ARTICLE

Developing a Minimalist Multilingual Full-text Digital Library Solution for Disconnected Remote Library Partners

Todd Digby

ABSTRACT

The University of Florida (UF) George A. Smathers Libraries have been involved in a wide range of partnered digital collection projects throughout the years with a focus on collaborating with institutions across the Caribbean region. One of the countries that we have a number of digitization projects within is Cuba. One of these partnerships is with the library of the Temple Beth Shalom (Gran Sinagoga Bet Shalom) in Havana, Cuba. As part of this partnership, we have sent personnel over to Cuba to do onsite scanning and digitization of selected materials found within the institution. The digitized content from this project was brought back to UF and loaded into our University of Florida Digital Collections (UFDC) system. Because internet availability and low bandwidth are issues in Cuba, the Synagogue's ability to access the full-text digitized content residing on UFDC was an issue. The Synagogue also did not have a local digital library system to load the newly digitized content. To respond to this need we focused on providing a minimalist technology solution that was highly portable to meet their desire to conduct full-text searches within their library on their digitized content. This article will explore the solution that was developed using a USB flash drive loaded with a PortableApps version of Zotero loaded with multilingual OCR's documents.

ABOUT THE PARTNERSHIP

The University of Florida (UF), George A. Smathers Libraries have been involved in a wide range of partnered digital collection projects throughout the years with a focus on collaborating with institutions across the Caribbean region. UF has been involved with the Digital Library of the Caribbean (dLOC), which began in 2006, and the university is the technical home. The dLOC brings together collections from countries around the Caribbean in order to provide researchers with greater online access to these physically dispersed collections.[1] This partnership reflects common interests of preservation, access, accessibility, discovery, and content management.[2]

One of the countries that we have a number of digitization projects with is Cuba.[3] The Cuban Judaica collection comprises materials held in the library of the Temple Beth Shalom (Gran Sinagoga Bet Shalom) in Havana, Cuba. The synagogue library collection contains over 10,000 books. The origin of this collection started with Abraham Marcus Matterin, the founder of the La Agrupacion Cultural Hebreo Cubana cultural group, who first gathered and arranged the materials. In addition to Matterin's own works, the materials in the library include many rare Yiddish publications from the early 20th century, as well as little-known works produced in Cuba beginning in the 1930s. The Temple Beth Shalom library as a whole provides a complete snapshot of Cuban Jewish intellectual, cultural, religious, and political life as it evolved and progressed during the 20th century.[4]

Todd Digby (digby@ufl.edu) is Chair, Library Technology Services Department, and Associate University Librarian, George A. Smathers Libraries, University of Florida. © 2021.

https://doi.org/10.6017/ital.v40i4.13319

Among the rare publications included in the Smathers Libaries collections are *Habanera Lebn*, the main Cuban Jewish newspaper published between 1932 and1960; *Israelia*, a Spanish-language newspaper which circulated as a monthly during the 1950s; as well as many other Cuban Jewish publications. This collection will also provide access to the synagogue library's wealth of Jewish publications from other parts of the Caribbean and Latin America.

The synagogue library digitization project is a partnership between the George A. Smathers Libraries, the Isser and Price Library of Judaica under the auspices of its NEH Challenge Grant, la Comunidad Hebrea de Cuba, and la Biblioteca Nacional de Cuba José Martí.

THE DIGITIZATION PROCESS

As part of this partnership, graduate interns who were fluent in Spanish travelled to Cuba to do onsite scanning and digitization of selected materials found within the institution. The digitized content from this project was brought back to UF and loaded into our University of Florida Digital Collections (UFDC) system. This digitization process involved taking images in a high-resolution TIFF format and then creating the appropriate metadata to accompany these records for use once they were loaded into the digital library system.

An additional step of this process was developed by Fletcher Durant, UF's preservation librarian, who cut strips of colorful acid-free paper, which were placed in physical items to indicate that they were digitized. These paper flags were used to tell local synagogue users which items were digitized and available locally in the synagogue and more broadly on the internet.

