# The Edward Blank YIVO Vilna Online Collections Project: A Case Study

**Stefanie Halpern**
*YIVO Institute for Jewish Research*

## Introduction

In January 2022, the YIVO Institute for Jewish Research completed the Edward Blank YIVO Vilna Online Collections Project (VCP or "Vilna Project" for short), a major international initiative to virtually reunite YIVO's prewar collection held in New York and three Lithuanian institutions: the Lithuanian Central State Archives, the Martynas Mažvydas National Library of Lithuania, and the Wroblewski Library of the Lithuanian Academy of Sciences. This project, the largest and most comprehensive digitization initiative YIVO had ever embarked on, took seven years to complete and had a budget of nearly 7 million dollars. It resulted in making approximately 1.5 million pages of archival documents and over 8,000 volumes of books freely available online, opening many avenues of research to users around the world. The project proved a great success for YIVO, allowing not only for the preservation of some of YIVO's most important and highly used collections, but also for forging positive relationships with the Lithuanian repositories who were now the custodians of those portions of YIVO's prewar collection that had not been restituted to YIVO in New York in 1947.

A complete history of what became of YIVO's prewar collections from its looting by the Nazis, to its partial restitution to YIVO in New York, to the discovery of additional material in Lithuania after the fall of the Soviet Union, to the several failed attempts at partnership with the Lithuanian archival institutions in possession of YIVO materials during the 1990s and early 2000s was covered in detail in "A History of YIVO's Prewar Archival Collections from 1925 to 2001," published in this journal (Halpern 2022).

The current article, presented as a case study, details the particularities of the Edward Blank YIVO Vilna Online Collections Project between its initial conception (2013–2015) and its completion in January 2022. This study outlines the methodologies YIVO and its colleagues in Lithuania employed during each phase of work, including project planning, budget creation, and the building of an overall project timeline; work done on the materials in New York and Lithuania, including conservation, processing, and digitization; and the building of a dedicated web portal through which patrons would be able to access digitized materials.

 Halpern, Stefanie. 2024. Judaica Librarianship 23, 25–57.
https://doi.org/10.14263/23/2024/1393

This case study attempts to promote transparency regarding a project of this scope and size, highlighting certain staffing, procedural, and workflow needs that YIVO failed to anticipate because it had little experience planning and implementing a conservation, processing, and digitization effort of such complexity. In that vein, the case study describes how, after several years of project work, YIVO began to fall behind its intended goals. The discussion covers roadblocks and missteps that led to these setbacks, many of which were discovered after a mid-project assessment. Ultimately, it was realized that many of the issues stemmed from inadequate planning during the initial phase of the project and a lack of proper tracking and reporting. Additionally, this case study details the procedures and workflows that were implemented following the mid-project assessment, including the creation of a new timeline based on surveys completed by conservators and processing archivists; the opening of an in-house digital lab rather than continuing to contract with a vendor; the implementation of tracking and reporting procedures; and abandoning the idea of creating a proprietary web portal, separate from the library systems already in place, which would have required dedicated funding for maintenance in perpetuity.

Though the project proceeded smoothly after a new timeline and procedures were implemented, it faced yet another obstacle in its last years: the COVID-19 pandemic. This study also discusses how YIVO's team was able to overcome the delays and hurdles brought on by this unexpected event.

Finally, the case study presents a set of guidelines that resulted from the lessons learned throughout the Vilna Project. YIVO now implements these guidelines for all its digitization initiatives, including building ample time and funding into the timeline and budget for adequate planning before any project work begins; basing the timeline and budget on comprehensive surveys of archival and library materials completed by project staff; building in additional time and funding to account for scope creep[1] as well as vacation and sick time for staff; and staggering the start of each phase of the project so that staff for each subsequent phase has a steady stream of work.

Though the materials in the Vilna Collections Project are unique to YIVO, the challenges, solutions, and guidelines presented in this case study may serve as a road-map for other institutions embarking on conservation, processing, and digitization projects for the first time, or as points to consider for those seeking to adjust their current practices.

❐❐❐

1. A project management term: "Adding additional features or functions of a new product, requirements, or work that is not authorized (i.e., beyond the agreed-upon scope)." See Richard Larson and Elizabeth Larson, "Top Five Causes of Scope Creep... and What to Do About Them," in *PMI® Global Congress* (2009): 59–62.

There is little evidence to suggest that YIVO had any ongoing contact with the Lithuanian Central State Archives (LCVA) or the Martynas Mažvydas National Library of Lithuania (MMNLL) following the six-year project allowing YIVO to photocopy its prewar materials held at the LCVA (1995–2001). YIVO's delay in returning materials to the LCVA created a contentious relationship between the two institutions. In 2011, YIVO's recently appointed executive director, Jonathan Brent, was invited to Vilnius to participate in an event commemorating the liquidation of the Vilna Ghetto. Aware of the long-standing dispute regarding YIVO's prewar materials found in Lithuania, Brent began a series of conversations regarding the possibilities of recovering these collections. With the understanding that the materials at the LCVA and the MMNLL would never make their way to New York, he instead planted the seed for another idea: putting aside questions of ownership and reuniting these collections virtually through digitization.

Throughout 2012 and 2013, Brent met several times with Emanuelis Zingeris, a longtime member of the Seimas of Lithuania, to discuss the viability of reviving the relationship between YIVO, the LCVA, and the MMNLL (Brent 2013). Zingeris had been the elected chairman of the Jewish community of Vilnius when YIVO's materials were first discovered after the fall of the Soviet Union. At the time, he had also played a role in YIVO's initial negotiations with their Lithuanian colleagues. Zingeris also helped set up a meeting with Renaldas Gudauskas, the recently appointed director general of the MMNLL. Gudauskas met Brent's ideas for a joint digitization and access project with much positive enthusiasm. With the help of Zingeris, high-profile members of the Lithuanian Jewish community and representatives of Lithuania's government warmed to the digitization project idea. With these pledges of support, YIVO began planning the Vilna Collections Project.

It was clear from the start that the Vilna Collections Project would encompass archival and library components pulled from the collections housed at YIVO in New York and those housed in Lithuania. The criteria for inclusion of those archival collections in New York was that they had been collected by YIVO in Vilna between 1925 and 1940 and eventually restituted to YIVO in 1947.[2] Included in the project were also the materials known as the Sutzkever-Kaczerginski Collection (RG 223), which had been smuggled out of the YIVO building and hidden in the Vilna Ghetto by members of the Paper Brigade as well as those materials that the Paper Brigade had collected within the Vilna Ghetto itself. These materials had not been restituted by the United States government but rather had been delivered to YIVO by various individuals in the postwar years. Initial plans projected the conservation, arrangement and description, digitization, and availability online of 64 separate collections from the YIVO Archives in New York, totaling approximately 515 linear feet of material.

2. For a list of collections included in the project, see https://vilnacollections.yivo.org/Browse.

In Lithuania, the archival component was to be composed of those materials which had been discovered at the LCVA in 1990 and which formed the photocopying project that took place between 1995 and 2001 (Halpern 2022). More materials would be uncovered at the Wroblewski Library and the National Library of Lithuania in 2015 and 2017 respectively, and these materials would eventually be included in the scope of the Vilna Project.

In total, initial estimates for the number of archival pages that were to be digitized amounted to 1,014,000 pages (794,000 in New York and 220,000 in Lithuania). By the end of the project, this number would grow to 1.5 million pages (1.1 million in New York and 400,000 in Lithuania).

The library portion of the project was originally conceived of to encompass 6,000 books in New York and 2,200 books in Vilnius held by the MMNLL. Though many more than 8,200 books had been restituted to YIVO, the Vilna Project was concerned with several different categories: books that would help YIVO "digitally reconstruct the historic private Strashun Library of Vilna, one of the great prewar libraries of Europe" (YIVO Vilna Collections Newsletter 2015), as well as books in Yiddish and other languages that YIVO had collected before World War II.

During the initial planning phases, YIVO decided to maintain a small project team in New York consisting of seven new hires that would be supplemented by volunteers and interns. These hires included a project manager, two processing archivists, a project librarian, two digitization specialists, and a quality assurance librarian. YIVO let their Lithuanian colleagues determine the number of staff members needed to complete their portion of the conservation, processing, cataloging, and digitization work in their respective institutions.

The original budget as set out in 2014 for the archives and library portions of the project in New York and Vilnius was estimated at $5.6 million (Reclamation 2014). After a major assessment of the project and subsequent restructuring in 2018, the budget was adjusted to $7 million.

## Initial Planning in Lithuania

Before any work on the project commenced in New York, plans first needed to be made with YIVO's partners in Lithuania, as one of the major goals of the project was to gain access to those materials that had not been made widely available to researchers. With seed funding provided by the Rothschild Foundation, YIVO undertook several trips in 2013 and 2014 to plan for the inclusion of YIVO materials at the LVCA (archives portion; the existence of archival materials in other repositories was unknown at that time) and the MMNLL (library portion). By mid-2014, YIVO obtained official letters in support of the project from both the director of the LCVA and the director of the MMNLL. Both institutions concurred that the project's goal was "to make

these important research materials as accessible as possible to researchers and the global community of scholars" (Gudauskas to Brent 2014).

