“Resources in Yiddish Studies: Digital Collections”

by Zachary M. Baker

In *geveb: A Journal of Yiddish Studies* (October 2016)

For the online version of this article:

“RESOURCES IN YIDDISH STUDIES: DIGITAL COLLECTIONS”

Zachary M. Baker

Introduction: Emerging scholars have limited opportunities for systematic orientation in the research resources of Yiddish Studies. As Zachary Baker commented elsewhere, “information literacy” is something that graduate students and faculty are likely to attain informally and on their own; classroom training is generally not available for this purpose. That’s why we’re publishing this online bibliographical series devoted to research resources in Yiddish Studies. It builds upon a day-long workshop devoted to resources in Yiddish Studies, which Baker led in April 2015 at the University of California-Berkeley.

This research guide will be divided into the following units, to be published in installments, each of which will take the form of a stand-alone post:

1. “Meta”-resources – bibliographies, web gateways, online scholarship, indexes, library and archival resources, encyclopedias.
2. Full-text electronic resources in Yiddish Studies.
3. Yiddish linguistic scholarship, including dictionaries.
4. Yiddish literature and culture.
5. Bibliographies of imprints (by country or region).
6. Anti-Semitism and the Holocaust (Yiddish focus).

Each unit is accompanied by a brief introduction. Where warranted, entries include brief annotations.
Since the first installment of this Research Guide came out we have received suggestions for the inclusion of additional sources, which are always welcome. The bibliographer faces a Sisyphean task which can never be completed. Some of the suggested sources will be listed in a future section of addenda. Readers are invited to contribute suggestions to be included in addenda.
2. DIGITAL COLLECTIONS

HOW WE GOT HERE

Some years ago a miracle technology emerged, one that promised to revolutionize scholarship while also enabling libraries to free up valuable real estate. That technology was known as . . . microfilm and it spawned a revolution that took off in the 1940s and lasted for close to half a century. Microfilming enabled libraries and archives to preserve and share their collections in ways that could scarcely have been imagined previously. Regrettably, it also led libraries to dispose of major segments of their print holdings—old newspapers, especially—on a truly massive scale and often indiscriminately. (Nicholson Baker famously excoriated this practice in his book *Double Fold* [2001].) Many librarians – myself included – were implicated in this practice.

For reasons that scarcely require elaboration, the “user experience” (to employ librarians’ argot) of reading from microfilm reels or microfiche sheets is far from ideal. Still, microfilms continue to play an important role in the research process—and nowadays many machines for reading and copying from microfilms offer the capability of downloading page images as PDFs. And in defense of microfilms, let it be said that under proper storage conditions they have a shelf life that can be measured in centuries. This is not the case with most paper-based publications or documents dating from the mid-nineteenth century and later.

By comparison, the permanence of files generated by—and stored on—computers has yet to be demonstrated:

- “Bit rot”—deterioration of even a tiny amount of coding—can render an entire file as unreadable as a blank slate.
- Obsolescence is a fact of life, as hardware and software applications come and go. Moreover, rapid advances in platforms and interfaces make yesteryear’s state-of-the-art digital project look and behave like it was produced in the horse-and-buggy era. (Some libraries use sophisticated forensic techniques to decipher files that were originally composed in programs that have been completely superseded—Word Perfect, for example.)
- We tend to overlook the fact that digital technology is utterly dependent on the availability of abundant and reliable sources of electricity.