These digital files were then transported back to the Digital Support Services group at the University of Florida with the returning personnel on USB hard drives, which were then appropriately scanned for viruses before extracting the digitized files. As part of the ingest process to UFDC we created derivative access files including JPGs, JPG2000, and thumbnail images from the high resolution. Additionally, these files are processed through an optical character recognition (OCR) process to generate full text searchable files. This OCR process involved a combination of using Adobe Acrobat or ABBYY FineReader. The unique aspect of the OCR process was the need to ensure multilingual character recognition that could recognize and generate full text files that may include Spanish, Hebrew, and English.

These scanned files, along with the derivatives and OCR files were then loaded and made publicly searchable with the UFDC system. These newly digitized files were then easily loaded into our UFDC system and made accessible to the wider internet audience around the world. Working with a partner in Cuba, however, presented additional challenges.

PARTNERSHIP CHALLENGES

Providing the Synagogue with access to their own content presented challenges that were not necessarily anticipated when the scanning project started. The Synagogue did not have a local digital library system that they could use to load and access the newly digitized content. We did provide the Synagogue with copies of all the digital files, but since this digitization effort focused on text-based printed materials, to make full use of this content access to the OCR files in a searchable format was the end goal. This in itself is not normally an issue since the digitized content would be loaded on UFDC systems and could be accessed online. Unfortunately, one challenge that presented itself when working with cultural heritage partners within Cuba is that limited technology infrastructure and internet connectivity can create issues in supporting the physical and digital initiatives needs of the project activities. With internet availability being intermittent and bandwidth speeds limited in Cuba, there needed to be creative ways to address the digitization work and subsequent file sharing needed to be developed.[5] Aside from the broader infrastructural challenges related to technology that are presented when working with Cuban partners, there are additional bureaucratic challenges presented with partnering projects in Cuba that can be in flux and change with changes in policy by either the US or Cuban governments.

These technological and political hurdles made our ability to offer ongoing remote support highly challenging. Additionally, there were barriers to how we could offer remote support because our respective IT technicians spoke either English or Spanish, but not both. The need for translation between languages is one that can be overcome, but does slow down the responsiveness. Also, translators may not be accustomed to translating technical jargon, which can further complicate providing support.

Given the challenges presented above we endeavored to provide the Synagogue with a solution that would provide a multilingual full-text search of the materials that had been scanned and put through the OCR process. Additional factors that influenced our planning, as discussed above, were the recognition that our Cuban partner did not have reliable internet and access to the content hosted in the US and that this solution would be run locally at the Synagogue so we would be able to provide minimal, if any, support in installing or supporting the system once it was deployed.

MINIMALIST COMPUTING SOLUTIONS

In the search for a solution, we were influenced by the concept of minimalist computing.[6] Borrowed from digital humanities, minimal computing "refer[s] to computing done under some set of significant constraints of hardware, software, education, network capacity, power, or other factors. Minimal computing includes both the maintenance, refurbishing, and use of machines to do DH work out of necessity."[7]

An important focus of minimal computing in the digital humanities, as noted by Risam and Edwards (2017), is how these practices can be used by those that find themselves with technological needs when they work outside larger macro structures of financial and technical support.[8] So basically, minimal computing is a solution for those individuals or institutions that are not positioned at a larger scale to support projects financially and technologically.

APPROPRIATE TECHNOLOGY

In addition to acknowledging the use of minimalist computing, additional related frameworks exist that we drew upon for our project. The most prominent of these is the concept of Appropriate Technology (AT) which is a concept that comes from the field of economics.[9] This concept was further adapted and used in the field of economic development and was seen through implementation pertaining to

> … small production units, appropriate technologies are small-scale but efficient, replicable in numerous units, readily operated, maintained and repaired, low-cost and accessible to low-income persons. In terms of the people who use or benefit from them, appropriate technologies seek to be compatible with local cultural and social environments.[10]

One of the main reasons for the use of appropriate technology is that advanced technologies were often inappropriate for the needs of the populations in countries that did not have the same level of technological infrastructure, support, and knowledge. This idea is composed of multiple facets, where in some cases it can be used to describe as using the simplest level of technology that is needed to meet the intended purpose of the user. It can also refer to how the system works in the way it is developed that takes into account the social and environmental factors for a given use.