The Ministry of Culture of the Republic of Lithuania also expressed its official commitment to the Project in April 2015 and reiterated the intent of both the LCVA and the MMNLL to participate in YIVO's project. The Ministry of Culture laid out the in-kind contributions both institutions were willing to make to the project: 347,090 EUR in total (approximately $367,160) from their budgetary allocations. The contributions from LCVA would consist of "human resources, relevant digitizing and restoration equipment and applying a discount for digitization services" (Jarockis to McCarthy 2015). In total, the cost of salaries for the Archives was estimated at 165,711 EUR. They also stated that they would provide a 50 percent discount on digitization services, amounting to approximately 50,567 EUR ($54,017) in savings. The MMNLL agreed to create three new positions (two metadata specialists and one restoration specialist) for the duration of the 4-year project, projecting that salary costs for the project would be 116,000 EUR plus 11,892 EUR in administrative costs. The letter also stated that if the LCVA and MMNLL needed additional financial resources they would apply to the Lithuanian Culture Council for support under the "Restoration and Preservation of Movable Cultural Properties Stored in the Institutions of Remembrance" program (Jarockis to McCarthy 2015).

In April 2015, YIVO and the LCVA signed an agreement that stipulated that Fonds 287, Opis 1-33, namely the materials that were sent to New York in the 1990s, photocopied, and then returned to the LCVA, would be digitized over the first two years of the project. Additional unsorted materials from that collection that had never been sent to New York would also be included in the project. In total, this amounted to approximately 157 linear feet of material.

YIVO agreed to reimburse LCVA every four months for staff salaries, conservation supplies, and digitization of documents (at $0.25 per page) for a total reimbursement cost of $39,700 for the first year of the project. In return, the LCVA agreed to send YIVO digital images every four months (YIVO-LCVA 2015). Two additional amendments followed in 2016 and 2018 that covered work through 2019. YIVO agreed to pay an additional $120,000 to cover work done by the LCVA.

The guiding principle of providing access to materials at the LCVA was to follow the organization of the materials as they were physically arranged in Lithuania. The hope was that this would minimize confusion among researchers who may wish to see the original materials. The order of the 33 series (what LCVA called "Opis," and YIVO referred to as "lists") as per the LCVA arrangement would be retained as well as the same folder numbering. A finding aid would be created to represent these materials as a single collection, with each list comprising a sub-series

of the overall collection. Because it was later discovered that these lists often included materials of mixed provenance, YIVO's archivists would later need to make additional intellectual arrangements of the materials within each list to more accurately reflect the corresponding YIVO collections in NYC.[3]

The library portion of the project in Lithuania was to come exclusively from books housed at the MMNLL. In April 2015, YIVO signed a first amendment to the Memorandum of Understanding with the MMNLL. This amendment set out a work plan for the library items that would be included in the project. The agreement prioritized the digitization of books in the following order:

1. Lithuania-published books from the Strashun Library which were never digitized by any library. That corpus included approximately 600 books, 200 of which had already been cataloged and would be conserved and digitized before the 400 books that had no associated metadata.

2. Lithuania-published Yiddish books which were never digitized by any library.

As with the archival materials from LCVA, digital images would be delivered to YIVO every four months on an external hard drive. YIVO would be given 300dpi color JPEGs and a bibliographic description of the original publication created within the National Bibliographic Data Bank which would be delivered in the ISO 2709 format. YIVO would be responsible for reimbursing the MMNLL for staff salaries and conservation supplies totaling $17,000 for the first year of the project (YIVO-MMNLL 2015).

A second amendment was signed in June 2016 which added an additional 350 books from the Strashun Library, specifically in Hebrew, every year for the next two years of the project, totaling at least 700 books; and Yiddish books published from 1851–1920 in year 2 and Yiddish books published from 1921–1940 in year 3 of the project. The amendment stipulated that at least 1,500 Yiddish books would be digitized over years 2 and 3 of the project. This agreement also included the cataloging and digitization of 400 as-of-yet uncatalogued Strashun Library books printed outside of Lithuania. The costs for year 2 and year 3 of the project, which was to be covered by YIVO, amounted to $61,820 each year (YIVO-MMNLL 2016).

3. Maintaining the titles as "List 1," "List 2," etc. provided no descriptive or contextual information for researchers to adequately use these materials. For this reason, it was later decided to name each folder according to the corresponding YIVO RG.

The third amendment, signed on March 27, 2018, included the conservation and digitization of 1,200 books, including approximately 500 remaining "Strashun books." The cost of this would be $61,820, again covered by YIVO (YIVO-MMNLL 2018). Moreover, it was discovered in 2015 that additional and yet unknown archival materials collected by YIVO before World War II were discovered at the Wroblewski Library of the Lithuanian Academy of Sciences. The collection consisted of approximately 18,950 pages (9,475 folios). In 2018, the two institutions signed an agreement stating that the materials at the Wroblewski Library were those that "the YIVO believes were formerly part of its archives in Vilnius," and that they would be included in the Vilna Project. The Wroblewski Library named this collection F424 of the YIVO Institute for Jewish Research. YIVO agreed to fund all aspects of this part of the project, which the Wroblewski Library quoted at $27,380. The entire collection was to be digitized and delivered to YIVO in the fall of 2018 (YIVO-Wroblewski 2018).

In 2017, another discovery had been made: the MMNLL revealed the existence of an estimated 170,000 pages of archival documents. The MMNLL described them as "previously unknown Jewish archival materials and manuscript books" (YIVO-MMNLL 2017). These documents were found in a locked basement of the MMNLL.[4] Based on YIVO's history with the MMNLL, it is now believed that this set of documents, which for purposes of clarity YIVO calls the "New Discoveries," had first been uncovered in the Book Camber of the Lithuanian SSR in 1991 (later renamed the Bibliographic Center of the Martynas Mažvydas National Library of Lithuania). However, by the time the MMNLL informed YIVO about their existence in 2017, no one in either institution knew that these materials existed, due in no small part to the various leadership changes experienced at both institutions over several decades. It seems that the MMNLL informed YIVO of the existence of these materials in 2017 in part because the relations between the two institutions had matured and improved.

Like those materials first discovered in the Book Chamber after the fall of the Soviet Union, the "New Discoveries" were in varying states of disarray, packed in 200 aging cardboard folders, 14 cardboard boxes, and in the form of loose materials with no enclosures, while others were placed into 109 acid-free archival boxes more recently.

The MMNLL shared highlights found within these materials, including several *pinkasim*; rabbinical manuscripts; materials from the Strashun library; letters of Simon Dubnow; and literary manuscripts of Yiddish writers. The materials also included YIVO administrative records; portions of the YIVO-created Esther Rokhl Kaminska Yiddish Theater Museum Collection; portions of the Vilbig collection; ORT Archives; materials on the Vilna Jewish Community; and materials relating to Tsemakh Szabad.

4. The new MMNLL building was opened in 1963 with an extension built in 2003. It is assumed that the materials were moved into the storage space sometime during the construction of the extension.

It was eventually decided that only those materials that MMNLL and YIVO could agree were of YIVO provenance would be included in the project. Here, the MMNLL considered "provenance" in the narrowest terms, meaning that only materials which YIVO created or were collected as part of a specific YIVO-project could be included in the digitization efforts. Those materials that were not created by YIVO formed much of the category of "documents and printed materials acquired or collected after World War II by the Jewish Museum in Vilnius." Though many of these materials corresponded to those prewar collections that had been restituted to YIVO after the war, the MMNLL made clear that only those materials that formed the "National Library YIVO Materials" would be explicitly included in the project (Categories 2019; YIVO-MMNLL 2017). Materials that YIVO claimed to have collected before the war would be included only at the discretion of the MMNLL.

## Initial Planning in New York

In order to determine a work plan for the archives portion of the project in New York, YIVO conducted an initial survey in 2015 detailing various preservation and processing needs.[5] The survey indicated that of the 39 collections that needed preservation, 29 simply needed rehousing into acid-free folders and boxes. Only ten collections were earmarked as needing more extensive preservation measures (Processing Priorities 2015). The survey also noted those collections that had an Encoded Archival Description (EAD) finding aid, those that had only a container list, the languages present in each collection, the extent of each collection based on information contained in the catalog record, and an estimated page count based on a calculation of 1,500 pages per linear foot. The resulting plan mapped out processing work for the project, taking into account that of the 502 linear feet of archival materials (715,358 pages), 40 percent already had an existing EAD finding aid, and therefore it would take 2,294 workdays to complete the processing. This calculation assumed that processing would proceed at a rate of 50 hours per linear foot for those collections which had no arrangement and description, and at a rate of 15 hours per linear foot for those collections which already had an existing finding aid of some sort. The work plan indicated that for the 40 percent of collections described in a finding aid, the only work the processing archivists were expected to complete was merely checking the physical collection against the finding aid for accuracy (Finding Aids 2015).

