

Building an Open Source Institutional Repository at a Small Law School Library: Is it Realistic or Unattainable?

Digital preservation activities among law libraries have largely been limited by a lack of funding, staffing and expertise. Most law school libraries that have already implemented an Institutional Repository (IR) chose proprietary platforms because they are easy to set up, customize, and maintain with the technical and development support they provide. The Texas Tech University School of Law Digital Repository is one of the few law school repositories in the nation that is built on the DSpace open source platform.¹ The repository is the law school's first institutional repository in history. It was designed to collect, preserve, share and promote the law school's digital materials, including research and scholarship of the law faculty and students, institutional history, and law-related resources. In addition, the repository also serves as a dark archive to house internal records.

In this article, the author describes the process of building the digital repository from scratch including hardware and software, customization, collection development, marketing and outreach, and future projects. Although the development

of the repository is ongoing; it is valuable to share the experience with other institutions who wish to set up an institutional repository of their own and also add to the knowledge-base of IR development.

Institutional Repository from the Ground Up

Unlike most large university libraries, law school libraries are usually behind on digital initiative activities because of smaller budgets, lack of staff, and fewer resources. Although institutional repositories have already become a trend for large university libraries, it still appears to be a new concept for many law school libraries.

At the beginning of 2009, I was hired as the digital information management librarian to develop a digital repository for the law school library. When I arrived at Texas Tech University Law library, there was no institutional repository implemented.

There were very few digital projects done at the law library. One digital collection was of faculty scholarship. This collection was displayed on a webpage with links to PDF files. Another digital project, to digitize and provide access to the Texas governor executive orders found in the *Texas Register*, was planned then disbanded because of the previous employee leaving the position.

I started by looking at the digitization equipment in the library. The equipment was very limited: a very old and rarely used book scanner and a sheet-fed scanner. The good thing was that the library did have extra PCs to serve as workstations. I did research on the book scanner we had and also consulted colleagues I met at various digital library conferences about it. Because the model is very outdated and has been discontinued by the vendor and thus had little value to our digitization project, I decided to get rid of the scanner. I then proposed to purchase an EPSON Perfection

V700 flatbed scanner, which was recommended by many digitization best practices in Texas. For software, we had all the important basics such as OCR and image editing software for the project to start.

For the following several months, I did extensive research on what digital asset management platform would be the best solution for the law library. We had options to continue displaying the digital collections through webpages or use a digital asset management platform that would provide long-term preservation as well as retrieval functions. We made the decision to go with the latter.

Generally speaking, there are two types of digital asset management platforms: proprietary and open source. In some rare occasions, a library chooses to develop its own system and not to use either type of the platforms if the library has designated programmers. There are pros and cons to both proprietary and open source platforms. Although setting up the repository is fairly quick and easy on a proprietary platform, it can be very expensive to pay annual fees for hosting and using the service. For the open source software, it may appear to be "free" up front; however, installing and customizing the repository can be very time consuming and these solutions often lack technical and development support. There is no uniform rule for choosing a platform. It depends on what the organization wants to achieve and its own unique circumstances.

I explored several popular proprietary platforms such as CONTENTdm and Digital Commons. CONTENTdm is an OCLC product, which has a lot of capability and is especially good for displaying image collections. Digital Commons is owned

Fang Wang (fang.wang@ttu.edu) is Digital Information Management Librarian, Texas Tech University School of Law Library, Lubbock, Texas.

by Berkley Press and is often used in the law library community. As a smaller law library, our budget did not allow us to purchase those platforms, which require annual fees of more than \$10,000. So we had to look at the open source options.

For the open source platforms, I investigated DSpace, Fedora, EPrints and Green Stone. DSpace is a Java-based system developed by MIT and HP labs. It offers a communities-collections model and has built-in submission workflows and long-term preservation function. It can be installed “out of the box” and is easy to use. It has been widely adopted as institutional repository software in the United States and worldwide. Fedora was also developed in the United States. It is more of a back-end software with no web-based administration tools and requires a lot of programming effort. Similar to DSpace, EPrints is another easy to set up and use IR software developed in the U.K. It is written in Perl and is more widespread in Europe. Greenstone is a tool developed in New Zealand for building and distributing digital library collections. It provides interfaces in 35 languages so it has many international users.