OPEN-SOURCE APPROPRIATE TECHNOLOGY

Further influences that influenced the design of this project can be found in a more granular approach to appropriate technology that has become known as open source appropriate technology (OSAT). As defined by Pearce (2012), OSAT refers to technologies that can be sustainably developed, while at the same time being developed using the concepts and principles of free and open source software. Additionally, Pearce (2012) further states that,

> OSAT is made up of technologies that are easily and economically utilized from readily available resources by local communities to meet their needs and must meet the boundary conditions set by environmental, cultural, economic, and educational resource constraints of the local community.[11]

DEVELOPING A SOLUTION

As mentioned previously, the digitized Synagogue content from this project was brought back to UF from Cuba and loaded into our local Digital Collections system. This made the content accessible to anyone with internet access around the world, yet due to the fact that internet availability and low bandwidth are issues in Cuba, the Synagogue's ability to access the full-text digitized content residing on UFDC could not be assured. Additionally, the Synagogue also did not have a local digital library system to load the newly digitized content into.

To respond to their desire to conduct full-text searches on their digitized content within their library, we focused on providing a minimalist technology solution that was highly portable, user friendly, open source, and sustainable without any or minimal technical support.

In scanning the library technological landscape, our first thought was to find a small digital library system that could be used to meet these needs. Although there are a number of open source digital library systems and some of these can be configured to work in a non-internet–connected environment, the level of customization and ongoing technical support posed a problem, especially when we may not be able to provide support due to both issues in the telecommunications systems and possible language barriers. Although we knew that they had a Windows laptop, there were still technical uncertainties about the local computing environment within the Synagogue.

Once we determined that a full digital library system was not going to be sustainable and deployable, we decided to look for alternative approaches. A solution that just involved providing the scanned materials on DVDs was considered, but this also presented a problem, because the OCR'd PDFs would need to be opened individually to search the text within, and there was no logical way to provide citation information or organize files in a meaningful way.

ZOTERO PORTABLE

Eventually, we looked at the various citation management systems, since many of these allow PDF files to be imported and allow for searching the full text of PDF files using the OCR’d text. We then focused on a solution that is open source; this is especially important given that we were providing this to an entity in a foreign country and we did not want to experience any licensing or update issues to the chosen software. The platform that was chosen was Zotero, primarily because of the open source license of the software, but also because of existing technical knowledge and experience using it.

With Zotero chosen as our platform, we then investigated how this could be made portable in a way that was already installed, configured, and populated for the end user. I had existing knowledge of the PortableApps platform (https://portableapps.com/), which is a fully open source and free platform that enables a broad range of Windows applications to be installed on a portable device, in our case a large flash drive. Once installed using the PortableApps platform these applications can be used without any additional installation, just by plugging in the flash drive to another computer. In the case of the Zotero application, there was already a deployment that was created to be used with PortableApps, which made the installation and configuration less of a hurdle. See figure 1.