An additional 175 working days were allotted for the writing and encoding of new finding aids based on a rate of two linear feet per day. Finally, revisions on existing finding aids were expected to take a total of 81 working days (Priorities and Workflow 2015). Assuming the two processing archivists hired for the project each worked 260 days per calendar year, the processing portion of the project was slated to take less than 5 years to complete, with all materials delivered for digitization by 2020. This processing timeline was used as the basis for the overall project budget and to determine the number of years needed to complete the project.

5. For definitions of archival terms used throughout the article, refer to https://dictionary.archivists.org.

A basic workflow for processing archival collections was distributed to the project team. A specific list of collections was assigned to each processing archivist, but the order in which they should work on those collections was not determined at this stage. After being assigned a single collection, the processing archivists were instructed to first survey the collection they were assigned, research biographical information, note bulk dates, list supplies needed, anticipate any privacy or copyright concerns, and determine the involvement of other staff needed to process the collection, including volunteers and interns. Processing archivists were instructed that processing the Vilna collections could take many different forms, from completely rearranging collections to expanding description, to simply checking that the contents matched the existing description. The processing archivists were also asked to write a processing plan as they began working on a collection. The plans were to include a proposed arrangement and any other observations gleaned during a survey of the materials.

The library portion of the project in New York was initially conceived of as comprising several components. The first was a reconstruction of the private library of the famed scholar and book collector Matisyahu Strashun. Originally housed in his private residence in Vilna, at the time of his death the contents of Strashun's personal library were willed to the Vilna Jewish community. These books became the core of the public Strashun Library, open to all wishing to use the collection, which grew to over 50,000 books by 1940. The Strashun Library was among those institutions looted by Nazis. Nearly half of the Library's collection was among those books restituted to YIVO by the United States Government in 1947.[6]

To manage the Strashun portion of the project, YIVO created a spreadsheet of *Liḳute Shoshanim*, the 1889 catalog of the 5,753 books held in Stashun's private library. The spreadsheet noted which books had already been digitized by other institutions, which ones were owned by YIVO, and which ones had already been digitized by YIVO for other unrelated projects.

The project also was meant to include Yiddish books in YIVO's collection that had been recovered from the Offenbach Archival Depot after the war, estimated at approximately 5,000 books, to complement the Yiddish books that were being scanned as part of the project in Lithuania. Some of the books were on microfilm and others would be scanned from the originals. Many of these did not have a catalog record and would need to be cataloged by YIVO (Outline Proposal nd). A list of the titles of Yiddish books from YIVO's collection that would be included in the project had not been compiled before the digitization of the materials began. YIVO intended that these scanned books would be made available in several different locations: the HebrewBooks

6. For a history of the Strashun Library and the contentious history of the restitution of the collection to YIVO, see Rabinowitz 2018.

website;[7] a new (but not-yet-designed) YIVO library catalog; and a special website and search portal dedicated to the Vilna Collections Project (Outline Proposal nd).

In April 2015, YIVO began making plans for digitizing the materials. Several options were pursued, including performing digitization in-house or contracting with an outside vendor. After receiving several vendor quotes and pricing out other options, YIVO decided that opening and operating its own digital lab would not be wise. Though staff members would be under the direct supervision of YIVO, and though there was capacity building potential, YIVO believed that purchasing equipment ($40,000), training staff members, and supervising the digitization work directly were not worth the effort or estimated cost of approximately 1 million dollars.[8]

In July 2015, YIVO opted to contract with an outside vendor who provided an initial quote of $678,246, well below that of the other vendors with whom YIVO had been in contact.[9] This particular vendor was not only appealing because of the low price quoted for the project and the assurance that YIVO's timeline could be met, but also because they were willing to have their technicians work on-site at YIVO. Having the scanning technicians working on-site was important to YIVO since the uniqueness of the materials, their history, and their fragility did not allow for the materials to be shipped off-site (Update 2015). The selected vendor's quoted work plan included scanning 7,000 library books (5,000 Yiddish books and 2,000 Strashun books) and 800,000 pages of archival documents. The cost estimate for the library materials was done on a per-page basis. YIVO provided the vendor with an estimate of 171 pages per book, which totaled 1,200,000 book pages to be scanned. Half of these were to be scanned using a Kirtas Robotic Book Scanner and the other half were meant to be scanned using a Minolta Manual Book Scanner. The vendor estimated that it would take 160 working days (8 months) to scan all book pages with two people working on this portion of the project (at an estimated 3,750 book pages scanned per person per shift). The number of archival materials to be scanned was estimated by YIVO to be approximately 800,000 pages. The materials would be scanned using a Plusten OpticBook A300 Flatbed Scanner. The archival materials, which could be scanned at a rate of 2,500 pages per person per shift, were estimated to take 160 working days for two individuals to scan. While the book scanning was to be done consecutively throughout an 8-to-10-month period, the archival portion would be done in large batches over six years. This would allow the processing archivists to accumulate enough processed materials to pass along to the scanning technicians in

7. https://hebrewbooks.org.

8. This estimate turned out to be less costly than some of the outside vendors YIVO requested quotes from, ranging from 1.3 million dollars for a vendor who offered to digitize materials in batches several times per year to a 3.2 million dollar quote from another vendor which included all digitization work from capture, derivative creation, quality assurance, descriptive metadata import, ingest, and post-ingest quality control (Digitization Options 2015).

9. The total cost of scanning the books was $380,646 and the cost of scanning the archival documents was $297,600. See Contract for Digitization 2015.

large batches. Though the contract stated that the scanning would only take a total of 10 months in work time, technicians would only be sent to YIVO when a batch of 125,000 pages was ready to be scanned (Contract for Digitization 2015).

All pages were to be scanned as 400dpi TIFF files. These would be copied to a removable hard drive and taken to the vendor's facility for image processing including cropping and straightening. The person scanning the documents would also perform quality assurance on the materials, noting poor image quality including pages that were too light; too dark; cut off, etc. These images were to be rescanned at no additional charge. However, there was an additional 20 percent fee charged to have the two vendor technicians working on-site at YIVO, a total of $113,000 (Contract for Digitization 2015).

Another essential part of the project as initially planned was building a separate web portal as an entry point for users to access digitized materials. YIVO created a request for proposals (RFP) that was sent out in 2016, specifying both the technical requirements as well as the total budget for the creation of the website. YIVO's wish was that the newly designed portal would allow for a federated search with results drawn from several different sites. Archival metadata from YIVO's EAD finding aids would be pulled from the *Guide to the YIVO Archives*, an online search portal built on a customized platform of the Archon open-source software. The Vilna portal also intended to pull metadata from a new Aleph-based catalog of YIVO library records which had not yet been built. Digital images would be drawn from DigiTool, the Digital Asset Management System in use at the time at the Center for Jewish History, of which YIVO is one of five partners. YIVO hoped that search results would also be pulled from related content in articles appearing in *The YIVO Encyclopedia of Jews in Eastern Europe*. Finally, the portal was meant to display exhibitions drawn from materials from the project that would be built on some type of open-source software such as Omeka (NEH 2015, 8). Based on responses to the 2016 RFP, YIVO chose a vendor to create the portal who presented a plan that closely aligned with YIVO's vision. The vendor did note, however, that they would not be able to deliver all the aspirations outlined in YIVO's RFP within YIVO's stated budget but did believe they would deliver a flexible platform that would be capable of adaptation in future stages of the portal's development.

## Project Work in New York, 2015–2017

The project suffered setbacks early in the process. As work commenced at YIVO, it became clear that the workflows and timeline had been inadequately planned. This would have a ripple effect on every aspect of the project, causing major delays to various deliverables. For example, though the processing archivists were asked to create processing plans, it seems that the overall project timeline did not take the additional information gleaned from these processing plans into account (At a Glance 2015). Because this pre-processing work was done on a collection-by-collection

basis as each processing archivist was assigned a new collection, it was ultimately difficult to determine a holistic understanding of the needs of all 64 collections that made up the project.

As part of this initial pre-processing work, the processing archivists were expected to consult with the project's management team for assistance in finding appropriate resources and in developing a workflow tailored to the needs of each collection, but some members of this team only worked part-time on the project (Digitization Workflow 2015). Before the start of the project, none of the finding aids, container lists, or authority files for each collection were gathered in a single easily accessible location, so each processing archivist was expected to locate all iterations of previous finding aids, of which there were often multiple copies with different descriptive metadata. This, combined with the need for frequent meetings with multiple members of YIVO's archival staff to make simple process decisions, caused additional holdups in the processing timeline.

