https://doi.org/10.2352/issn.2168-3204.2024.21.1.7

Digitizing the Library of Congress Hebrew Manuscripts Collection

Ariel Segal; Library of Congress; Washington, D.C./United States

Hana Beckerle; Library of Congress; Washington, D.C./United States

Abstract

This paper will present an overview of a project to digitize the Library of Congress Hebrew Manuscripts collection, which spanned from 2021 through spring 2023. It will describe the historical/cultural importance and breadth of the collection, as well as the workflow and processes used to digitize and display the manuscripts.

Introduction

This paper will present an overview of the Hebrew Manuscripts digitization project at the Library of Congress, including motivations, approach, challenges encountered, imaging techniques, results, and conclusions. It will discuss 3 areas in particular: 1) The unique content of the manuscript collection; 2) The workflow which the Digital Scan Center developed to ensure accurate and timely digitization of the collection, and 3) The innovative "Golden Convert" process which enabled the manuscripts to display with color and texture closer to their physical original than previously possible, as shown by analysis from target data.

Project Motivations

The Library of Congress's Hebraic Section, part of the African and Middle East Division (AMED), possesses 226 manuscripts in Hebrew and in similar languages such as Judeo-Arabic, Judeo-Persian, and Yiddish. The manuscripts date from the 11th to early 20th centuries. Most of the manuscripts were donated to the Library in the 1910s by Jacob Schiff, acquired along with extensive printed Hebraica from the book collector and polymath Ephraim Deinard. However, the Library has continued to acquire manuscripts (MSS), and some of the most remarkable ones came into the Library's possession in the last 15 years.

The collection contains items documenting almost every aspect of Jewish life and culture. Some unique items in the collection include MS 181, the famous Washington Haggadah, an illuminated treasure from the late 15th century. The Haggadah has striking illustrations depicting the Passover service as well as charming details of daily life, such as women preparing the Passover feast with a dog at their feet (See Figure 1). MS 229, "Order of Prayers before Retiring at Night", is a mid-18th century illustrated collection of common prayers, purchased by a man for his wife (See Figure 2). The Hebrew prayers are in a square Hebrew script, while the Yiddish instructions are in a font often used for books meant for women, "vayber-taytsch" (translation for women).

Figure 1a. *Leaf from MS 181, Washington Haggadah, depicting cleaning out leaven;* ***Figure 1b.*** *Leaf from same MS, depicting preparation of Passover feast by women with beggar and dog [1]*

Figure 2. *Folio from MS 229 "Order of Prayers before Retiring at Night" [2]*

Other manuscripts of note include MS 157, "Fragment from a collection of responsa by [Rabbi] Solomon ibn Adret," known as the "Rashba," who is considered one of the most prominent authorities on Jewish law of all time. The Library has discovered that the watermark on this paper makes this manuscript the oldest example of Fabriano paper in its collections. MSS 77 A&B, "Sefer Eitan Ha Ezrahi", is an unpublished novel in Hebrew based upon the then-recently concluded first Zionist Congress in 1896. The author,

Samuel b. Hayyim Acatriel ha-Levi Levine, was an ancestor of the musician Jorma Kaukonen, who played for Jefferson Airplane. These manuscripts were extremely meaningful to Ariel Segal, one of the authors, given their rich Judaic content; the two most recent manuscripts were authored by one of his friends' ancestors. [3]

These manuscripts had been digitized at relatively low resolution by a Library of Congress Junior Fellow in 2015 and some had been previously published on the Library's website. The Library has since acquired updated, state-of-the-art digitization equipment as part of the relocation and renovation of its Digital Scan Center (DSC) in 2020-2021, and new techniques for color management were now available. The Library therefore decided to digitize the manuscripts at higher resolution and release them to the public in a digital form much closer to the actual physical objects in terms of both resolution and color.

Approach

The Library partnered with the David Berg Foundation, which provided funding for this effort. The Foundation seeks to "provide greater access to cultural treasures" [4] such as the Library's Hebrew Manuscripts. The agreement between the Berg Foundation and the Library specified that the digitized manuscripts needed to be openly available online by May 2023 for Jewish American History Month (or before) and laid out reporting requirements for Berg Foundation funding. The funding provided by the Berg Foundation covered staff time and general project resources, such as conservation supplies. The funding was not allocated for specific digitization equipment.