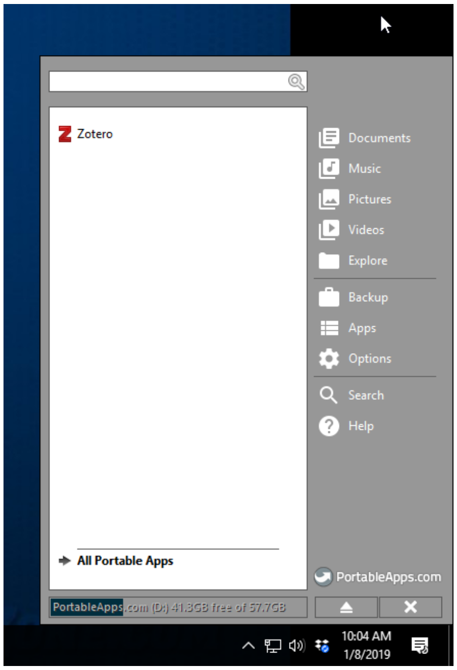
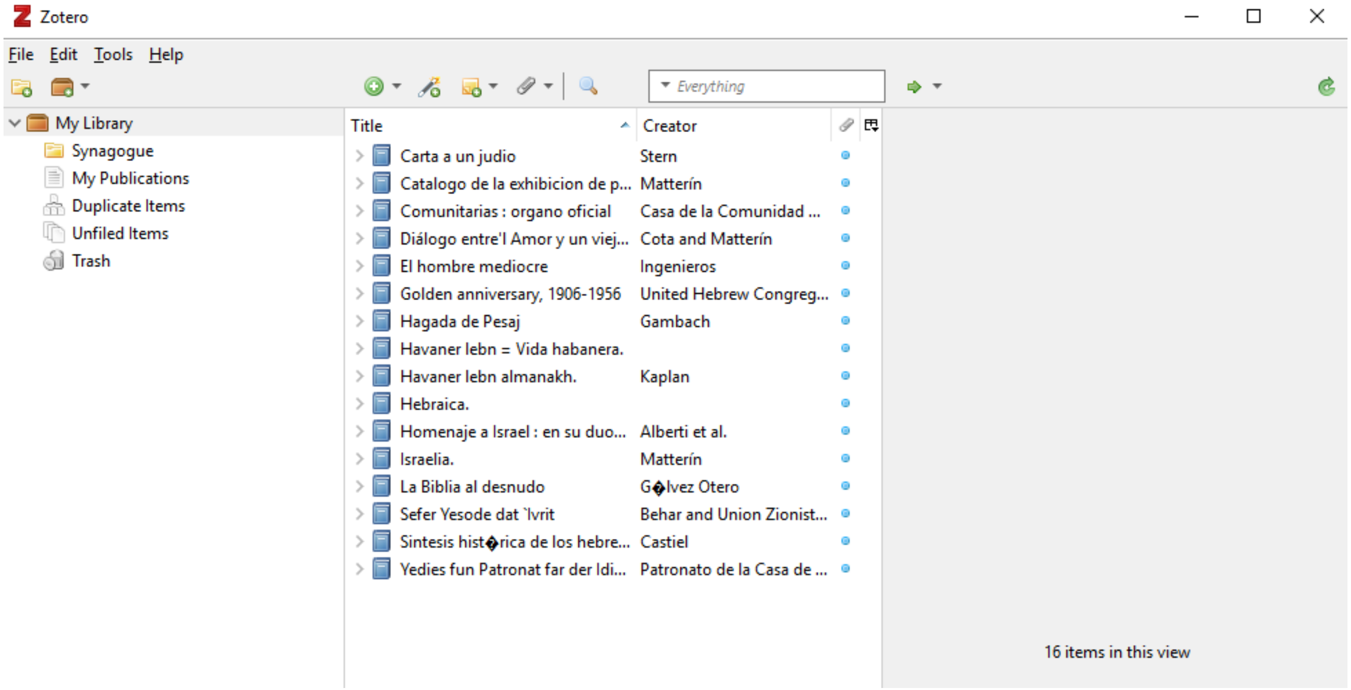


Figure 1. Once plugged into the PC, the PortableApps.com drive would present itself; double-clicking on the icon would load the menu for the user with only the Zotero application available to choose.

Once installed we started to load the digitized files into Zotero. This consisted of a two-step process. First, because the content was already added to our digital library system, we imported citations for each item or volume into Zotero, as you normally would when adding citations into Zotero. Then we went into each of these citations and added the corresponding digitized file(s) into each entry. Some of these items consisted of multiple volumes that would be placed under an existing citation. In addition to this, some of the materials contained both Spanish and Hebrew text, so during the OCR process there was a separate file created for each of these languages.