As processing commenced, it was discovered that the overall timeline was based on several false assumptions. The one that affected the timeline most assumed that for those collections that already had an existing finding aid the only work that needed to be done was checking for accuracy. Though many of these collections had indeed been processed more than once over the years, the processing archivists found that there were widespread issues with many of the already processed collections. These included many missing or misplaced documents; finding aids whose folder descriptions did not necessarily match with the contents of the collection; large portions of the collection that had not been processed or accounted for in the finding aids; and undocumented processing decisions made over the years that deviated from the principles of provenance and original order.

In addition, though the overall survey had tried to assess preservation needs, there had been no systematic preservation survey conducted by YIVO's lead conservator. Therefore, there was only a limited understanding of the needs of the collections. Moreover, no full-time conservators were assigned to the Vilna Project in the initial stages, so it fell to the processing archivists to perform basic preservation (open folds, remove staples, and flatten and clean items) as they were processing the collections.

Once the processing archivists began looking at the physical collections, it was determined that far greater time needed to be spent on preserving each of the 64 collections than initially described in the survey. A new procedure was put into place, directing the processing archivists to flag any materials they could not ameliorate by the basic preservation steps for review by YIVO's conservator. The conservator would then remove the folder from the collection, evaluate the item, and perform treatment (Digitization Workflow 2015). The removal of items and folders from the collection led to issues of physical control, as it was not always adequately documented.

This method of flagging items for conservation often confused the conservator, who did not have a schedule and order of priorities.

In addition to accommodating a longer timeline based on the processing archivists' plan, the added preservation tasks increased the number of working days the archivists would need to complete the arrangement and description. What is more, because those items needing additional treatment were evaluated and treated by the conservator on an ad hoc basis, and because a processing archivist often could not fully complete the arrangement and description of a collection until the folders were returned to them, the timeline for the completion of processing became not only protracted but also unpredictable very early on in the project.

By the end of the first year of the project, the processing archivists had prepared only approximately 19,000 of the 226,500 pages (about 8 percent) that were projected to be ready for digitization according to the project timeline. Though processing was on target in both 2016 and 2017 (101,000 of 108,000 pages each year for both years), there were still setbacks that stemmed from the first year's complications. In total, by the end of 2017, only 50 percent of a projected 442,500 pages were fully processed and ready for digitization (Progress 2018).

The digitization of archival documents also suffered some major setbacks as early as 2015. Though it was projected that 70,000 pages of archival documents would be scanned by the end of that year, only 50 percent of this total was completed by the end of 2015. Though some gains were made by the end of 2017, only about 80 percent of the goal of 406,000 total archival pages had been scanned (Progress 2018). The ongoing issues with processing and preservation often prevented materials from being sent for digitization in a timely fashion. Because conservators and archivists often worked on an unpredictable schedule, there was no way of telling which collections materials would be ready to be sent to the on-site scanner. In addition, because materials were being removed from collections for preservation, boxes with missing folders could not be moved along for digitization. That often caused collections to be scanned out of order while the processing archivists were waiting for preservation to be completed to finish the tasks necessary to be sent to the scanners.

Other factors contributed to the slower-than-expected pace of archival scanning. For example, because the scanners had limited or no familiarity with Yiddish, Hebrew, and Cyrillic characters, YIVO decided to have the processing archivists tag each page with a paper flag, to provide scanning instructions for each document. Instructions included noting items in which there was writing in multiple directions on a single side to indicate the preferred or dominant orientation; duplicate items or any items that, for whatever reason, should not be scanned; and page orientation based on language. This ultimately proved time-consuming for both the archivists and the

scanners, as it took additional time to not only write out and affix the flags but also to remove and re-affix them before and after the scanning process (Scanning Instructions 2016).

Despite the issues with the digitization of the archival materials, during the first years of the project, the number of books prepared for digitization far exceeded initial estimates. By all accounts, the library portion of the project was proving a great success. By the end of the first year, over 712,000 book pages had been scanned, over five times the estimate of 126,000 pages YIVO had anticipated. By the end of 2017, the vendor had scanned almost 1.486 million pages, over three times as many pages as YIVO's plan accounted for (Progress 2018).

The exceptional pace at which the library materials were being scanned was predicated in part on the assumption that the library books needed very little physical preservation work before they were sent to the vendor for scanning. In fact, the master plan for the project did not have dedicated time built in for library preservation efforts. In addition, though it was acknowledged at the start of the project that many of the books to be scanned lacked sufficient intellectual control, only minimal work was done during the first years of the project to correct any cataloging issues. Rather, it seemed that these were to be left for a later stage of the project.

By the vendor contract, all post-production work on library books and archival items, including cropping and straightening, was part of the work plan. In July 2015, the digitization vendor informed YIVO that rather than having the scanning technicians perform post-production, they would subcontract this work out to a company based in China. Files were shipped to this company on hard drives that were then shipped back to the vendor who would deliver them to YIVO. Though the vendor assured YIVO that they had worked with this subcontractor for over a decade and were impressed with their work, the result of the post-production work turned out to be subpar in YIVO's estimation. There were several different errors including improper cropping, incorrect splitting of items that should have been shot as spreads, and mistakes in item sequencing. Many of these errors were due to the subcontractor's general lack of familiarity with both unpredictable archival sources as well as the foreign languages involved in the documents (Scanning Instructions 2016). It was discovered that approximately three percent of every digitized archival collection that had been scanned between 2015 and 2017 required rescans due to errors that occurred during the scanning and/ or post-production processes. The rescan work at the time demanded that YIVO staff tag each page with a physical flag and written instructions for the problems that needed to be corrected during the rescan process.

There were also major issues with the digital images taken of the books. Quality assurance had been done on 1,059 library books as of November 2017. Of these, 22 percent had problems including dark or blurry pages, files with bound books that needed to be separated, or pages that

were scanned multiple times. Another 27 percent of the scans had missing pages; some type of information obstruction; or improper orientation; and 3 percent had other issues that needed to be addressed via rescans. This amounted to problems with over 50 percent of this batch of books from the vendor. A report from July 19, 2018, indicated that of another 1,595 books that had gone through the quality assurance process, 51 percent had similar issues as the first batch, and needed some type of remediation. Quality assurance performed by a YIVO staff member on 578 books over the course of two years yielded 2,000 notes indicating the need for rescans of certain pages. By the end of 2018, that number had jumped to 661 books which went through quality assurance with 2,273 rescan notes (Book Rescanning 2018).

Although the original agreement stated that rescans would be done by the vendor at no additional charge, neither the vendor nor YIVO had considered the number of needed rescans. Once the number of essential rescans was realized, the vendor agreed to allow one scanner to spend 20 hours every other week on necessary rescans. That proved inadequate to complete the number of rescans necessary to keep the project moving forward, as ingesting the materials relied on complete collections. Without proper and timely rescans, ingest fell behind. YIVO eventually negotiated a new agreement with the vendor so that they would provide rescans for $34 per hour (Geller and Strykowski 2022, 3).

## Project Assessment and Restructure in New York, 2018–2020

A summary of activities prepared in early 2018 for the first three years of the project indicated that in almost every area, the project had fallen behind its intended goals (Progress 2018). For the archival portion of the project in New York, for example, only 40 percent of the materials had been processed and were ready for digitization, with less than 40 percent of materials having received conservation treatment. In addition, nearly 55 percent of the archival documents still needed to be digitized, and 83 percent still needed to go through quality assurance and be ingested (Progress 2018).

This realization, along with the addition of new project leadership, led to an overall assessment of the entire project in 2018. The 2018 assessment included further data gathering on the scope of the project and the amount of work completed, an evaluation of work processes and systems of reporting, and compiling accurate information about revenues and expenses in order to precisely calculate budgetary needs.

In addition to what had been revealed in the 2018 summary of activities in New York, the assessment soon exposed that page estimates made at the start of the project were inaccurate, with the initial numbers underestimating the scope of the archival portion by nearly 300,000 pages.

Moreover, though the summary of activities indicated that digitization numbers for the library portion of the project had exceeded projected expectations, the assessment revealed that only 2,897 volumes had gone through quality assurance and 2,274 had been put online (Strykowski to Brent 2019). That revelation left more than 70 percent of the books that had been scanned in New York inaccessible to researchers online and more than 80 percent of the total number of books scanned in Lithuania and New York inaccessible to researchers (Progress 2018; Strykowski to Brent 2019). It was also discovered that there were widespread cataloging and metadata issues that needed to be addressed before digital images of books could be properly linked to their catalog records.

By the conclusion of the assessment, it was discovered that not only was the project running behind schedule, but an additional $1.75 million would be needed to complete it. Though falling behind the projected timeline had something to do with this, the increase in the scope of materials that had been discovered at both the Wroblewski Library and the MMNLL would require additional staff time and resources in both Lithuania and New York. The new and final budget for the project was set at $7 million, which was raised from grants, foundations, and private donations. Though the projected timeline for the overall completion of the project was still set by the end of 2021, the timeline of certain aspects had to be expanded past their originally intended end dates to accommodate the additional work that needed to be done.

