Preserving Librarian History in the Digital Age: Developing the American Library Association Institutional Repository

Cara S. Bertram
University of Illinois at Urbana-Champaign

The American Library Association Archives at the University of Illinois at Urbana-Champaign is the official repository of the American Library Association (ALA), including its offices, divisions, committees, round tables, and members. In 2013, the ALA Archives started a pilot project to launch a digital repository to manage and preserve born-digital records. This paper describes the project's background and outcomes. It provides an overview of the repository development process in addition to describing its structure, deposit process, and preservation software. It will also address the use of library driven software to compliment archival practices.

History

For the past forty years, the University of Illinois has managed the American Library Association's archives. In the 1970s, the ALA recognized the value of the organization's records as scholarly interest was piqued by the approaching ALA centennial in 1976. At that time, the archives of the ALA were housed in a warehouse in Chicago and not easily accessible to both staff and researchers. With a need for better access and care for its records, the ALA sought out an institution to house and manage its archives.

University Archivist Maynard Brichford made a case for the University of Illinois at Urbana-Champaign. Even though the Chicago campus was closer to ALA Headquarters, Urbana-Champaign was a more desirable location because the collection complemented Urbana-Champaign's Graduate School of Library Science (now the Graduate School of Information and Library Science). In addition, the University Library already held strong library related records at the University Archives. In 1973, the first shipment of ALA records arrived at the University of Illinois at Urbana-Champaign, and the American Library Association Archives was established (Brichford, 1979, p. 2-3). The development of the ALA Archives at the University of Illinois began a long term partnership between the association and university.

Project Background

Part of the ALA Archives mission is to collect and preserve records of the ALA that have continuing administrative, legal, or research value, regardless of format. This has created a variety of records in the ALA Archives holdings of approximately 3,000 cubic feet of archival materials. Materials include paper documents, scrapbooks, photographs, glass slides, posters, and a wide range of audiovisual materials. However, electronic records have presented the usual challenges of how to deal with accessions, preservation, and access to the ALA and the ALA Archives. While the ALA Archives has actively collected electronic records from depositors for the past several years, there has been no easy way for ALA staff and members to deposit them, nor a clear understanding of the ALA Archives abilities to support such materials.

Electronic records are received by the ALA Archives via email, CDs, flash drives or other hard drives, along with legacy formats, such as floppy disks and zip drives. Those records are processed to scan for viruses, uncover any sensitive information, produce checksum metadata, and arranged for access purposes and then placed on the University Archives server. Access copies are produced from the preservation copies and are made available online if appropriate (University of Illinois Archives). This has been an effective way to preserve electronic records and their metadata, as the records on the University Archives server benefit from the University Library Digital Preservation unit's initiatives in backing up the digital content of the University Library. The ALA's website has also been folded into the unit's web archiving program using the California Digital Library Website Archiving Service. But the process of accessioning and

preserving electronic records has often been reliant upon ALA staff and members to back up records onto portable devices or media and sending them to the Archives.

The ALA Archives faces similar problems that many archives have encountered with electronic records in that "important electronic records are simply maintained on the Web space or desktops of the creating office" and that "personnel have destroyed valuable departmental records while getting rid of their office files." This is especially problematic as records are increasingly produced and maintained only in digital formats (Bicknese, 2003, p. 89). The ALA produces many of their records in electronic form only and units are not always conscious of transferring them to the Archives as they would a filled filing cabinet.

The Association for Library Collections and Technical Services (ALCTS), an ALA division, began exploring options for a commercial service to create an institutional repository to manage the electronic-born publications of ALCTS and its members, particularly those who did not have access to other institutional repositories. This was in an effort to manage their electronic records and preserve the work of the division. The University approached ALCTS, offering to help develop the digital repository as a more cost effective option and to build upon the existing partnership between the ALA and the University.

The University of Illinois has since collaborated with ALCTS in developing a digital repository for the ALA as a whole, focusing on a means to preserve publications and archival materials. Seed money was provided by the ALA to launch the project. The funds allowed for the initial setup of the repository, its storage, and the hiring of a fulltime digital archives assistant for one year to help configure the site, process electronic records, and help create workflow and training documentation. The repository is called the American Library Association Institutional Repository (ALAIR), which collects, permanently stores, and provides open access to the publications and born-digital records of the ALA.

Development

The ALA Archives has worked with DuraSpace to create ALAIR using their DSpaceDirect service. Initially, we proposed to establish a locally hosted DSpace Repository, modeled after the University of Illinois's own institutional repository, the Illinois Digital Environment for Access to Learning and Scholarship (ideals.illinois.edu). Further review of repository options found that the DSpaceDirect service would be more efficient in terms of cost and staff hours.

While an in house implementation of a DSpace Repository was an option, all technical support, updates, maintenance, and preservation and storage would fall upon the ALA Archives. The Archives has a small staff of a one halftime professional and one halftime graduate assistant for the day to day functions of the Archives, with the Assistant University Archivist overseeing the Archives. Taking advantage of the DSpaceDirect's technical support, automatic updates, site maintenance, and preservation services helps to cut staff resources needed to maintain ALAIR. The budget provided by the ALA has allowed for a fulltime digital archives assistant for the start up, and then for an additional halftime graduate assistant to help maintain outreach efforts, training, reworking documentation as needed, and working with ALA units. The graduate assistant will also help with other Archives initiatives, such as digitization projects and processing electronic records on legacy formats. And in the case the ALA wants to manage ALAIR, the transfer from the University to ALA will be easier if it is on a third party's server than if the repository was on the University's servers.