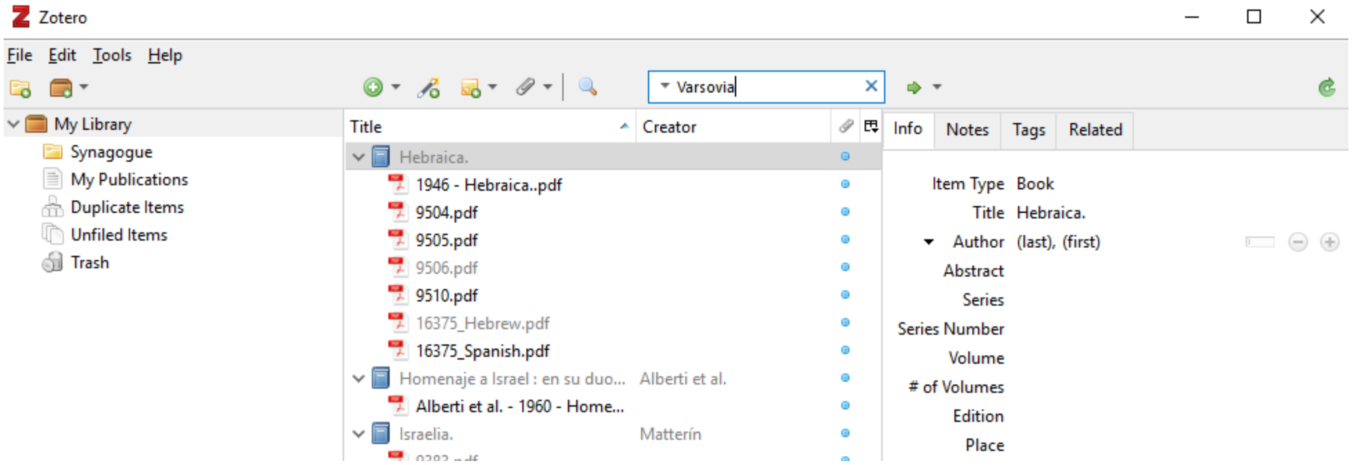
At this point we were able to test the full-text search capabilities of the Zotero against our multilanguage OCR'd PDF files.



Title	Creator
Carta a un judio	Stern
Catalogo de la exhibicion de p...	Matterín
Comunitarias : organo oficial	Casa de la Comunidad ...
Diálogo entre'l Amor y un viej...	Cota and Matterín
El hombre mediocre	Ingenieros
Golden anniversary, 1906-1956	United Hebrew Congreg...
Hagada de Pesaj	Gambach
Havaner lebn = Vida habanera.	
Havaner lebn almanakh.	Kaplan
Hebraica.	
Homenaje a Israel : en su duo...	Alberti et al.
Israelia.	Matterín
La Biblia al desnudo	G�lvez Otero
Sefer Yesode dat ʻIvrit	Behar and Union Zionist...
Sintesis hist�rica de los hebre...	Castiel
Yedies fun Patronat far der Idi...	Patronato de la Casa de ...

Figure 2. This image shows a set of citations of the scanned materials. To perform a full-text search within these documents a user uses the Everything search box in the Zotero toolbar.

At this point in our testing, we had determined that our method was successful in being able to perform a full-text search across all the loaded PDF documents (see figure 2). However, a limit of Zotero at this point was that the search would only identify which files the search terms were located in and not the exact location of the located terms within each file (see figure 3). Although this was a limitation of our solution, we were able to provide a second step for searching within the PDF files for the exact location of the search terms.