To help rectify the overall delay, the project was restructured. Additional project staff were hired, and all non-project staff were asked to no longer have any involvement in the project.[10] This structure and additional staff allowed tasks which had been assigned to interns and volunteers, and whose completion timeline was therefore unpredictable, to be reassigned to full-time project staff. Though it took some time to work out the exact makeup of the project team and to hire the staff necessary, by 2020 the final structure included the following 12 positions:

- A member of YIVO's archival team leadership who would act as project manager, devoting 40 percent of work time specifically to managing and tracking the overall timeline, workflows, and budget. As this person had managerial responsibilities for other archival projects, they were not expected to do any work on specific collections included in the project.

10. In addition to the small project staff devoted to the project, YIVO staff members whose salaries were only partially funded by the project or who were not connected to the project at all were regularly assigned project-related tasks while performing their regular duties. Some of these staff members were also taking the lead on higher-level decision-making for the project. While the expertise of non-project staff was valuable, no clear chain of command was followed. Project staff were often unclear on whose decision they were to defer to, while other decisions were being made in a group setting during prolonged meetings, which caused even the smallest decision to become a drawn-out process, further adding to the increase in the timeline.

- A preservation manager who, in addition to working directly on the treatment of collections, would spend approximately 5–10 percent of their time monitoring the work of a team of two project conservators, answering any questions, and meeting regularly with the project manager to discuss progress and any potential problems.

- An archival processing manager who, in addition to working directly on the arrangement and description of collections, would also spend approximately 5–10 percent monitoring the work of a team of three project archivists, answering any questions, and meeting regularly with the project manager to discuss progress and any potential problems.

- A digital preservation manager who, in addition to performing all aspects of digitization, would also spend approximately 10–15 percent of their time directing and monitoring the work of three digital lab specialists. This responsibility included tracking the number of pages undergoing various digitization stages and the number of TIFF files and final PDF files ingested. Like the other managers, the digital preservation manager would meet regularly with the project manager.

As this new structure was being worked out, YIVO emphasized hiring collections staff who either had or were pursuing an MLIS or a master of archival studies. YIVO still prioritized both content and language knowledge and privileged those candidates who also had degrees in a field related to Jewish history and culture and knowledge of at least one of the languages most prevalent in YIVO's collection. However, some of the challenges encountered thus far in the project led YIVO to believe that a certain degree of archival literacy was important. YIVO's goal during this restructuring was to assemble project staff based on a holistic approach, making sure that each member of the team understood intimately what everyone else was doing and could help with other aspects of the project if needed. The decision to reorganize the team proved prescient as the world shut down for the pandemic in March of 2020.

In order to create the necessary timelines and workflows, YIVO's conservation team was tasked with assessing each of the remaining collections. The lead conservator surveyed each of the archival collections that had not received any preservation treatment and ranked them on a difficulty scale of 1 to 3. Collections in the first rank could be conserved at a rate of 1 box per week, while collections in the second and third ranks could be conserved at a rate of 2 or at least 3 boxes per week, respectively. Using these rankings, a new timeline for the completion of each collection was created. It was also decided that rather than have collections processed before they received treatment, all preservation work on a collection would need to be completed before it moved onto processing. To ensure that the processing archivists would have enough material to work with, the new timeline was frontloaded with those collections that needed only limited preservation (rank 3). This way, a surplus of materials would be ready for processing, allowing the conservation team more time to work on collections which would take longer to treat without the fear that they were holding up the flow of materials.

While the conservators were assessing and ranking the collections, the processing archivists were asked to provide an estimated timeline for the complete arrangement and description of each collection, including the encoding of the finding aid. These assessments, coupled with the new conservation timeline, provided enough information to assign all remaining collections to the processing archivists. This timeline included dates the processing archivists could expect to receive collections from conservation, a date by which processing of the collection must be started, and a firm end date for when the collection would need to be passed on to the digitization team. To ensure the work stayed on track, each team was provided with a pacing schedule for the weekly number of linear feet needed to be completed. To monitor the progress, each team member was asked to submit daily tracking forms for the work that was completed. An Excel sheet that aggregated the daily reports and calculated the completion percentage of each stage facilitated the detection of pacing problems, which allowed for needed interventions, as well as determining what aspects of the project were proceeding ahead of schedule, which allowed staff members with additional time to pitch in with any task that might be falling behind.

Though initial project reports indicated that 100 percent of the books intended to be part of the project had been scanned based on the number of digital objects received by the vendor, the 2018 assessment revealed that these reports were not entirely accurate. Other than the work that had been done to determine which of the books listed in *Liḳute Shoshanim* existed in YIVO's collection, no master list of books flagged to be scanned had been created before the start of the project; instead, lists were being created during or after separate batches of books had been sent for scanning. It was therefore impossible to determine if "all" the books had been scanned, especially the scanned books that fell outside the scope of *Liḳute Shoshanim*, including widely available mass-published Yiddish books.

The 2018 assessment also demonstrated that during the first years of the project, difficulties arose in maintaining intellectual and physical control over library materials while they progressed from one project step to the next. The unresolved catalog records issues, including missing or duplicate records and incomplete or incorrect metadata, made matching digital objects to the correct record difficult. In addition, books that had been pulled for scanning from a variety of YIVO's library collections[11] were randomly reshelved: some were placed back in their previous location after scanning with a white paper flag indicating they had been scanned, while others were put away in separate and often unmarked shelf locations with no indication of a changed physical location on any of the project spreadsheets. These reshelving practices interrupted the quality assurance process since finding the physical books and comparing them with the digital objects proved to be challenging at times. These physical control and reporting problems also led to misunderstandings about what library materials had been scanned by the vendor.

11. For the history of the YIVO Library and its call numbering systems, see Abramowicz 1968; Baker 1995.

To gain a better understanding of where the library portion of the project stood, a master list based on Aleph records to which YIVO staff had chosen to add a "VCP" code in the 583 field (action note) was generated in June 2019. Checking this list against the digital files that had been received by the vendor, it was discovered in early 2020 that rather than the 8,300 books that YIVO believed had been scanned based on vendor reports, there were only 6,700 physical volumes that had been scanned. When considering bound volumes, this number amounted to 6,900 titles. Further investigation revealed that of those volumes which had been scanned, 86 volumes (17,729 files) were duplicate scans, with some volumes having as many as 28 different scans; 22 volumes (1,613 files) were scanned though they were outside the scope of the project; 147 volumes had been uploaded to YIVO's tertiary digital storage system multiple times; 75 books on the scanned items list were in fact not scanned; 32 scans were listed as having gone through quality assurance but the digital files could not be located; 39 titles were cataloged multiple times; 1,585 undigitized microfilm reels had been tagged as part of the Vilna Project in Aleph, although it was digital scans of physical volumes which were sometimes attached to the microfilm record by mistake, instead of the physical book in the library; records did not all clearly define "bound with" information; records contained incomplete holdings information; the catalog record did not always clearly indicate what YIVO Library collection books were from in case there were multiple copies of them; 901 books had been incorrectly cataloged; and many volumes contained duplicate records (Strykowski to Halpern 2020).

With this information in hand, the rest of the books that had been marked in the catalog record as having been intended to be part of the project were added to the digitization plan. This new information also allowed for the creation of an updated timeline and workflow to perform quality assurance. The new plan also covered the cleaning up and deduping of catalog records so that digital objects could be linked to the proper Aleph record. The project team also worked to locate each physical volume that had been or was supposed to be scanned, and a physically separate location with new shelf markings was prepared for these books. When the process of working with the book was fully complete, it was then reshelved in its original location. Finally, like the archival portion of the project, the library portion was given a pacing schedule and a reporting system to keep track of the progress of the remainder of the work.

In addition, YIVO had not been satisfied with its digitization vendor for some time at this point. The widespread need for rescans of both archival and library items due to mistakes made during the scanning and post-production process had added to the increase in both the timeline and budget of this portion of the project. It also became apparent that how digitization had been operating for the past three years was not only ineffective but also no longer sustainable. YIVO had been compelled to produce a certain number of archival pages each week for the digitization vendor to adhere to the terms of the contract and attempt to remain on track. As YIVO consistently fell short of meeting the agreed-upon page quotas, it was in breach of contract with the digitization

vendor, who as a result received payments that were far below what YIVO had agreed to. To meet the payment goals, a second contract was negotiated with the vendor in 2018. In order to maintain its yearly profit margins, the new contract had the vendor charging double the cost per page than had at first been negotiated. This, coupled with the additional hourly costs of rescans as well as the 20 percent on-site fee the vendor was charging, created budgetary implications that threatened to further increase the cost of the project for YIVO.