These caveats aside, in the digital era the research experience truly has undergone a sea change, given the ease and speed of access to so much of the literature. The ubiquity of digitized and born-digital resources bring information right to your desktop and portable devices—when in the past you’d have to travel to repositories far and wide (or sit in front of a microfilm reader in a library). The sheer convenience of digitized books, newspapers, manuscripts, images, and audiovisual media—amplified by websites, discussion forums, and social media—has been a tremendous boon to researchers. All of this has happened within the last quarter century or so—the historical blink of an eye.
Yiddish Studies scholarship has greatly benefited from these advances. Digital scholarship in Yiddish began with the Mendele discussion forum, an e-mail-based forum for announcements, questions and answers among scholars of Yiddish, which dates back to 1991. Early experiments such as the Yiddish Typewriter also pointed the way to future progress. The real game changer has been the development, during the past dozen years or so, of large corpora of digitized texts in Yiddish, especially the thousands of books digitized by Google, the Yiddish Book Center, and research libraries (including national libraries), along with—most recently—the rapidly growing online presence of newspapers and journals. Audiovisual content in Yiddish has also been brought online. And scholarly platforms such as In geveb itself have been added to the mix.

This section of the guide is devoted to digitized text and audio corpora that are now available to researchers in Yiddish Studies. A couple of observations should be made with respect to these online resources:

- **Copyright:** To reiterate what I wrote in the first section of the guide, much of what has been digitized is not fully accessible online. Google has scanned millions of volumes for its Google Book Search service, but due to copyright restrictions only a minority are accessible from cover to cover. Audio recordings, too, are not always fully accessible on the Internet and can sometimes be listened to only on the premises of host institutions. One notable exception to this super-cautious approach is the Internet Archive, which provides full access to the books and audiovisual content that it has digitized. The Internet Archive removes digital files only when copyright holders object to their having been put online. Content digitized by the Yiddish Book Center—close to 12,000 books and over 1,000 audio recordings—is accessible via the Internet Archive.

- **Searching:** Until recently, the experience of reading a digitized Yiddish text has offered the researcher the online equivalent of turning the crank of a microfilm reader. Essentially, it has remained a page-turning experience, albeit with some welcome enhancements—such as the ability to jump quickly to different sections of a book or to select a specific date in a newspaper run. Optical-character recognition (OCR), the key ingredient to indexing and searching texts, has come late to digitized Yiddish texts and is still quite imperfect (“dirty,” to use the geeky term), especially in comparison with OCR for texts in English and other Western European languages. Search capabilities have improved for Yiddish but remain far from optimal.

As OCR for Yiddish improves, it will become more useful not only for searching texts but also for digital-humanities applications—text mining in particular, or what literary scholars such as Franco Moretti refer to as distant reading. For example, the Digital Yiddish Theatre Project describes itself as “a research consortium dedicated to the application of digital humanities tools and methods to the study of Yiddish theatre and drama,” though on the
textual front, at least, the digital humanities focus remains largely aspirational. Although the group’s internal discussions (in which I have participated) have touched upon spatial data as presented and analyzed via Geographic Information Systems (GIS) technology; text mining has been mentioned in largely hypothetical terms. A key reason why this is the case is that Yiddish OCR isn’t quite “there” yet.

NATIONAL AND INTERNATIONAL WEB PORTALS

Europeana collections: Gateway to over 53 million (as of September 2016) “artworks, artefacts, books, videos and sounds from across Europe.” Using the keyword “Yiddish,” Europeana includes links to nearly 25,000 results in various formats—only 1,105 of which are freely accessible online (i.e., in the public domain), with an additional 213 items accessible for “limited re-use.” Of the freely accessible items, 922 are audio files—primarily music recordings from collections in France, especially the Bibliothèque Medem. The 23,477 items for which “no re-use” is available include descriptive information (metadata) but no direct online access to the digitized content. Related portal:

Judaica Europeana: “A network of archives, libraries and museums working together to integrate access to the most important collections of European Jewish heritage and make them discoverable to more people . . . inspired by the vision of Europeana . . .” Content digitized by the partners can be searched in Europeana.

The National Library of Israel – Digital Collections: The search interface is rather cumbersome. A global search in the database (September 2016), limited to Language-Yiddish, yielded 1,186 results, virtually all for sound recordings. However, these results provide only the descriptive metadata for the digitized recordings, which apparently are accessible only on site in Jerusalem. (See the “Audio” section, below, for sites that offer streaming of audio tracks.)