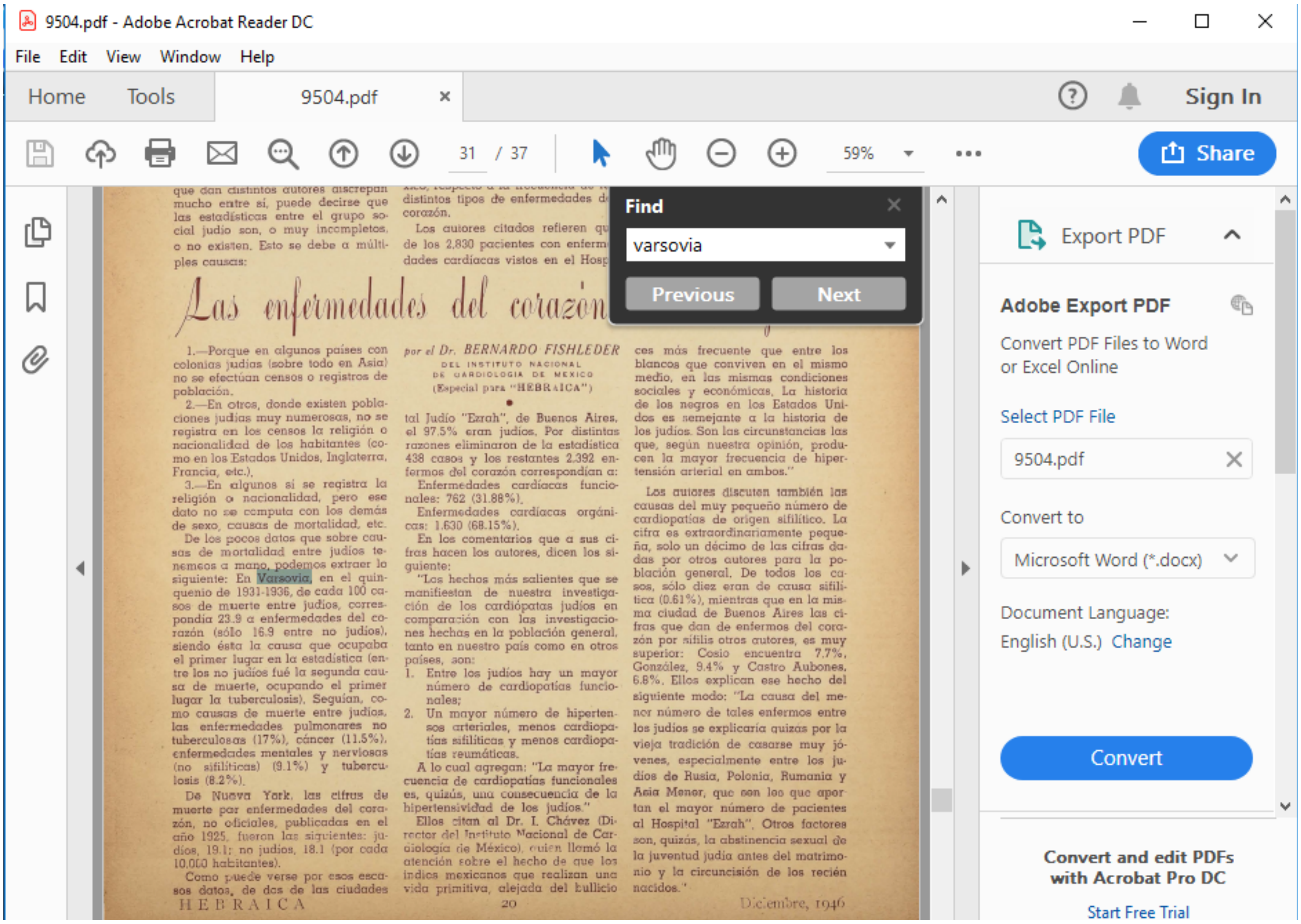


Search: Varsovia

Title	Creator
Hebraica.	
1946 - Hebraica..pdf	
9504.pdf	
9505.pdf	
9506.pdf	
9510.pdf	
16375_Hebrew.pdf	
16375_Spanish.pdf	
Homenaje a Israel : en su duo...	Alberti et al.
Alberti et al. - 1960 - Home...	
Israelia.	Matterín

Item Type	Book
Title	Hebraica.
Author	(last), (first)
Abstract	
Series	
Series Number	
Volume	
# of Volumes	
Edition	
Place	

Figure 3. The Zotero search results, which highlighted the files in which the search text could be found.