Shortly after the negotiation of this second contract, YIVO began investigating an avenue it had previously rejected: opening its own digital lab and bringing all work including capture, post-production, quality assurance, ingest, and quality control in-house. Developing an in-house digitization lab would mean not only that YIVO would have full control over workflows and timelines but also that individuals working there could be trained to do other steps in the digitization process. This would help to ensure that the work continued at the pace necessary to complete the project and that YIVO's team would no longer need to meet a quota for the digitization vendor. Based on a thorough cost analysis, YIVO determined that the total cost of continuing with an outside vendor for the duration of the project would be an additional $320,000, while the cost of creating an in-house digital lab would be approximately $215,000, inclusive of equipment and additional staff hires. Beyond the cost-saving measures, this would also ensure that the budget was predictable rather than what had come to be expected of the vendor: rescans and potential renegotiation of the contract should YIVO again fall short of the promised page quota (Geller and Strykowski 2022, 4).

YIVO canceled its contract with the digitization vendor in early 2019 and in March that year, it announced the opening of its newly created in-house digital lab. Following the advice of colleagues at the Center for Jewish History, YIVO purchased its first camera setup which included a copy stand with an 82-inch adjustable arm, 5400K fluorescent lights, and a DSLR camera body tethered to an iMac.[12] The digital lab staff members hired were not only experienced with the types of materials being digitized but also had some relevant language and content knowledge. Having staff with such background knowledge meant that the processing archivists would no longer need to tag collections for digitization and provide specific instructions to the cameraman, such as the orientation of the page.[13]

Having an in-house digital lab also meant that a single, unified team could rectify the problems that arose from the somewhat arbitrary distinction that had been set up at the start of the project between library and archival quality assurance, ingest, and post-ingest quality control. In es-

12. For more about the creation of YIVO's digital lab, see Geller and Strykowski 2022.

13. The only exception to this would be any document(s) where the sequence was contrary to how materials were arranged in the folder or if materials could be skipped because they were duplicates. See Geller to Halpern, Sklar, Podhorcer, Lutz 2019.

sence, the workflows for these materials were the same (Geller 2018). That way, in addition to a cohesive and communicative team, there would never be any downtime in work as there had been in the past. When the digital lab was waiting on capture or quality assurance of archival materials, for example, they could now turn their attention to ingest and metadata remediation of library materials. As a result, a similar pacing schedule as had been set up for conservation, processing, and library work was also implemented for digitization, quality assurance, ingest, and quality control in December 2018. By the first months of 2020, just before knowledge of the COVID-19 pandemic became widespread, the pace for each of the teams had been met or exceeded (Pace 2020). By all accounts, it looked as if the project would succeed in being completed by the end of 2021.

As work progressed with the project web portal it began encountering problems that project staff or vendor consultants could not easily overcome. One of the major hurdles was a general incompatibility between what YIVO wished the portal could do and the actual capabilities of Archon. For example, code issues rendered parts of the finding aids illegible from the front end when ingested into Archon and had to be manually corrected within Archon each time a finding aid was ingested. Archon also provided no way of knowing which version of a finding aid was the most recently uploaded. To determine which finding aid needed to be deleted if a corrected version was ingested, the archivist needed to click through and review both the new and the old versions (Archon 2019). There were also widespread problems with suppressing and tagging materials in the portal. Each time the code would need to be updated to correct one of these issues, new problems would arise with display on the front end.

YIVO had also pursued the idea of creating a separate photo and poster viewer to display jpeg images, in addition to the standard viewer that displayed the access PDFs of documents. The decision to create this separate viewer assumed that Archon could easily support item-level finding aids, as did the digital asset management system in use at the time (DigiTool). With the finding aids in DigiTool, each item could be described individually, and each link was added manually. However, YIVO had yet to successfully implement its vision of the portal during the first several years of the project. When it became clear that the portal would not be able to support the item-level cataloging of the nearly 10,000 photographs and posters as part of the project, it was decided to return to YIVO's previous system of folder-level description. That decision was made since the time involved in manually linking items for use only in DigiTool and not in the portal seemed not worth the staff time and cost involved. At the earliest planning stages, it was acknowledged that in many ways the functionality of the portal for the end user would be somewhat duplicative when it came to retrieving library records and collection-level records for archival materials. This understanding was a consequence of the fact that YIVO and the other partners at the Center for Jewish History used Primo, Aleph, and DigiTool, with plans to migrate the digital asset management system to Rosetta and employ ArchivesSpace as the finding aid

repository. The development of the Vilna portal would mean that YIVO would be responsible for the upkeep and maintenance of a separate system that in no way communicated with the infrastructure already in use.

The continued reliance on Archon and the *Guide to the YIVO Archives* to retrieve archival-level metadata was also problematic, for several reasons. The first was that Archon had not been supported since 2014, with the rollout of ArchivesSpace. The second was that the *Guide to the YIVO Archives* had been customized by consultants with whom YIVO continued to work in order to maintain the site. These developers therefore became a necessary part of the creation of the new portal as YIVO was unable to maintain the *Guide* on its own. The integration of the *Guide* into the portal meant that these consultants would need to continue to be paid in perpetuity to maintain not only the *Guide* but certain aspects of the portal as well. These pitfalls, in addition to the budgetary implications of needing to pay for long-term maintenance, led ultimately to the decision to stop any further development on the portal as it currently existed. In addition to Archon's obsolescence, the functionality of ArchivesSpace and its compatibility with Primo and the other systems in place at the Center for Jewish History opened new possibilities of discoverability and access that would allow YIVO a turnkey solution.

YIVO still desired to provide patrons with a simple way to search only those materials that were included in the Vilna Project. It was decided that YIVO would customize a Primo skin to pull only Vilna collection-level records and library items. That would provide patrons with a way to search both library records and collection-level descriptions of collections. The search functionality of ArchivesSpace would allow patrons to experience the folder-level search that the portal was attempting to create while still presenting materials within a hierarchical order. The Primo skin was customized with only a nominal cost and because it was fully integrated into the already existing Center for Jewish History systems, there would no longer be any need for separate upkeep and maintenance.

## Project Assessment in Lithuania, 2018–2020

In addition to the setbacks with the project in New York, various issues in Lithuania demanded attention from YIVO staff. Though YIVO staff made at least several trips each year to visit their partners in Lithuania, there had always been some issues in monitoring the work progress from afar. YIVO also experienced issues maintaining proper control over invoicing. By the end of 2016, for example, there had been no invoices produced by Lithuania for the first two years of the project. Invoicing and payment continued to be an issue through 2018, which created problems for YIVO in maintaining proper budgets and expenses. These delays also made it difficult for YIVO to accurately report on project activities to grantors and other funders.

Work on the Lithuanian library books was not proceeding apace. Though the MMNLL had created a list of possible books to be scanned during the planning phases of the project, there was no finalized master list for YIVO to check digital files against. The lack of a list detailing the books chosen to be scanned in New York and Lithuania meant that sometimes the same books were scanned in both institutions, often multiple times. There were also problems with the cataloging of the Lithuanian books. For example, 920 books that had been scanned in Lithuania and delivered to YIVO had no metadata or catalog record. A YIVO staff member was tasked with creating catalog records in addition to the work that already needed to be done to remediate the records for the New York books. Due to the scope of the work on cataloging the books from both Lithuania and New York, it was decided to hire an outside contractor to help complete these tasks.

Another concern was how each institution was creating metadata for the archival materials and delivering them to YIVO. Though there was an excellent team of experts in Lithuania working on the arrangement and description of the materials, there was no consistency in the way the three Lithuanian institutions were describing materials or structuring their metadata, resulting in widespread discoverability challenges. For that reason, it was decided that YIVO's archivists would review each of these digital images and create new descriptive metadata for each folder. That included creating finding aids that imposed an intellectual arrangement on the materials, as it was discovered that much of the material being scanned and sent from Lithuania had not necessarily been arranged in a way that reflected the provenance or original order of the materials (to the extent this could be recreated). Because of differences in descriptive practices, there was also no easily recognizable connection between the materials in Lithuania and those in New York. To reconnect these disparate materials, YIVO's archivists also eventually needed to add additional metadata that directly referenced the collections held by YIVO by their record group numbers. Originally this descriptive work was being assigned to a part-time volunteer, resulting in an unpredictable timeline for describing, proofreading, and encoding the finding aids (Report 2018). Following the 2018 assessment, that work was reassigned to full-time project staff, allowing for additional oversight of the workflow and timeline.

With the discovery of additional materials at the MMNLL, the scope of the archival portion of the project in Lithuania greatly increased. Not only did YIVO need to figure out funding for this increased work in Lithuania, but it also needed to determine a path forward for checking, correcting, and creating descriptive metadata for these new materials. To further complicate matters, YIVO received an NEH grant to work on these newly discovered materials. Because NEH funding could only be used for work done in the United States, it was decided that all arrangement and description would need to be done by YIVO's project staff. To maintain a proper timeline, YIVO and the MMNLL developed a plan for the two institutions: the MMNLL would perform conservation and digitization and provide folder titles only or broad descriptions of grouped materials. YIVO would then complete the descriptive work, the creation of a finding aid, and all

ingest and quality control on the digital files after they had been delivered to New York. The plan dictated that between December 2019 and January 2021, YIVO would need to receive all digital copies of the new materials from the MMNLL. To speed the process along, YIVO would also take responsibility for carrying out copyright clearance, a process which was often time-consuming for YIVO's Lithuanian colleagues. YIVO and MMNLL agreed that the materials would be delivered on a rolling basis over the 14 months with regular deliveries every quarter. That arrangement would allow YIVO's archivists to work through the materials over a year, fitting the descriptive work into their other project responsibilities.