This project followed the Library's special collections digital access workflow. A Hebraic Section librarian developed a proposal for the Digital Library Content Group (DLCG) in partnership with the Berg Foundation. Digitization project stakeholders, including representatives from the Digitization Services Section (now Collections Digitization Division), Conservation Division (CD), Office of General Counsel (OGC), and others conducted a feasibility assessment to discuss the project approach and level of effort for each project phase. This was deemed a high-priority project given the cultural and scholarly importance of the collection and the partnership with an external organization.

Once the project was initiated, the DSC worked closely with the Hebraic Section and Conservation Division as manuscripts were examined, treated as necessary, and digitized. The DSC designated an internal project manager to track progress and monitor all items as they moved through a multi-step digitization workflow (see Imaging Techniques for more detail), who also served as a liaison to stakeholders in other areas of the Library. For the first year of the project, only one digital conversion technician and the project manager worked on imaging and processing the manuscripts. At the end of that year, two digital imaging specialists were hired by the DSC. Among other projects, they helped introduce efficiencies to the workflow, corrected issues with the imaging cameras, and imaged the manuscripts themselves. Several other digital conversion technicians and specialists began to work on the project, which substantially increased progress toward completion.

DSC equipment is set up and calibrated to meet Federal Agencies Digital Guidelines Initiative (FADGI) 3-star levels of image quality, at a minimum [5]. Target images from this and other projects were checked regularly to ensure that all images produced in the Scan Center met the Library's image quality standards. Digitization project team members also conducted a visual inspection quality review of their images, checking for focus, skew, completeness, order of the images, and more. Some of the manuscripts do not have pagination, requiring staff to review them page-by-page in comparison to the images to ensure that no leaves were missing from the digital surrogate. Additionally, since the manuscripts are in Hebrew, the image order (right to left, rather than left to right), was checked. A Hebraic Section Specialist performed a final quality review on all images using the Library's Sampler review platform.

Once images were accepted, they were ingested into the Library's Content Transfer Services (CTS) long-term storage repository and prepared for public access. The Library's metadata and web teams performed necessary catalog records and metadata updates, and created a new collection framework and contextual materials for the online display. This is consistent with the Library's current Special Collections web access workflows.

Challenges

Digitization began in December 2021 and proceeded efficiently at first. However, the head of the Digitization Services Section determined that substantive construction was needed to alleviate temperature issues in the DSC following its relocation and renovation. This construction began in spring 2022 and lasted over a year, which considerably reduced the number of imaging devices and space available to project team members.

The workflow was further interrupted in spring and summer 2022 by two other high-profile, short deadline projects that were also planned in agreement with external organizations. This meant time and access to capture devices for digitization was greatly reduced. However, the deadline for completing the work (stipulated in the written agreement with the Berg Foundation) was not adjusted when these interruptions occurred. One advantage of working on these other projects with short completion timeframes was that the project manager and the digital conversion technician were able to adjust workflows to produce excellently digitized work products under the short deadlines. This experience helped the team work more efficiently on the completion of the Hebrew manuscripts, especially as one of the other short deadline projects involved a substantial number of Hebrew books. In the earlier phases of the project, when there were issues with capture devices or post-processing software, documentation on how to troubleshoot or correct the issue was

absent. This resulted in some setbacks as the project team figured out solutions. DSC documentation is more complete now but should still be revised regularly to serve as a better resource for staff.

Because of the external partnership with the Berg Foundation, the project required a higher level of tracking and reporting than other special collections digitization projects. The project team spent a considerable amount of time tracking data related to the time to digitize each manuscript, individual team member activities, and overall project progress. This was one of the first times these reporting requirements were in place for a DSC project, which required some project management process adjustment. Other Library partners involved with this project, including AMED, the Conservation Division, and the now-Digital Collections Management and Services Division (DCMS), were all required to report staff time (hours per pay period for all staff members involved) and other expenses for supplies, etc. to the Library's Financial Management Directorate (FMD) to report to the Berg Foundation several times throughout the lifetime of the project.

The Library also reported digitization progress (number of images produced, items completed, items in progress, etc.) and Conservation status of items to the Berg Foundation regularly. AMED was the primary author/compiler of the reports, with input from other partners. FMD managed the formal financial relationship with the Berg Foundation. The Library was not required to provide specific internal project plans or strategy documents (such as the web launch strategy) to the Berg Foundation but did provide general information about the project approach and progress in its regular reporting.

