JS/DH: Books and Computing

Michelle Margolis

Columbia University Libraries

As co-director of a digital project focusing on the history of the book,[1] the area of digital methods employed in the study of Jewish book is of particular interest to me. Libraries have been at the forefront of digital humanities undertakings, largely because metadata is a fundamental part of library work, and bookish projects often align with computational methods. Considering that many book projects are not quite completed, this column describes in-process projects as they are, and gestures toward possible enhancements or future work that may be useful. To be clear, no textual analysis projects are examined, but rather undertakings that focus on books and their contents as objects and data, using digital tools such as optical character recognition (OCR) for typewritten texts, textual markup to add metadata to various images, data visualization, and mapping.

One project that is still in its infancy is the MIDRASH project: Migrations of Textual and Scribal Traditions via Large-Scale Computational Analysis of Medieval Manuscripts in Hebrew Script.[2] The principal investigators in the MIDRASH project have been working to perfect an OCR model for all kinds of Hebrew manuscripts. This has obvious ramifications for those studying manuscripts writ large, as a working OCR system for handwritten texts would significantly enhance the research of Hebrew manuscripts and make it more efficient. It also will be interesting to see whether the paleographical structures conceived in the last few decades by scholars such as Malachi Beit-Arié, Colette Sirat, or Judith Olszowy-Schlanger will hold firm or require adjusting based on computational analysis of the various scripts.

MIDRASH will be working with the Friedberg Geniza Project (FGP)[3] and KTIV: The International Database of Hebrew Manuscripts,[4] both of which provide access to significant corpora of

---

1. See "Footprints: Jewish Books Through Time and Place," footprints.ctl.columbia.edu, accessed November 3, 2024.

2. For more information, see "AI and Hebrew Manuscripts," https://www.biu.ac.il/en/article/11914 (archived December 2022), accessed August 25, 2024. The MIDRASH project does not yet have a standalone website, although one is under development (personal communication, Daniel Stoekl ben Ezra, August 22, 2024).

3. "The Friedberg Geniza Project," http://pr.genizah.org, accessed November 3, 2024.

4. "Ktiv: The International Collection of Digitized Hebrew Manuscripts," https://www.nli.org.il/en/discover/manuscripts/hebrew-manuscripts, accessed November 3, 2024.

manuscripts spanning the globe (both in their place of creation and current locations) as well as millennia. In more recent months, MIDRASH has also ventured into analysis of printed texts.[5]

Another project relating to Jewish book history, by Luigi Bambaci, illustrates how the process of encoding a critical edition of the Hebrew Bible (Benjamin Kennicott's 1778 edition) is exercised to "reverse engineer" the book and determine which text variants came from which manuscripts.[6] Once the manuscripts and variants were encoded, the team could look for similarities among manuscript traditions or trace the way Hebrew manuscripts were accessed from various locations. At present, Bambaci's data files for the Five Scrolls are available on Github for use and reuse in various formats, such as TEI, XML, and HTML.[7] A future expansion of Bambaci's work to the entire Hebrew Bible, and its availability to all, would likely be a game-changer for the critical study of this text.

Another markup project of a digitized published text is Yael Netzer and Eliezer Baumgarten's endeavor to map connections within *Shem ha-gedolim* (first edition Livorno, 1774),[8] the bio-bibliographical work of Hayyim Joseph David Azulai (ha-Ḥida; 1724–1806). Netzer and Baumgarten identified individual manuscripts and collections listed by Azulai, and linked authors and their connections throughout the corpus. This project is not available online at present.[9] An accessible version would be a useful tool to those studying connections among historic Jewish texts and the people who created and published them, as well as to those surveying Azulai's perspectives and biases in putting together this bibliographical work. Netzer has continued to apply this method transforming another bibliographical reference work into a computational format, for example, in her work with Amalia Levi on the *Quellen zur Geschichte der Juden in den Archiven der neuen Bundesländer* (Sources for Jewish History in Archives in the New Federal States of Germany).[10]

---

5. See, for example, the following invitation to a printed texts hackathon: "Week of the Digital Hebrew Book 2024," https://docs.google.com/forms/d/e/1FAIpQLSdHXAjaVENWnU1AcFrU7kd_D69ijQrsYc6_3ojGd5YuUZEaow/viewform, accessed August 22, 2024.