Lithuania, via ePaveldas: Web portal for the digital cultural heritage of Lithuania (“paveldas” is Lithuanian for “heritage”); includes almost 300 items partially or entirely in Yiddish: proclamations, newspapers, journals, and record labels. The texts do not appear to be searchable. Use Advanced Search and limit by Language jidiš. The site includes both Lithuanian- and English-language interfaces.

Poland, via polona.pl: Portal for digitized content from Poland, sponsored by the Biblioteka narodowa (National Library) in Warsaw. For links to Yiddish items that have been digitized, click on Biblioteka (library), go to the Obiekty (objects) tab, and then limit by język (language) jidysz (Yiddish). As of September 2016, there were results for 21,468 items. Various document types have been digitized, e.g., periodicals (czasopisma), books (książki), and printed ephemera (druki ulotne). Titles can be searched by name, though their contents do not appear to be searchable. The Kolekcje
(collections) tab includes one topic of specifically Yiddish content: *Literatura jidysz* (Yiddish literature), with 232 titles. There is also a tab for *Sztetl* (shtetl), with postcard images as the predominant format.

**Poland, Digital Libraries Federation, via FBC**: “The largest Polish data provider in Europeana,” with over 4 million digital objects from Polish cultural and research institutions, as of September 2016. A keyword search on the term *jidysz* (Yiddish) yielded close to 900 results for various formats and document types, though relatively few of these are for texts in Yiddish.

**NEWSPAPERS**

**Historical Jewish Press Project.** This is a joint project of the National Library of Israel and Tel-Aviv University. The collection has grown from a handful of nineteenth- and early twentieth-century Hebrew journals to over 115 titles in ten languages, from various countries, languages, and time periods. Texts are searchable, though the OCR is inadequate. Thirty-three Yiddish titles were online as of September 2016, among them these important publications:

- *Der fraynd* (St. Petersburg), 1903-1913.
- *Haynt* (Warsaw), 1908-1939.
- *Kol mevaser* (Odessa), 1862-1873.
- *Lebnsfragn* (Tel-Aviv), 1951-2011.
- *Literarishe bleter* (Warsaw), 1924-1939.
- *Der moment* (Warsaw), 1910-1939.
- *Der morgen zhurnal = The Jewish Morning Journal* (New York), 1906-1922.
- *Unzer ekspres* (Warsaw), 1928-1939.
- *Der yud* (Krakow), 1899-1902.

**Forverts/The Yiddish Daily Forward**

Founded in 1897, the *Forverts* now lives primarily on the web, though a print edition is published once a month. Among its regular features are book reviews, feuilletons, and original works of fiction and poetry (in the section *Penshaft—New Yiddish Writing*).

**BOOK COLLECTIONS**

Collections of digitized Yiddish books have proliferated in recent years. Some of these collections are aggregated by hosts such as Europeana or HathiTrust, while others are discoverable only on their home institutions’ websites.
Ale verk fun Sholem-Aleykhem: Raphael Finkel, a professor of computer science at the University of Kentucky, has overseen a number of digital initiatives in the field of Yiddish, including this edition of the collected works of Sholem Aleichem. Ale verk fun Sholem-Aleykhem is an extraordinarily versatile online collection, with a very clean interface (entirely in Yiddish). The texts can be read in the following modes: original (scanned from the print edition), Unicode (Yiddish characters for OCR), transliteration (YIVO standard), and gloss (English translations are provided by hovering the cursor over words in the text). Some texts also include links to recordings that were made at the Jewish Public Library of Montreal. In addition, there is a separate tab for searching the entire corpus, to which OCR has been applied.