When choosing an IR platform, it is not a question of which software is superior to others but rather which is more appropriate for the purpose and the content of the repository. Our goal was to find a platform that had low costs and did not involve much programming. We also wanted a system that was capable of archiving digital items in various formats for the long term, flexible for data migration, had a widely accepted metadata scheme, decent search capability, and was easy to use. Another factor we had to consider was the user base. Because open source software relies on the user themselves for technical support for the most part, we wanted a software that had an active user community in the United States. DSpace seemed to satisfy all of our needs. Also, according to repository

66, the majority of the repositories worldwide were created using the DSpace platform.²

For the installation, we looked at the opportunity to use services provided by the state digital library consortium Texas Digital Library (TDL) and tried to pursue a partnership with the main university library, which had already implemented a digital repository. However, because of financial reasons and separate budgets, those approaches did not work out. So we decided to have our own IT department install DSpace.

Installation and Customization of Our DSpace

Unlike large university libraries, smaller special libraries face many challenges while trying to establish an open source repository. After making the decision to use DSpace, the first challenge we faced was the installation. DSpace runs on PostgreSQL or Oracle and requires a server installation. Customizing the web interface requires either the JSUI (JavaServer Pages user interface) or XMLUI (Extensible Markup Language user interface). The staff in our IT department knew little about DSpace. However, another special library on campus offered their installation notes to our system administrator because they just installed DSpace.

Although DSpace runs on a variety of operating systems, we purchased Red Hat enterprise Linux after some testing because it is the recommended OS for DSpace. Then our system administrator spent several months trying to figure out how to install the software in addition to his existing projects. Because we did not have dedicated IT personnel working on the installation, the work was often interrupted and very difficult to complete. Our IT staff also found it very difficult to continue with the installation because the software requires a lot of expertise.

Two months later, we discovered that a preconfigured application called JumpBox for DSpace was released and approved to be a much easier solution for the installation. The price was reasonable too, \$149 a year (the price has jumped quite a bit since then). However, using JumpBox would leave our newly purchased Red Hat Linux server of no use because JumpBox runs on Ubuntu, therefore after some discussion we decided not to pursue it.

We were a little stuck in the installation process. Outsourcing the installation seemed to be a feasible solution for us at this point. We identified a reputable DSpace service provider after doing extensive research including comparing vendors, obtaining references, and pursuing other avenues. After obtaining a quote, we were quite satisfied with the price and decided to contract with the vendor. While waiting for the contract to be approved by the university contracting office, I began designing the look and feel that is unique to the TTU School of Law with some help from another library staff member. The installation finally took place at the beginning of January 2010. I worked very closely with the service provider during the installation to ensure the desired configuration for our DSpace instance. Our repository site with the TTU Law branding became accessible to the public three days later. And with several weeks of warranty, we were able to adjust several configurations including display thumbnails for images. Overall, we are very pleased with the results. After the installation, our IT department maintains the DSpace site and we host all the content on our own server.

Collection Development of the IR

Content is the most critical element to an institutional repository. While we were waiting for our IT department

to install DSpace, we prepared and scanned two collections: the "Texas Governor Executive Orders" collection and the "Professor J. Hadley Edgar's Personal Papers" collection. The latter was a collection donated by Professor Edgar's wife after he passed away in 2009. Professor Edgar taught at the Law School from 1971 to 1991. He was named the Robert H. Bean Professor of Law and was twice voted by the student body as the Outstanding Law Professor. The collection contains personal correspondence, photos, newspaper clippings, certificates, and other materials. Many of the items have a high historic value to the law school. For the scanning standards, we used 200 dpi for text-based materials and 400 dpi for pictures. We chose PDF as our production file format as it is a common document format and smaller in size to download.

After the installation was completed at the beginning of January, I drafted and implemented a digital repository collection development policy shortly after to ensure proper procedures and guidance of the repository development. The policy includes elements such as the purpose of the repository, scope of the collections, selection criteria and responsibilities, editorial rights, and how to handle challenges and withdrawals. I also developed a repository release form to obtain permissions from donors and authors to ensure open access for the materials in the repository. Twelve collections were initially planned for the repository: "Faculty Scholarship," "Personal Manuscripts," "Texas Governor Executive Orders," "Law School History," "Law Library History," "Regional Legal History," "Law Student Works," "Audio/Video Collection," "Dark Archive," "Electronic Journals," "Conference, Colloquium and Symposium," and "Lectures and Presentations." There will be changes to the collections in the future as the digital repository collection development policy will be

reviewed each year.

Based on the collection development policy, we made a decision to migrate the content of the old "Faculty Scholarship" collection from webpages into the digital repository. It was intended to include all publications of the Texas Tech law school faculty in the collection. We then hired a second-year law student as the digital project assistant and trained him on scanning, editing, and OCR-ing PDF files; uploading files to DSpace; and creating basic metadata. We also brought another two student assistants on board to help with the migration of the faculty scholarship collection. The faculty services librarian checked the copyright with faculty members and publishers while I (the digital information management librarian) served as the repository manager handling more complicated metadata creation, performing quality control over student submissions, and overseeing the whole project.