6. Luigi Bambaci, "Digitizing Kennicott's Collation of the Hebrew Bible: Experiences of Encoding and of Computer-assisted Stemmatic Analysis," in *Jewish Studies in the Digital Age*, ed. Gerben Zaagsma, Daniel Stökl Ben Ezra, Miriam Rürup, Michelle Margolis, and Amalia S. Levi, (De Gruyter Oldenbourg, 2022), 299–334. https://doi.org/10.1515/9783110744828-014.

7. "Luigi Bambaci/Kennicot," https://github.com/LuigiBambaci/Kennicott, accessed November 3, 2024.

8. "Shem ha-gedolim: ...," https://babel.hathitrust.org/cgi/pt?id=hvd.hn91fd&seq=1, accessed November 3, 2024.

9. A recording of Netzer's presentation on this project (2021) is available here: "Shem ha-Gdolim: Mapping Hebrew Bibliography," https://www.morressier.com/o/event/5fd2237e54bbb7f516f76f1b/article/5fd8d8c83d762219be-34f4ed, accessed August 22, 2024.

10. Yael Netzer and Amalia S. Levi, "Legacy Catalogs as Data. The case of Quellen zur Geschichte der Juden in den Archiven der neuen Bundesländer," *Medaon* 17 (2023): 32, https://doi.org/10.17613/32hc-4n62. See also "Quellen zur Geschichte der Juden in den Archiven der Neuen Bundeslander," https://undusting.org/quellen/s/jewish_sources/page/home, accessed October 28, 2024.

Jewish book historians are fortunate to have access to the Bibliography of the Hebrew Book (BHB),[11] a database of nearly all books printed in Hebrew prior to 1960. Sinai Rusenik took the data in the bibliography a step further in an interactive map of the BHB,[12] allowing users to visualize Jewish book production since the premodern era. Users can limit the data by date and language (both are color-coded) to show the multifariousness of Jewish book history.

Wikidata has also proven to be an important resource in Jewish studies projects that trace names of people. The National Library of Israel regularly creates and updates records in Wikidata, as do various scholars working with digital materials relating to Jewish studies. The platform's flexibility and openness allow authority records from library catalogs and other standard databases to be connected with individual projects related to Jewish books, authors, and other actors within the book history ecosystem. In but one example, the entry for Hayyim Joseph David Azulai has no less than 37 identifiers linked from various systems.[13] Further use and engagement with Wikidata sources will serve to link additional data and thus tighten the connections among bibliographical projects.

Linked data is an area that has a great potential for expansion within the world of Jewish books as well. It has been discussed in this journal in the past, but there is still a great deal of work to be done to utilize its full potential.[14] Looking forward, we can expect that projects like MIDRASH and open knowledge bases such as Wikidata, among some of the others listed above, will continue to grow and share information about the Jewish book in many different ways.

11. "Bibliography of the Hebrew Book," https://merhav.nli.org.il/primo-explore/search?vid=MBI&lang=en_US, accessed November 3, 2024.

12. "Bibliography of the Hebrew Book: map and visualisation," https://lookerstudio.google.com/reporting/73a17d-cc-f092-4dec-8c57-58b606862b2d, accessed August 25, 2024 (note that it sometimes takes a few minutes to load).

13. "Chaim Joseph David Azulai," https://www.wikidata.org/wiki/Q1058991, accessed August 15, 2023,.

14. Dov Winer, "Judaica Europeana: An Infrastructure for Aggregating Jewish Content," *Judaica Librarianship* 18 (2014): 88–115, doi.org/10.14263/2330-2976.1027. Winer demonstrated the significance of linked data with relevant examples for Jewish Studies. Sadly, the Judaica Europeana project no longer exists as such and its URL no longer works.