In some cases, the complex nature of the individual manuscripts necessitated consultation with staff both within the DSC and other parts of the library, such as the Conservation Division. Several manuscripts were found to have iron-gall ink and cracks which needed further treatment from the CD before they were stable enough to image without damage. One manuscript contained a white powdery substance, confirmed by CD to be toxic lead arsenate, which had to be handled by CD while DSC staff carefully completed the imaging. The Library budgeted as many as 100 hours for conservation treatment of these manuscripts in its agreement with the Berg Foundation. Conservation treatment was not always accomplished in a steady, rolling cadence as planned due to extensive treatment needs for certain manuscripts and other priority projects for the Conservation Division.

In most cases, the manuscripts had page numerals in Hebrew letters corresponding to the number of folia, and corresponding Arabic numerals later penciled in by catalogers; for example, the Hebrew letter aleph א equals the cardinal number 1, bet ב equals 2, and so on [6]. However, in some cases, page numerals were absent or the pages were bound out of order. In these instances, Ariel Segal, the digital conversion technician, who is fluent in Hebrew, would ensure that the text was in correct order, by seeing which words were at the end of a page and if they corresponded to the text on the next page. The odd numbered page image names, such as 0001, 0003, usually corresponded to the recto (actually left in Hebrew text) pages unless a page was imaged twice to capture text hidden under flaps or hinges, which would change whether the image numbers corresponded to recto or verso pages. On one occasion, Ariel compared the text of a digitized online printed version of a book with its manuscript to ascertain correct page order.

As an example of complexity, MS 2, Tokhehot Mussar ("Moral Reproofs") had many subtle text hinges under which alternative Hebrew text was found. (See Figure 3). In the end, the head of the Digitization Services Section said that it would have too much potential for damage to image all the hinges both open and closed. In the future, if the DSC digitizes Hebrew printed books from the 15th and 16th centuries, it would be useful to sit down with Hebraica section staff and get their advice as to what to do in such special cases.

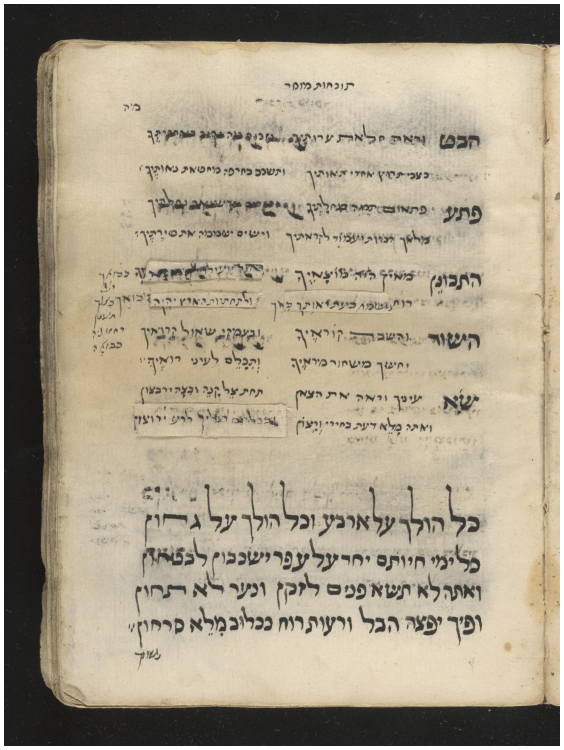

Figure 3. *MS 2, "Tokhekhot Mussar," leaf showing paste-on insertions of Hebrew text, hinged to allow for observation of alternate text underneath. [7]*

Imaging Techniques

In the newly renovated Scan Center, the project team developed a comprehensive imaging workflow. All of the manuscripts were imaged on Phase One iXG 100-megapixel camera systems. If the manuscripts had a wide enough gutter and could withstand gentle pressure in the spine, they could be imaged on a Digital Transitions BC100 station, which can image both leaves in a spread at once, and also uses Phase One cameras. Images were post-processed using a variety of tools, including Capture One 20 Cultural Heritage software.