Later Development and Promoting the IR

During the faculty scholarship migration process, we discovered a need to customize DSpace to allow active URLs for publications. We wanted all the articles linked to three widely used legal databases: Westlaw, LexisNexis, and Hein Online. Because the default DSpace system does not support active URLs, it requires some programming effort to make the system detect a particular metadata field then render it as a clickable link.

We outsourced the development to the same service provider who installed DSpace for us. The results were very satisfying. The vendor customized the system to allow active URLs and displayed the links as clickable icons for each legal database.

In April 2010, "Professor J. Hadley Edgar's Personal Papers" collection was made available in conjunction with his memorial service,

hosted by the Lubbock County Bar Association at the TTU Law School. We made the initial announcement to the law faculty and staff and later to the Lubbock County Bar about the new digital initiative service we have established. We received very positive feedback from the law community. Professor Edgar's family was delighted to see his collection made available to the public.

Following the success of the initial launch, I developed an outreach plan to promote the digital repository. To make the repository site more visible, several efforts were made: the repository site URL was submitted to the DSpace user registry, the Directory of Open Access Repositories (OpenDOAR), and Registry of Open Access Repositories (ROAR); the site was registered with Google Webmaster Tools for better indexing; and the repository was linked to several websites of the law school and library. The "Faculty Scholarship" collection and the "Texas Governor Executive Orders" collection became available shortly after. I then developed a poster of the newly established digital repository and presented it at the Texas Conference on Digital Libraries held at University of Texas Austin in May 2010.

Currently, our digital repository has more than eight hundred digital items as of August 2010. With more and more content becoming available in the repository, we plan on making an official announcement to the law community. We will also make entering first-year law students aware of the IR by including an article about the new repository in the library newsletter that is distributed to them during their orientation. Our future marketing plan includes sending out announcements of new collections to the law school using our online announcement system TechLawAnnounce and promoting the digital repository through the law library social networking pages on Facebook and Twitter. We also plan

on hosting information sessions for our law faculty and students to learn more about the digital repository.

Future Projects

There is no doubt that our digital repository will grow significantly because we have exciting collections planned for future projects. One of our law faculty, Professor Daniel Benson, donated some of his personal files from an eight-year litigation representing the minority plaintiffs in the civil rights case of *Jones v. City of Lubbock*, 727 F. 2d 364 (5th Cir. 1984) in which the minority plaintiffs won the case. The lawsuit changed the City of Lubbock's election system for city council members from the "at large" method to the "single member district system," which allowed the minority candidates consistently being elected.

This collection contains materials, notes, memoranda, letters, and other documents prepared and utilized by the plaintiffs' attorneys. It has significant historical value because a Texas Tech Law Professor and five Texas Tech Law graduates participated in that case successfully as pro bono attorneys for the minority plaintiffs.

In addition, we plan on adding Social Science Research Network

(SSRN) links to individual articles in the faculty scholarship collection. After that, the next collections we will work on are the law school and law library history materials. We also plan to do some development on the DSpace authentication to integrate with the TTU "eRaider" system to enable single log-in. In the future, we want to explore the possibilities of setting up a collection for the works of our law students and engage in electronic journal publishing using our digital repository.

Conclusion

It is not an easy task to develop an institutional repository from scratch, especially for a smaller organization. Installation and development are certainly a big challenge for a smaller library with limited number of IT staff. Outsourcing these needs to a service provider seems to be a feasible solution. Another challenge is training. We overcame this challenge by taking advantage of the state consortium's DSpace training sessions. Subscribing to the DSpace mailing list is necessary as it is a communication channel for DSpace users to ask questions, seek help, and keep up to date about the software.

All roads lead to Rome. No matter what platform you choose, whether open source or not, the goal is to pick a system that best suits your organization's needs. To build a successful institutional repository is not simply "scanning" and "putting stuff online." Various factors need to be considered, such as digitization, IR platform, collection development, metadata, copyright issues, and marketing and outreach.

Our experience has proven that it is possible for a smaller special library with limited resources and funding to establish an open source IR such as DSpace and continue to maintain the site and build the collections with success. Open source software is certainly not "free" because it requires a lot of effort. However, in the end it still costs a lot less than what we would pay to the proprietary software vendors.

References

1. "The Texas Tech University School of Law digital repository," <http://repository.law.ttu.edu/> (accessed Apr. 5, 2011).
2. "Repository Maps," accessed <http://maps.repository66.org/> (accessed Aug. 16, 2010).