Detskie knigi na idish iz fondov OLSAA [Yiddish children’s books from the collections of the Russian National Library]: This collection consists of about fifty titles, the vast majority of them published in the Soviet Union during the 1920s and 1930s. The often-colorful covers are reproduced in the bibliographical records, but online access to most or all of the texts is limited due to copyright restrictions. The interface is in Russian.

Fenno-Ugrica – Hebraica: “The Hebraica collection contains digitized works in Yiddish. It is a selection of literary works, which were published in Yiddish during the last decades of the Russian Empire and which the library received as deposit copies. In addition to literary works of the Hebraica collection of the National Library of Finland, the online collection contains digitized text books in Yiddish. The items were originally published in former Soviet Union during the 1920s and the 1930s.” It includes 169 monographs and sixteen serials from the National Library of Finland and the National Library of Russia.

Google Books: Google launched its massive book digitization project in 2004 and has scanned over 25 million book titles since then, among them several thousand books in Yiddish (the precise number is uncertain). Online access to most of these texts is limited due to copyright restrictions. Searching is via the familiar Google interface.

HathiTrust Digital Library: HathiTrust is a partnership of academic & research institutions, offering a collection of millions of titles digitized from libraries around the world.” Probably a majority of the digitized books in the HathiTrust corpus were scanned as part of the Google Books project. HathiTrust offers a more “library-like” search interface; books for which “full view” is permitted (meaning, they are in the public domain) use a different reading platform from their manifestations in Google Books. As of September 2016, only one-third of the approximately 5,000 Yiddish items in the HathiTrust database were accessible in full view.

HebrewBooks.org: Although this is primarily a corpus of religious works in Hebrew, it does include a sprinkling of Yiddish titles, including issues of Yivo-bleter from 1931 to 1942.
Internet Archive: “A non-profit library of millions of free books, movies, software, music, websites, and more.” Among these are the 11,000-plus books digitized by the Yiddish Book Center, and over 1,000 Yiddish lectures and Yiddish Talking Books from the Jewish Public Library of Montreal (also digitized by the Yiddish Book Center; described below). The search interface is rather cumbersome, to put it mildly.

Jiddische Drucke (Goethe-Universität, Frankfurt-am-Main): “This database contains nearly 800 very valuable Yiddish books printed in Hebrew letters in West, Central and East Europe. It is outstanding because of its variety of Yiddish dialects and themes and the impressive number of extremely rare books including several unique editions. Publication dates range from the middle of the sixteenth century to the beginning of the 20th century, the earliest print being a Hebrew Bible of 1560 from Cremona, followed by a print from Basel from 1583.” The digitized works can be downloaded as PDFs but their texts are not yet searchable.

Poland, via polona.pl: see National and international web portals (above).

Yiddish Book Center – Digital Library & Collections. The Yiddish Book Center’s digital collections include the following subdivisions:

Digital Yiddish Library: “The Steven Spielberg Digital Yiddish Library includes more than 11,000 Yiddish titles available to read online or download free of charge.” The texts are not yet OCR-searchable. The books are available for purchase from the Yiddish Book Center, in print-on-demand. They are also accessible online via the Internet Archive, albeit with its super-clunky search interface.

Yiddish Audio Books: “The Sami Rohr Library of Recorded Yiddish Books is a collection of Yiddish books and short stories read aloud by native speakers, recorded in the 1980s and ‘90s at the Jewish Public Library of Montreal.”

Archival Recordings: “The Frances Brandt Online Yiddish Audio Library contains lectures by and interviews with Yiddish writers, recorded at the Jewish Public Library of Montreal between [1951] and 2005.”

Oral Histories: “The Wexler Oral History Project is a growing collection of in-depth video interviews with people of all ages and backgrounds, whose stories offer a rich and complex chronicle of Jewish identity.”

Yizkor Books: “The David and Sylvia Steiner Yizkor Book Collection contains memorial books that document Jewish life before World War II, along with vivid firsthand accounts of the Holocaust and its aftermath.” The books are from the collections of the Dorot Jewish Division of the New York Public Library and are accessible online via NYPL’s Yizkor Books web page. Hard-copy reprints are available for purchase from the Yiddish Book Center.