## Work During the Covid-19 Pandemic and Project Completion

In early March 2020, YIVO was informed by building management that its offices would be closing on March 13. The Vilna Project team began making immediate plans for a prolonged absence from the building, which at that time, no one was certain as to how long it would be closed. During the week leading up to the closure, all materials that had been removed from the climate-controlled stacks for project work were returned. All individuals on the team were asked to bring their personal laptops so that VPN could be installed on them for remote access to YIVO's servers. A plan and timeline were also made to reassign project staff to activities that could be completed remotely. YIVO's preservation team was perhaps most affected by the shutdown, as none of their work could be completed remotely. However, because the team had some content and language knowledge, they were trained on tasks that the digital lab would have normally performed, including creating PDF access copies from master files, quality assurance of library materials, and post-ingest quality control of archival collections. YIVO's digital lab team was able to devote their time to post-production, ingest, and quality control work. Although no capture was being performed, the digital lab team was able to address any backlog that existed for these other processes and continued to make materials available to researchers according to the timeline produced after the project assessment. The processing archivists were tasked with the biographical and historical notes and scope and content notes for collections that had been processed with full container lists but did not have a full finding aid. In addition, the processing archivists were asked to check, correct, and create descriptive metadata for all materials that had been delivered to YIVO from the LCVA, the Wroblewski Library, and the MMNLL.

These Lithuanian institutions also lived through a shutdown beginning on March 16, 2020. Though not all digital images had been delivered before March 2020, it was decided that with the uncertainty of when normal activities could resume, the project timeline could not afford to wait until YIVO's colleagues in Lithuania returned to the office to complete the digitization. To help expedite the process, it was decided that they would provide only minimum metadata and arrangement and that the archivists in New York would enhance the physical description as needed. That allowed YIVO's Lithuanian partners to digitize materials at a quicker pace and

therefore send more materials to the YIVO staff. In order to facilitate the description of these materials, YIVO's processing archivists were given access to YIVO's staging server, where digital materials were located until being ingested into YIVO's digital asset management system. The processing archivists first went through and identified languages present in each folder as well as types of materials. Then, based on the language knowledge of each of the archivists, as well as their familiarity with other project collections, they created descriptive metadata for the materials using a shared Google Sheet, which allowed more than one archivist to simultaneously work on description. When descriptive metadata and an intellectual arrangement for the collections were completed and the full finding aid created, YIVO's digital lab team remotely ingested these materials and linked them with the finding aid.

Since access to YIVO's offices was cut off, the team needed to find new ways to communicate, pass along information to team members, and problem solve. All project members were asked to send reports twice a week documenting the work that they had completed. A full project report based on these individual reports was aggregated by project leadership and sent out weekly to YIVO project staff and YIVO administration. During weekly Zoom meetings for the entire project team, each team passed off information to members of other teams that directly related to their work. Individual team meetings were also held regularly to collaboratively address problems and workflows. Lab managers and project leadership also met weekly to discuss any issues that needed to be addressed.

By early July 2020, it became clear that the at-home work the Vilna Project team could do would soon come to an end. Though the processing tasks the team had completed over nearly four months had helped to keep the project on track, the project would not be able to continue without access to the physical materials. In consultation with the entire project staff, it was decided that the team would return to the office at the end of July, albeit in a modified way and with tasks reassigned. The entire team was broken up into two groups who would alternate going into the office to perform those tasks that could only be done in person: arrangement, conservation, capture, and quality control of digital images. All descriptive metadata and the creation of finding aids would be done remotely using the digital images of the collections. All post-production work, including color correction and cropping as well as all ingest and post-ingest quality control, would also be done remotely. In essence, the work that was done in person each week would sustain the team while they were working remotely.[14] To make the most of the time in the office, all team members were trained on various tasks so that they could help move the work along. For ex-

14. To keep all staff as safe as possible, not only did YIVO put into place a strict mask mandate, but it also ensured that individuals would be able to work in separate offices. In addition, YIVO closely monitored the infection and hospitalization rate in New York City, knowing that any staff members coming in were putting themselves at risk by having to take the subway daily. In late December 2020, with an increase in cases, YIVO decided to once again fully suspend in-person operations until infection rates went down. By this time, there was enough surplus in digital images that needed descriptive metadata, post-production work, ingest, and quality control, that YIVO's project team could once again be reassigned work and continue remotely until it was again safe to return to the office.

ample, the processing archivists learned how to digitize materials or help the preservation team with paper flattening and mending. Because the priority would be to move digitization along so that processing work could be done remotely, YIVO decided to purchase two additional camera setups, which greatly increased the overall digitization capacity. Between July 2020 and December 2021, over 400,000 pages were digitized and went through post-production and quality assurance, and nearly 500,000 pages were ingested. The newly implemented workflows allowed YIVO to maintain the pace it needed to keep the project on track and to successfully complete certain aspects of the work covered by large grants, including an NEH grant and an IMLS Save America's Treasures Grant.

The Edward Blank YIVO Vilna Online Collections Project officially ended the first week of January 2022. Press coverage in news outlets including the BBC, NPR, the New York Times, and CBS 60-Minutes helped tell the story of the unique history of YIVO and its collections and celebrate that major accomplishment.[15] Perhaps more importantly, coverage in these and other local outlets helped position YIVO's digitization project as a leading effort in a broader movement to make historical materials available via large scale preservation and access programs.

Though YIVO's team was proud of their accomplishment, they also knew that no digitization project done at this scale would end without its share of loose ends. There would be technical glitches and ongoing maintenance activities that would need to be performed. In the final weeks of the project, the team compiled a list of cleanup activities, which would be performed in the background. Over the next year, these activities covered checking library records for dead or incorrect links, scanning archival materials that had been removed for YIVO exhibitions and projects over the years and never returned to their proper collection, fixing access restrictions that prevented the delivery of digital objects via YIVO's discovery layer, and returning oversized materials to their proper location.

YIVO's partners in Lithuania digitized all materials that were considered to have belonged to YIVO prior to World War II. In the years since the completion of the project, these institutions have agreed to continue to digitize other prewar materials in their collections that are of Jewish origin. The conservation, processing, and digitization of these collections will be done at the expense of each respective institution, which has agreed to share digital images of these materials with YIVO.

In January 2022, YIVO began its next major initiative: the Jewish Labor and Political Archives Digitization Project, an estimated 3.5 million pages of archival documents acquired from the Bund Archives in 1992. Using the framework developed by YIVO's team and building on the

15. For press coverage, see https://yivo.org/2022.

successes and lessons learned during the Edward Blank YIVO Vilna Online Collections Project, YIVO is set to complete this major initiative in June 2030.

## Key Takeaways and Future Plans

After the completion of the Vilna Project, YIVO's team had a series of post-mortems to discuss the lessons that could be learned from that large-scale undertaking. One of the key takeaways was that YIVO should create a written set of guidelines that could serve as a road-map for planning other digitization projects. As this case study has demonstrated, many of the major missteps that took place during the Vilna Project stemmed from inadequate planning which could have been avoided had YIVO followed a prescribed framework for surveying collections and creating procedures, timelines, and budgets.

Though some aspects of the following planning guidelines are specific to the makeup of YIVO's team and its collections, the principles gleaned from this case study can apply to other institutions, particularly because there is no standard set of practices for planning these types of projects. The basic steps outlined below can be implemented when planning a digitization project and the challenges that YIVO encountered can serve as questions for consideration.[16]

**1.** Before a final timeline and budget for a project can be created, it is important to be given ample time to adequately survey the archival and library collections. Some of the initial funds that are raised for an overall project must be devoted to a planning phase in which the following steps take place.

**1.1** An initial survey is conducted of all archival materials noting location information, extent including oversized materials, types of materials, and languages present in each collection.

**1.2** Based on the extent and using an average of 2,100 pages per 5-inch archival box, a page count is determined for each separate record group included in the project (based on the average number of pages in a box in YIVO's collection)

**1.2.1** All information about each collection is gathered, scanned, and placed in a shared folder that each member of the project team can access. These documents include information on provenance; donor and donation information; and custodial history.

16. Since the end of the Vilna Project, YIVO has only embarked on digitization projects that have focused mainly on archival materials. As such, the guidelines for library materials have not yet been tested on any project.

**1.3** For library items, a complete list of all books is created noting call number, physical location, physical description including page count, condition information, bound-with information, and a link to the correct catalog record or another item identifier.

**1.4** The preservation team surveys all books and archival collections using a ranking system and documents this in a spreadsheet.