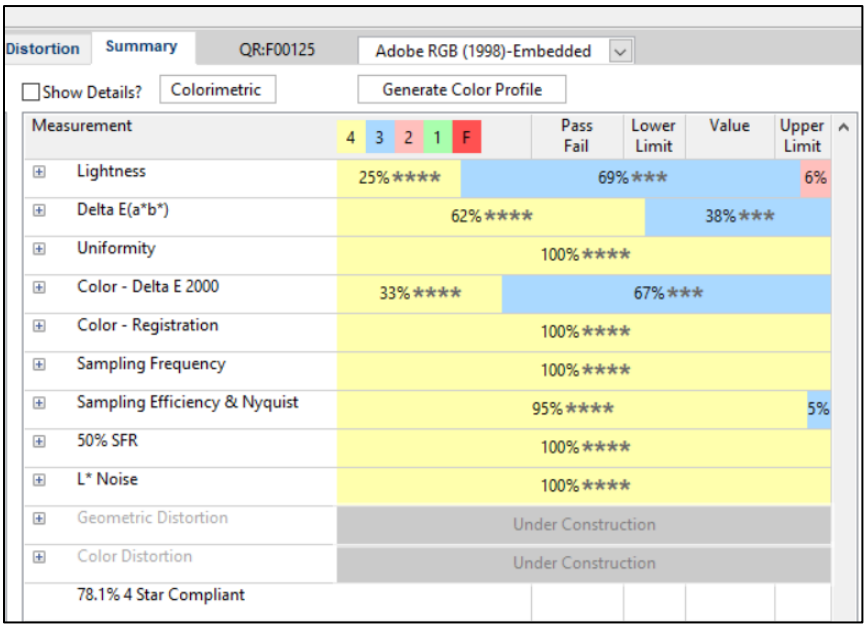

***Figure 4**. Target image analysis data summary from an unconverted target image captured during this project. The data indicates that this unconverted image is 78.1% FADGI 4-star compliant.*

After post-processing and an initial quality review, project team members put all image files through the Golden Convert program developed by Image Science Associates. The software uses images of FADGI 19264 targets for device calibration. Each target has a unique QR code and 168 color patches that are "read" by the program and used to create a color profile specific to the target and the imaging session where the calibration image was produced. The output files include an ICC profile (.icm) file and a .ti3 file with a table containing colorimetric data. The software can now use different types of targets as the source calibration image; however, the Library of Congress team only used FADGI 19264 targets for this project.

The program then applies the custom color profile to each of the TIFF image files from that session, and "converts" them to produce new, color-corrected TIFFs in a user-selectable working color space (Adobe RGB 1998, etc.). The program also embeds color profile information into the metadata for the output TIFF files. The converted output images consistently appeared more true-to-life than the original images. This visual observation of improvement in the color of the images was supported by target image analysis data (Figures 4 and 5, right). The biggest improvements were indicated in the Lightness, Delta E(a*b*), and Color – Delta E 2000 image quality evaluation parameters, as indicated by analysis data from the Golden Thread NXT software by Image Science Associates.

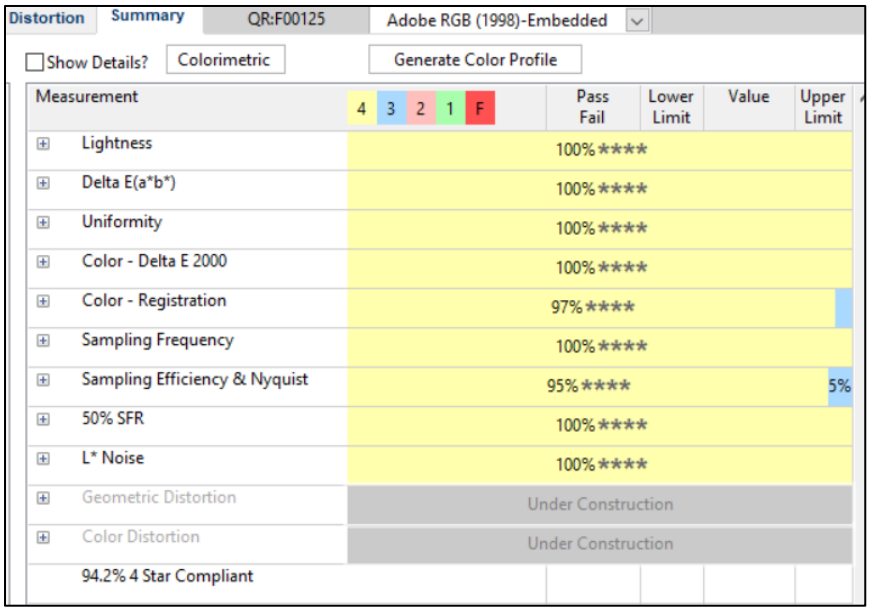

***Figure 5**. Target image analysis data summary from the same target image from Figure 4 above, captured during this project, after going through the Golden Convert software improvements. The data indicates that this converted image is 94.2% FADGI 4-star compliant.*

A striking example of the improvement in detail and color made possible during this project by both initial imaging and post-processing using Golden Convert can be seen in 3 images of the same page of MS 31 (Figure 6, below).