Find: varsovia

Select PDF File: 9504.pdf

Convert to: Microsoft Word (*.docx)

Document Language: English (U.S.)

que dan distintos autores discrepan mucho entre si, puede decirse que las estadísticas entre el grupo social judio son, o muy incompletos, o no existen. Esto se debe a múltiples causas:

Las enfermedades del corazón

por el Dr. BERNARDO FISHLEDER
DEL INSTITUTO NACIONAL
DE CARDIOLOGIA DE MEXICO
(Especial para "HEBRAICA")

1.—Porque en algunos países con colonias judias (sobre todo en Asia) no se efectúan censos o registros de población.

2.—En otros, donde existen poblaciones judias muy numerosas, no se registra en los censos la religión o nacionalidad de los habitantes (como en los Estados Unidos, Inglaterra, Francia, etc.),

3.—En algunos si se registra la religión o nacionalidad, pero ese dato no se computa con los demás de sexo, causas de mortalidad, etc.

De los pocos datos que sobre causas de mortalidad entre judíos tenemeos a mano, podemos extraer lo siguiente: En Varsovia, en el quinquenio de 1931-1936, de cada 100 casos de muerte entre judios, correspondia 23.9 a enfermedades del corazón (sólo 16.9 entre no judíos), siendo ésta la causa que ocupaba el primer lugar en la estadística (entre los no judíos fué la segunda causa de muerte, ocupando el primer lugar la tuberculosis). Seguían, como causas de muerte entre judíos, las enfermedades pulmonares no tuberculosas (17%), cáncer (11.5%), enfermedades mentales y nerviosas (no sifilíticas) (9.1%) y tuberculosis (8.2%).

De Nueva York, las cifras de muerte por enfermedades del corazón, no oficiales, publicadas en el año 1925, fueron las siguientes: judíos, 19.1; no judíos, 18.1 (por cada 10.000 habitantes).

Como puede verse por esos escasos datos, de dos de las ciudades

distintos tipos de enfermedades d corazón.

Los autores citados refieren qu de los 2,830 pacientes con enferm dades cardiacas vistos en el Hosp tal Judío "Ezrah", de Buenos Aires, el 97.5% eran judíos. Por distintas razones eliminaron de la estadística 438 casos y los restantes 2.392 enfermos del corazón correspondían a:

Enfermedades cardíacas funcionales: 762 (31.88%).

Enfermedades cardíacas orgánicas: 1.630 (68.15%).

En los comentarios que a sus cifras hacen los autores, dicen los siguiente:

"Los hechos más salientes que se manifiestan de nuestra investigación de los cardiópatas judíos en comparación con las investigaciones hechas en la población general, tanto en nuestro país como en otros países, son:

1. Entre los judíos hay un mayor número de cardiopatías funcionales;
2. Un mayor número de hipertensos arteriales, menos cardiopatías sifilíticas y menos cardiopatías reumáticas.

A lo cual agregan: "La mayor frecuencia de cardiopatías funcionales es, quizús, una consecuencia de la hipertensividad de los judíos."

Ellos citan al Dr. I. Chávez (Director del Instituto Nacional de Cardiología de México), quien llamó la atención sobre el hecho de que los indios mexicanos que realizan una vida primitiva, alejada del bullicio

ces más frecuente que entre los blancos que conviven en el mismo medio, en las mismas condiciones sociales y económicas. La historia de los negros en los Estados Unidos es semejante a la historia de los judíos. Son las circunstancias las que, según nuestra opinión, producen la mayor frecuencia de hipertensión arterial en ambos."

Los autores discuten también las causas del muy pequeño número de cardiopatías de origen sifilítico. La cifra es extraordinariamente pequeña, solo un décimo de las cifras dadas por otros autores para la población general. De todos los casos, sólo diez eran de causa sifilítica (0.61%), mientras que en la misma ciudad de Buenos Aires las cifras que dan de enfermos del corazón por sifilis otros autores, es muy superior: Cosio encuentra 7.7%, González, 9.4% y Castro Aubones, 6.8%. Ellos explican ese hecho del siguiente modo: "La causa del menor número de tales enfermos entre los judios se explicaría quizás por la vieja tradición de casarse muy jóvenes, especialmente entre los judíos de Rusia, Polonia, Rumania y Asia Menor, que son los que aportan el mayor número de pacientes al Hospital "Ezrah". Otros factores son, quizás, la abstinencia sexual de la juventud judía antes del matrimonio y la circuncisión de los recién nacidos."