**1.4.1** The system for archival collections treatment is based on a 1–3 ranking system: Rank 1:1 box per week; Rank 2: 2 boxes per week; Rank 3: 3 boxes per week.

**1.4.2** The system for library materials treatment is based on a 1–4 rank system: Rank 1 (requiring extensive treatment): 2–3 books per week; Rank 2 (requiring moderate treatment): 5–7 books per week; Rank 3 (requiring light cleaning or straightening of pages only): 50–65 books per week; and Rank 4 (requiring only a visual inspection and occasional treatment on a small number of pages per book): 80–100 books per week.

**1.5** The archival processing team creates processing plans for each collection.

**1.5.1** Processing plans are assigned to team members based on the archivist's language skills and familiarity with the topic.

**1.5.2** Processing plans include the following data points:

- ◊ Record Group number
- ◊ Collection title
- ◊ Dates of the collection (estimated)
- ◊ Collection creator
- ◊ Extent, including item number and size of boxes
- ◊ Physical location
- ◊ Provenance
- ◊ Separated materials
- ◊ Restrictions
- ◊ Scope and content note
- ◊ Special formats
- ◊ Languages
- ◊ Current folder titles, description, arrangement
- ◊ Proposed arrangement
- ◊ Time needed for arrangement, description, and creation of full finding aid, including front matter

- ◊ Supply needs
- ◊ Preservation needs

**1.5.3** The processing team reviews processing plans together to ensure that each team member is aware of the specifics not only of the individual collections for which they are responsible but also of the overall project.

**1.6** The digital lab conducts a digitization survey on each archival collection and library item. These surveys allow the digital lab team to anticipate any special equipment considerations and decide which team member should work on a particular collection, based on the specifications of the materials.

**1.6.1** Digitization surveys for archival materials include the following data points:

- ◊ Record Group number
- ◊ Extent, including page count
- ◊ Physical condition
- ◊ Types of materials (i.e., correspondence, manuscripts, diaries, photographs, objects, sound recordings, etc.)
- ◊ Oversized materials
- ◊ Special considerations

**1.6.2** Digitization surveys for library items include the following information, much of which can be gleaned from the initial survey:

- ◊ Call number
- ◊ Bound-with information
- ◊ Link to the correct catalog record
- ◊ Physical location
- ◊ Page count
- ◊ Condition
- ◊ Special considerations

**2.** A timeline for each stage of the project is created based on the findings of the surveys and plans.

**2.1** The start date of each stage is staggered so that a surplus of materials is available before being passed along to the next team.

**2.1.1** For smaller projects that include only a single collection, the conservation of materials must be completed before processing begins. Processing is completed before digitization begins.

**2.1.2** For larger, multi-year projects that include multiple collections and library materials, conservation is fully planned and begins during the time in which processing archivists are surveying the collections. Thus, processing takes place only on collections that have already undergone preservation treatment. Digitization on materials begins at least six months to a year after processing begins. That allows the digital lab team to do capture, post-production, quality assurance, PDF creation, ingest, and post-ingest quality control on collections and library items without any gaps in work that would have been generated if they were waiting on the processing archivists.

**2.1.3** For library items, all catalog remediation must be performed before a book is sent to be digitized.

**3.** Extra time is built into each stage of the project to account for sick time, vacation days, holiday closures, and potential problems that may arise.

**4.** A full budget for the project is created only after a timeline for the entirety of the project has been completed, reviewed, and signed off on by all members of the team.

**5.** The project timeline must dictate what grant and foundation funding to apply for, rather than allowing grant opportunities that may fall outside the current project priority to upend the flow of the project.

## Sources

Abramowicz, Dina. 1968. "The YIVO Library."*Jewish Book Annual* 24: 87–102.

Baker, Zachary M. 1995. "The Yiddish Collections of the YIVO Library: Their History, Scope and Significance," *YIVO Annual* 22: 267– 69.

Brent, Jonathan. 2013. "The Last Books," *Jewish Ideas Daily,* 1. May 1, 2013. http://www.jewishideasdaily.com/6413/features/the-last-books/.

Geller, Devora and Strykowski, Jill. 2022. "Lessons in outsourcing: Creating in-housedigitization infrastructure in Jewish heritage organisations." *Journal of Digital Media Management* 11 no.3: 1–7.

Halpern, Stefanie. 2022. "A History of YIVO's Prewar Archival Collections from 1925 to2001." *Judaica Librarianship* 22: 20–40. https://doi.org/10.14263/22/2022/707.

Rabinowitz, Dan. 2018. *The Lost Library: The Legacy of Vilna's Strashun Library in the Aftermath of the Holocaust.* Lebanon: Brandeis University Press.

### YIVO Vilna Project, RG 100. YIVO Institute for Jewish Research

"Book Rescanning." [Book Rescanning 2018.] 13 November 2018. Box 3.

"Contract for Digitization." [Contract for Digitization 2015]. 8 July 2015. Box 2.

"Digitool Finding Aids." [Finding Aids 2015]. 27 March 2015. Box 1.

"New York Processing Priorities." [Processing Priorities 2015]. 25 March 2015. Box 1.

"Outline Proposal for YIVO Library Digitization Projects." [Outline Proposal.] No date. Box 2.

"Pace by Month." [Pace 2020]. 2020. Report, 8 April 2020. Box 3.

"Processing Priorities and Workflow." [Priorities and Workflow 2015]. 22 April 2015. Box 1.

"Saving Finding Aids in Archon." [Archon 2019]. 26 November 2019. Box 3.

"Scanning Instructions." [Scanning Instructions 2016]. 26 October 2016. Box 2.

"Stats for QA Books." [Stats for QA 2018.] 19 July 2018. Box 3.

"VCP Project Progress." [Progress 2018]. Feb 28, 2018. Box 3.

"Vilna Collections Digitization Workflow." [Digitization Workflow 2015]. December 9, 2015. Box 1.

"Vilna Project Archival Processing Workflow at a Glance." [At a Glance 2015.] 2012. March 27, 2015. Box 1.

"Website Search Portal." [Search Portal 2016.] 27 January 2016. Box 2.

"YIVO Launches International Project of Cultural and Historical Reclamation." [Reclamation 2014.] Press Release, 23 September 2014, https://vilnacollections.yivo.org/Project-Launch-Announcement.

"YIVO Vilna Collections Initiative Update." [Update 2015.] 2015. May 28, 2015. Box 1.

"The YIVO Vilna Collections Newsletter" [YIVO Vilna Collections Newsletter 2015], 2015, Box 1.

"YIVO Vilna Collections Project—Digitization options." [Digitization Options 2015]. 2015. April 2015. Box 1.

Geller, Devora. "Post Digitization Report." [Post Digitization 2018]. 2018. Memo, 21 June 2018. Box 3.

Geller, Devora to Halpern, Sklar, Podhorcer, Lutz. "New Digitization Workflows." [Geller to Halpern, Sklar, Podhorcher, Lutz 2019.] 2019. Email, 9 April 2019. Box 3.

Gudauskas, Renaldas to Jonathan Brent. [Gudauskas to Brent 2014]. 2014. Letter, 22 May 2014. Box 1.

Jarockis, Roman to H.E. Deborah A. McCarthy. [Jarockis to McCarthy 2015]. 2015. Letter, 21 April 2015. Box 1.

"New Discoveries Categories, 2019." [Discoveries 2019]. 2019. Box 1.

NEH Application. [NEH 2015]. 2015. https://www.neh.gov/sites/default/files/inline-files/NEH-FOIA-19-17-Responsive-Records.pdf.

Strykowski, Jill to Jonathan Brent. [Strykowski to Brent 2019.] 2019. Email, 7 March 2019. Box 3.

Strykowski, Jill to Stefanie Halpern. "Library Info." [Strykowski to Halpern 2020]. Email, 2 February 2020. Box 3.

YIVO and Lithuanian Central State Archives, "First Amendment to Memorandum of Understanding." [YIVO-LCVA 2015]. 2015. Agreement, 15 April 2015. Box 1.

YIVO and the Martynas Mazvydas National Library of Lithuania. "First Amendment to Memorandum of Understanding." [YIVO-MMNLL 2015]. 2015. Agreement, 23 April 2015. Box 1.

YIVO and the Martynas Mazvydas National Library of Lithuania. "Memorandum of Understanding." [YIVO MMNLL 2017]. 2017. Agreement, 18 June 2017. Box 1.

YIVO and the Martynas Mazvydas National Library of Lithuania. "Second Amendment of Understanding." [YIVO-MMNLL 2016]. 2016. Agreement, 16 June 2016. Box 1.

YIVO and the Martynas Mazvydas National Library of Lithuania. "Third Amendment of Understanding." [YIVO-MMNLL 2018]. 2018. Agreement, 27 March 2018. Box 1.

YIVO and the Wroblewski Library of the Lithuanian Academy of Sciences. "Memorandum of Understanding Relating to the Vilna Collections Project." [YIVO-Wroblewski 2018]. 2018. Agreement, 8 June 2018. Box 1.