*Contrast between 3 versions of MS 31, "Responsum on the validity of the kiddushin (betrothal) of two daughters…". **Figure 6a,** leaf 26r microfilmed prior to project (top); **Figure 6b,** captured during project before undergoing Golden Convert process (bottom left); and **Figure 6c**, subsequent to Golden Convert process (bottom right). Note the exposure of previously unseen autograph gloss by R. Moses Benveniste including the Hebrew words "safek kiddushin" (doubtful betrothal). [8]*

Results

The team was able to use their unique talents to complement each other. The imaging team was expanded mid-project and included many talented people to share the work and meet project goals. Despite the extensive construction in the Scan Center and pauses to the project for the completion of other priority projects, the Hebrew Manuscripts were all digitized and available online before the desired deadline of May 2023 to coincide with Jewish American Heritage Month. The digitization of the collection was officially announced in August 2023 [9].

The project team held an after-action meeting to discuss lessons learned during project execution and documented the strengths and areas for improvement. Several elements of the workflow and project management have been carried over to other large, long-term special collections digitization projects in the Scan Center. This includes project tracking and collaboration, as well as the Golden Convert step in the digitization workflow. The DSC now also has more comprehensive documentation for its workflows, equipment operation, and project tracking.

Conclusions

This project was significant to the Library's Scan Center, as proof of concept for new and improved equipment and workflows, and as a demonstration of the team's capability to execute a high-profile project on a deadline with the added attention that comes with an external partner agreement. Subsequent projects have benefitted from the lessons learned and technical elements that were refined during this project. Researchers and the general public have also benefitted from the availability of high-quality digital surrogates for these rare and valuable manuscripts. It is possible that the improvements in color and detail made available by Golden Convert could prove useful to ongoing efforts to use artificial intelligence to enable optical character recognition for Hebrew manuscripts [10].

References

[1] To access these leaves, see https://www.loc.gov/resource/amedscd.2018757799/?sp=5 and https://www.loc.gov/resource/amedscd.2018757799/?sp=32 respectively.

[2] To access this folio, see https://www.loc.gov/resource/amedscd.2018757847/?sp=15 . For more information on MS 229, see Portrait of the Artist as Rain(bow) Maker: Joseph ben Meir Schmalkalden (https://blogs.loc.gov/international-collections/2020/11/portrait-of-the-artist-as-rainbow-maker-joseph-ben-meir-schmalkalden/ , November 13, 2020).

[3] These most recent manuscripts, by Rabbi Mnachem Risikoff, can be accessed here: https://www.loc.gov/collections/hebraic-manuscripts/?q=risikoff ; They were donated by his grandson, a military chaplain and uncle of Ariel's friend.

[4] David Berg Foundation (https://bergfoundation.org/, 2024).

[5] *FADGI Technical Guidelines for Digitizing Cultural Heritage Materials, Third Edition* (https://www.digitizationguidelines.gov/guidelines/digitize-technical.html , 2023).

[6] For more details on Hebrew numerals, see https://en.wikipedia.org/wiki/Hebrew_numerals .

[7] To access this leaf and see the paste-on insertions/hinges, see https://www.loc.gov/resource/amedscd.2018757619/?sp=91 .

[8] To access this leaf and see the gloss in the gutter, see https://www.loc.gov/resource/amedscd.2018757649/?sp=54 .

[9] "Library of Congress Releases Newly Digitized Hebrew Manuscripts" (https://newsroom.loc.gov/news/library-of-congress-releases-newly-digitized-hebrew-manuscripts/s/f2a83449-6dbf-4d58-a8db-ebac70c51a85 , August 22, 2023).

[10] For an example of such efforts, see Daniel Stoekl Ben Ezra (2023, February 3). ERC Synergy MiDRASH Awarded to EPHE and partner institutions. eScripta. Retrieved February 29, 2024, from https://doi.org/10.58079/oieq

Author Biographies

Ariel Segal is a Digital Conversion Technician at the Library of Congress. Notable collections he has digitized include LC's Hebrew manuscripts, and 19th century credit rating books. He has an MLS, specializing in Archives, and an BA and MA in History from the University of Maryland.

Hana Beckerle is a Digital Imaging Specialist at the Library of Congress. She previously worked in Electronic Resources and digitization at the Catholic University of America's (CUA) Mullen Library. She has degrees in journalism and political science from Penn State University, and a Master's in Library and Information Science from CUA.