HEBRAICA 20 Diciembre, 1946

Figure 4. Acrobat search conducted on a file to locate the exact location of the text in the digitized materials.

Since we were aware that you can search within a full-text PDF file using Adobe Acrobat Reader, we decided that for a more granular search we would instruct the users to click on the file that included the identified search term, which would load Acrobat Reader and open the file. Then the user could search for the exact term using Adobe Acrobat Reader's search capabilities to locate the location in the document where the term was situated (see figure 4 for an example).

Although this two-step process is not ideal, for a minimal technological solution that addresses all the concerns, it would meet the overall goals of the project and provide a workable searching solution for our partner,

With the workflows, installation, and configuration of the flash drive complete, we next created documentation in both English and Spanish to guide the user through the search process. We then provided the flash drive with accompanying documentation to the next staff member who was travelling to Cuba.

Because of federal rules implemented shortly after we transported the flash drive to Cuba, our ability to travel to Cuba to work with our partners was limited. This has substantially reduced the information flow between our institution and our Cuban partners and has limited how much we can know about how actively this resource is being used. It is hoped that in the near future we can once again be able to travel to Cuba and re-engage with our partners to determine the success of this and other projects we have been working with them on.

CONCLUSION

In the realm of library technology, we often implement and support complex and highly costly systems as part of our regular oversight. By working on this project, we have been given a chance to take a step back and design a solution that uses open source and free tools that are readily available and require low support. Looking broadly across the technology platforms, systems, and software that are used, there is a tendency to find these include a plethora of features and functions that are rarely used but add additional complexity. Focusing on solutions that reduce this complexity and still meet user needs has been a rewarding experience.

ENDNOTES

[1] Brooke Wooldridge, Laurie Taylor, and Mark Sullivan, "Managing an Open Access, Multi-Institutional, International Digital Library: The Digital Library of the Caribbean," *Resource Sharing & Information Networks* 20, no. 1–2 (2009): 35–44, https://doi.org/10.1080/07377790903014534.

[2] Miguel Asencio, "Collaborating for Success: The Digital Library of the Caribbean," *Journal of Library Administration* 57, no. 7 (2017): 818–25, https://doi.org/10.1080/01930826.2017.1362902.

[3] "Celebrating Cuba! Collaborative Digital Collections of Cuban Patrimony," University of Florida Digital Collections, accessed February 15, 2021, https://ufdc.ufl.edu/cuba.

[4] "Cuban Judaica," University of Florida Digital Collections, accessed February 15, 2021, https://ufdc.ufl.edu/cuban_judaica.

[5] Xuefei Deng, Nancy Armando Camacho, and Larry Press, "How Do Cubans Use Internet? The Effects of Capital," in *Proceedings of the 52nd Hawaii International Conference on System Sciences* (2019), https://doi.org/10.24251/HICSS.2019.617.

[6] Jentery Sayers, "Minimal Definitions," Minimal Computing: A Working Group of GO::DH, October 2, 2016, https://go-dh.github.io/mincomp/thoughts/2016/10/02/minimal-definitions.

[7] "About: What Is Minimal Computing?" Minimal Computing: A Working Group of GO::DH, https://go-dh.github.io/mincomp/about/.

[8] Roopika Risam and Susan Edwards, "Micro DH: Digital Humanities at the Small Scale," *Digital Humanities 2017,* http://works.bepress.com/roopika-risam/27/.

[9] Ernest F. Schumacher, *Small Is Beautiful: Economics as if People Mattered* (London: Blond and Brigggs, 1973).

[10] Peter Thormann, "Proposal for a Program in Appropriate Technology," in *Appropriate Technologies for Third World Development* (New York: St. Martin's Press, 1979): 280–99.

[11] J. M. Pearce, "The Case for Open Source Appropriate Technology," *Environment, Development and Sustainability* 14 (2012): 425–31, https://doi.org/10.1007/s10668-012-9337-9.