Yiddish Children’s Literature: “The Noah Cotsen Library of Yiddish Children’s Literature comprises 800 titles, including works by major Yiddish writers and Yiddish translations of classics.”

YIVOlibrarybooks.org – Ryzman edition: This collection largely comprises Hebrew-language rabbinica from the YIVO Library’s holdings, but it also includes some
Yiddish titles, among them the YIVO publications *Yidishe shprakh* (1941-1984) and *Yivo-bleter* (1932-1962).

**AUDIO**

A number of Yiddish audio collections—both spoken word and musical—have been digitized in recent years as well. Some have already been mentioned in previous sections of this guide. As far as access is concerned, audio files face similar copyright restrictions as digitized print materials. Often, only the descriptive details (aka metadata) for the digital files are provided online. Content available via the Internet Archive offers the main exception to this approach.

**The Grosbard Project:** Hertz Grosbard (1892-1994), “actor, word artist, elocutionist, bard and reciter, often called the ‘Master or Maestro of the Jewish Word’,” was a practitioner of the vanished art of the *vort-kontsert* (“word-concert,” i.e., literary declamations and recitations). The Project includes biographical information on Grosbard and listings of his recordings. Unfortunately, based on a September 2016 spot check of the website the links to the recordings themselves appear to have been deactivated.

**Der leyenzal:** A small online repository of academic lectures in Yiddish on literary topics, by scholars who are currently active in the field.

**Milken Archive of Jewish Music,** Santa Monica, CA: Founded in 1990 by Lowell Milken, the archive now comprises “over 600 pieces of music by roughly 200 composers.” Exponents of different musical genres are represented in the Archive, among them numerous Yiddish theater performers and composers. The Archive provides tracks of selected recordings by these musicians, as well as music albums. The website also includes music videos, live performances, documentaries, interviews, oral histories, and background reading.

**Recorded Sound Archives** (formerly: Judaica Sound Archives), Florida Atlantic University, Boca Raton, FL: “Originally established in 2002 as a small project dedicated to the preservation of Jewish music, the RSA has matured into a robust digitization operation for all types of sound recordings . . . Not all performers or songs are made available due to copyright restrictions in these instances only a listing for the recording will appear or a [45-second-long] snippet can be heard.” FAU affiliates have complete access to all 99,000 audio tracks; other “educators, students, and serious researchers” may apply for access through the RSA Research Station. The RSA’s Judaic Collection “boasts one of the largest and most extensive collections of Judaic music in the world,” with recordings of Yiddish music very well represented in its holdings. Many of the audio tracks of these recordings are fully accessible online via the RSA.

**Robert and Molly Freedman Jewish Sound Archive,** Schoenberg Center for Electronic Text & Image (SCETI), University of Pennsylvania, Philadelphia, PA: “This
musical research library, international, and multi-lingual in scope, is a collection of approximately 5,300 Judaic sound recordings, in various formats . . . The three major Judaic languages, Yiddish, Hebrew and Ladino are well represented as well as translations in various languages . . . The satellite collections are some five hundred publications in which original text, translation, transliteration and melody line of the recorded songs and poems are available; a sheet music collection of some 1,000 pieces (with thousands of pieces of sheet music yet to be catalogued); and ephemera, over 1,300 items of newspaper and magazine articles, concert programs, images, playbills, song pamphlets and assorted memorabilia.” Sample recordings are streamed online.

As noted in the previous section, the Yiddish Book Center’s Digital Library & Collections include audio recordings in Yiddish. Click on: Yiddish Audio Books, Archival Recordings, or Oral Histories. Many of the recordings available through the Yiddish Book Center are also accessible via the Internet Archive. And, as previously mentioned, audio recordings are also available through Europeana and The National Library of Israel.