

William & Mary Law School

William & Mary Law School Scholarship Repository

Library Staff Publications

The Wolf Law Library

11-2012

Transitioning Your Institutional Repository into a Digital Archive

Lauren Seney

lauren.seney@colorado.edu

Follow this and additional works at: <https://scholarship.law.wm.edu/libpubs>



Part of the [Law Librarianship Commons](#)

Repository Citation

Seney, Lauren, "Transitioning Your Institutional Repository into a Digital Archive" (2012). *Library Staff Publications*. 83.

<https://scholarship.law.wm.edu/libpubs/83>

Copyright c 2012 by the authors. This article is brought to you by the William & Mary Law School Scholarship Repository.

<https://scholarship.law.wm.edu/libpubs>

Trends

IN LAW LIBRARY MANAGEMENT AND TECHNOLOGY

Edited by Philip C. Berwick ♦ For academic, firm, corporate, and government law librarians

Transitioning Your Institutional Repository into a Digital Archive

By LAUREN SENEY, Access/Technical Services Librarian, Wolf Law Library, College of William and Mary, Marshall-Wythe School of Law, Williamsburg, Virginia

Developing a digital archive that provides access to materials about and related to the history of your school seems simple enough. After all, it is often easy to locate physical collections that meet these criteria, so it should be fairly straightforward to convert materials and make them universally accessible, right? This might be true in the perfect world, but there are far too many hurdles for this to be a reality. Many libraries are creating digital archives, but because every institution is different, this translates into unique implementations from library to library.

When the Wolf Law Library at the College of William & Mary set out to create an institutional repository, our efforts were focused on our faculty scholarship and law reviews. We had digital content readily available and quickly incorporated these collections into the archive with the aid of student workers. As these collections reached a maintenance level, we began

Trends

including items that were not available digitally. Our Annual Reports and Admissions Brochures were the first materials to make this transition because multiple copies of older editions were accessible for digitization. To move beyond these, we reached out to stakeholders within the law school.

Content

Promoting the repository quickly engendered excitement from our faculty, administrators, and alumni, initiating conversations about potential content. Some materials simply showed up in the library's mailbox with a "for the repository" note. While some of the ideas presented were not feasible due to copyright issues or a lack of content, we did add collections such as the Graduation Programs and Class Photos—where one administrator provided more than 40 years of content.

Outreach with the law school's Communications Department has also been beneficial, both to keep collections current and to generate new content. Communications provides materials when they are available and also directs us to the departments who are the most likely to have missing materials. This, in turn, leads to additional content as we build relationships throughout the law school. Also, we developed a workflow for keeping continuous collections up-to-date by raising awareness in the law school about the materials the library collects and what formats they should be delivered in.

With each relationship forged, the potential for new content multiplies rapidly. For example, for each item that becomes a new "collection," there is the possibility that back files are also available, increasing the new collection's size dramatically. As a result, in the past year our digital archive has exploded, and now includes materials from the Office of Development and Alumni Affairs, the Office of Career Services, and the Office of Admissions. Additional content covers the Law School and (now-defunct) Law Library newsletters, class photos, graduation materials, student newspapers, materials from Law School-sponsored conferences and lectures, and a collection dedicated to the history of the Marshall-Wythe School of Law.

While the majority of these items are documents, we also have photograph and video collections in their infancy.

Process

The law library maintains physical collections of law school documents that range from the materials mentioned above to documentation from one-time conferences to more than four decades of student newspapers. Some of these materials were already in the library's print collection, and others were scattered throughout different law school offices. Our digital archive began with the low-hanging fruit—the materials identified in the library's catalog. Once we exhausted these, we inventoried the law school materials in closed stacks and reached out to law school offices. Related materials that could be grouped into a collection of items were given priority over single items in the digitization process. As for single items, we perform research to see if additional materials are available somewhere in the law school, or in the main university's library archives.

The digitization process has been a learning experience for everyone involved, as it introduced new technology and procedures in the library. Adobe Acrobat Professional was already installed on several computers, but we acquired Adobe Photoshop, Omnipage Professional, and a Plustek OpticBook 3600 Plus Scanner for the digitization process. Items are scanned page-by-page and then edited down to a standard size. Additional editing is performed as needed to increase the readability of the text for both the end-user and the optical character recognition (OCR) software, usually to compensate for poor scan quality or to remove handwriting. Next OCR is performed on the documents. Those that are graphic heavy are OCR'd in Omnipage while text-based ones with neutral backgrounds are OCR'd in Acrobat.

