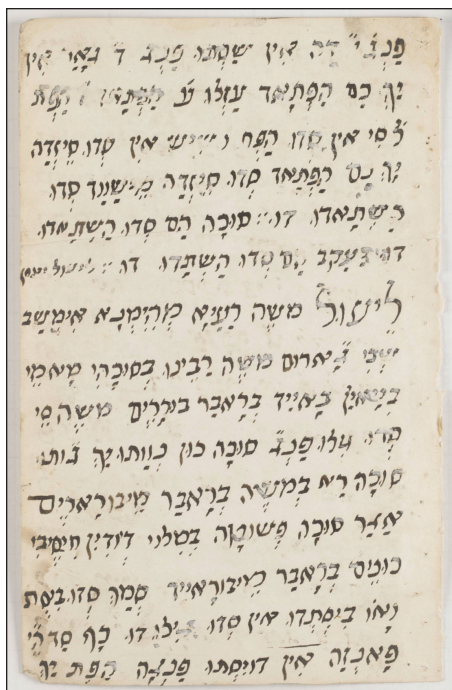


FIGURE 4. ENA 2329.3, recto. Courtesy of the Jewish Theological Seminary.

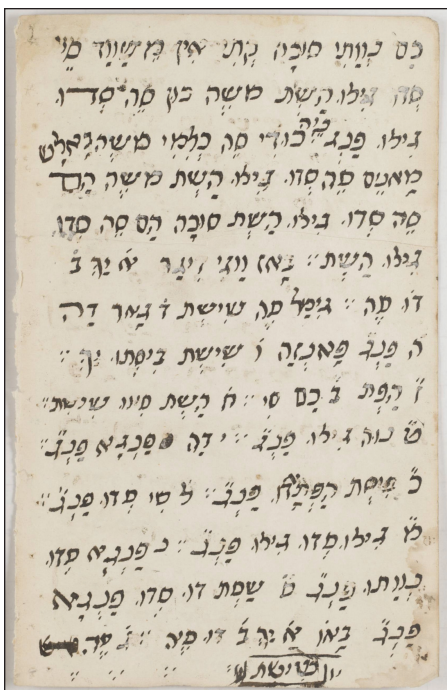


FIGURE 5. ENA 2329.3, verso. Courtesy of the Jewish Theological Seminary.

Person 4: Hebrew.

Person 5: Wait how do you know the language is Hebrew [*sic*]?

Because the script is definitely Hebrew [*sic*] . . . but it is difficult to determine which language unless you are an expert, which you very well could be. If so, we would love to add you to our list of moderators.

Person 4: I understood the Hebrew/Arabic question that we are answering for each document to be about the letters used. They might be transliterating Arabic words into Hebrew letters, but that is not what I thought we were supposed to be classifying. It is true that if you read the words, they don't seem to be Hebrew words at all and are more Arabic sounding than Hebrew, but I can't recognize them as Arabic words either except for maybe 'kun'. My Arabic is not that strong though, and is not Egyptian.

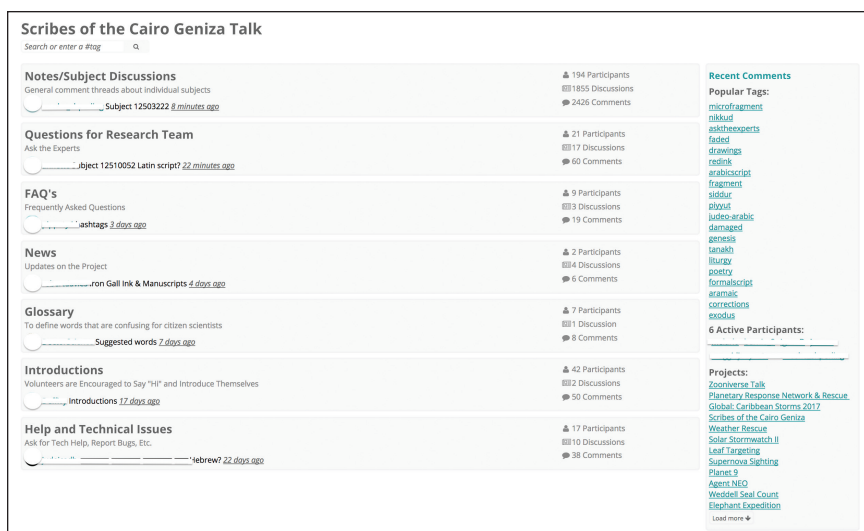


FIGURE 6. Scribes of the Cairo Geniza talk boards. Digital image. Scribes of the Cairo Geniza, 8 August 2017, <https://www.zooniverse.org/projects/judaicadh/scribes-of-the-cairo-geniza/talk>, accessed 27 September 2017.

Person 2: The script is Hebrew but the language isn't. Nor is it Arabic. I have a theory about what it is but I'm going to confirm with a colleague before sharing. Stay tuned!

Person 6: The language is Judeo-Persian and the content seems to be related to the kabbalistic [*sic*] custom of inviting Key Biblical figures to be hosted in the Sukkah.

Who knew that fragments from an attic could offer so much potential? These Cairo Geniza documents have the potential to rewrite the history of the premodern Middle East, of Mediterranean and Indian Ocean trade, and the Jewish diaspora. If citizen-scientists can successfully sort and transcribe these Geniza fragments, this will change Geniza scholarship forever. But the success of this project would have implications far beyond Geniza studies, and transform the way in which the decipherment of a large corpus of difficult texts is undertaken the world over.