It took a certain amount of trial and error to get to this workflow. There is a lot going on with each document (scan, edit, OCR) and we had about a dozen staff and students working on different stages of the process. Ultimately, we added checkpoints after scanning, editing, and OCR. At each stage, the document is examined for completeness and readability. If these

Trends

criteria are not met, individual pages will be re-processed. If the problem is with the original document, or due to the binding, we attempt to locate a cleaner copy to remedy the problem. We perform one final check to ensure that the document is complete and as true to the original as possible before it is added to the Scholarship Repository.

The workflow for documents that do not require scanning is a lot simpler. Documents are received from the creator, preferably as a PDF, but not always. If the document is received in an alternate format we will convert it to a PDF. Many of the documents have a physical counterpart, which we use to confirm that the conversion is done correctly. OCR is performed as necessary, though many documents do not require it because they have been converted from a document format and are already full-text-searchable. Even if we do not run the OCR, we spot check its accuracy.

The Scholarship Repository runs on bepress' Digital Commons platform, and we modify metadata fields we developed for this repository for each new collection. Basic information such as the title, creator/author, date of publication, and document type are collected from the individual items. Subject terms are assigned uniformly to some collections, such as the Admissions Brochure, and on an item-by-item basis for other materials, such as conference documentation. The collections are organized by the type of material, such as Class Photos and Commencement Activities, or by event, such as the George Wythe Lecture or the William & Mary Annual Tax Conference, so a description is presented at the collection level. For lectures, an additional description is given within the item record to provide more detail about that particular presentation.

Challenges

One of the first challenges we encountered as we began curating a digital archive was that our initial collection development policy for the repository did not account for all materials. While we had a broad array of ideas from the start, the policy only dealt with items of scholarly nature. Our policy is

continuously evolving as the repository grows to account for a wide variety of materials.

Then there were the mechanics. We purchased a book-edge scanner knowing we would be dealing with bound items, but did not realize that the binding on some of these items would be too tight to get a decent scan or that some of our materials were too large for the scanner. This resulted from the narrow focus of our original collection development policy—the vast majority of materials falling within the initial scope were printed on standard 8.5 x 11" paper. Oversize materials or single volumes that are bound too tightly for a book-edge scanner have made us think about outsourcing and alternative scanning techniques.

The physical condition of the items presented another hurdle. Fragile items were assessed and processed by one of the librarians. A larger issue was that many of the documents we digitized were bound as collections. When we searched for unbound copies, we discovered that we could not readily locate them. We worked with the documents we had while attempting to locate additional copies of materials. In many cases, we discovered unbound copies; however, it took an inventory of the unclassified materials in storage as well as requests to the offices that produced publications to procure them.

Multimedia is our most recent foray, and the most complex. We do not have many printed photographs requiring digitization, but we do have a large collection of digital photos, as well as a fair number of VHS tapes and DVDs from law school events. All of the library's photos are maintained on Flickr, and we have sought to archive a small portion of them in the Scholarship Repository. The sheer volume of our Flickr collection makes working with the photos unwieldy. The challenge of deciding which are archive-worthy is complicated by a limited display within the repository.

A video conversion project has been in the works for the better part of a year, and Final Cut Pro, an iMac, and a VCR have been added to our digitization toolkit. Getting videos from a VHS tape to a digital file presents its own challenges, but the larger hurdle is determining how and where digital

Trends

videos will be stored and how we will display them within our collections. Currently we are keeping copies of the files in the library on both a hard drive and a DVD and are using Vimeo for additional storage and as an embedded streaming media player in the repository.

This process has been far from seamless. Even with research into best practices, procedures, and equipment, we have still hit unanticipated hurdles in processing materials. While some trial and error is necessary to identify the most effective workflow, we have also had to work through unforeseen setbacks with content and technology. The biggest lesson we have learned is to be flexible and adaptable in both our policies and procedures in order to create a functional product. Fortunately a large supply of rich content, strong support from within the law school community, and a dedicated, flexible staff means that we are able to continue to curate a robust digital archive.

Creating a digital archive is a lot like raising a child. You can do all of the reading and talk to everyone you know with experience, but you will still make many decisions on the run. We continuously explore other digital collections and are in regular contact with bepress about how our content will be displayed and what changes can be made. Building a digital archive, like raising a child, is a lot of work. But you get a lot out of it as well.

Lauren Seney Access/Technical Services Librarian, Wolf Law Library, College of William and Mary, Marshall-Wythe School of Law, Williamsburg, Virginia. Email: lsene@wm.edu.