The Archives was able to join the DSpaceDirect beta and once the contract went through, the repository was quickly set up. DSpaceDirect allows it customers to customize the colors and logo of the site to better reflect the ALA's main website. It provides a simple interface that includes multiple search options to access records. Further customization has allowed the ALA to retain its domain name in the address of the repository (https://alair.ala.org/).

ALAIR is organized into individual communities and sub-communities that reflect the structure of the committees, divisions, offices, and round tables as represented on the ALA's main page and in ALA itself, rather than using the ALA Archives record series structure. This decision was made so the repository content could be found more easily by ALA staff and members who are more familiar with the ALA website. The community system lends itself well to duplicating the complex structure of the ALA, allowing to nest down into several layers of sub-communities to represent the hierarchy of the organization's individual units.

Each community and sub-community page is customized with a description and logo of the ALA units to reflect their individuality and varying missions. Nested in each community and sub-community are collections that ALA members can deposit their records into. These collections and communities must be created via a designated administrative user. This kind of restriction is put into place to help prevent duplication of collections within the same community and to maintain a level of vocabulary control.

The communities roughly correspond with the top level record groups within the ALA Archives and efforts have been made to crosslink the holdings of ALAIR with that of the Archives. This will potentially lead researchers from the Archives to ALAIR to materials that they can readily access or from ALAIR to the Archives to find analog records that help with their research. There has been a problem with outdated record groups within the Archives, meaning that the current listing of ALA units within ALAIR does not always correspond with the listing of units within the Archives catalog. This is the result of unit name changes, absorption into other units, and units becoming defunct. Through research, many of these issues were resolved.

One concern frequently voiced when establishing an IR is the ability of the repository to preserve the records. Elizabeth Yakel and her co-authors noted in 2008 that, "IRs have not become the equivalent of trusted digital repositories with 'a mission to provide reliable, long-term access to managed digital resources to its designated community'" (Yakel et al, 2008, p. 328). The situation has changed somewhat since then. When looking into DSpaceDirect's services, the Archives considered the preservation options available, choosing one that best fit the needs of the Archives. DSpaceDirect provides preservation services via DuraCloud, which is DuraSpace's cloud preservation service (DuraSpace, 2014). All content is mirrored in two separate Amazon services, S3 and Glacier. These preservation services will allow for the retrieval of records even if they are deleted from the repository. And in the case that DSpaceDirect can no longer support ALAIR, all of the data and records will be returned to the Archives. Currently the repository has assigned to it one terabyte of storage space, with ability to add more if needed.

Many of the backend functions of the repository are accessible only to ALA Archives staff, who will manage the site in partnership with DSpaceDirect. ALAIR uses a Shibboleth authentication that will recognize ALA accounts to allow users to deposit records. Right now only ALA staff and ALA Archives staff can log into the repository, but the end goal is to have access available to all ALA members so that they can deposit into their respective committees, round tables, and sections.

Backend access to ALA staff and members is limited to depositing. In considering the level of access provided to users, archival principles had to be taken into consideration. Users are able to upload and describe each item. Edits for further and more accurate descriptions can be done by administrators as needed. While some IRs allow contributors to delete records as needed, while leaving a tombstone to indicate that the record was once there (Blankenship and Haines, 2008, p. 24), deleting is an administrative function. Only Archives staff can remove records, typically in cases such as copyright violations or privacy issues. This is to maintain the authenticity of the records once they are deposited into the repository. As with analog archives, records cannot be deaccessioned without serious consideration on the part of archivists.

For depositing, ALAIR uses Dublin Core metadata standards. To make the submission process as easy as possible and recognizing the diversity of records to be deposited, there are minimal

requirements for metadata. Only title and date are required, with creator and format recommended. The software adds minimal metadata of its own and automatically generates provenance information, date it was deposited, and an identification number. Additional fields for description, such as subject terms and abstracts, are also provided. Metadata fields can also be adjusted to fit the needs of depositors. Being conscious of the diverse work done by the ALA units, a controlled vocabulary was not chosen for the subject terms. This will allow for local terms to be applied without administrator intervention, however inconsistent terminology has already populated ALAIR, such as "bylaws" and "by-laws."

The partnership between DuraSpace and the Archives has been a beneficial one. Already updates to the software have addressed issues faced in the usability of the repository. The technical support has already proven to be prompt in addressing user problems, resolving issues within 24 hours.

Archives and Institutional Repositories

The development of institutional repositories (IR) began with the need for a place to deposit the scholarly works of faculty, and was originally a tool to provide access and disseminate journal articles and other publications (Spindler, 2008, p. 65). While IRs are mostly associated with colleges and universities, they also exist in governmental agencies, museums, corporations, and other organizations (Shreeves and Cragin, 2008, p. 89). Considering the scholarly output of its members and staff, an IR has been a good fit for the ALA and its members, along with being a complementary extension of the Archives.

Clifford Lynch has noted that the creation and management of an IR should require the "collaboration among librarians, information technologists, archives and records managers" along with administers and policy makers (Lynch, 2003, p. 2). Despite the call for collaboration, conversations about IRs have mostly involved librarians, information technologists and contributors (Burckel, 2008, p. 21). However, many institutions have found that IRs have not become a haven for scholarly research and publications, and have broadened their collecting policies, showing a lean towards archival materials.

In their analysis of the *Census of Institutional Repositories in the United States*, Yakel and her co-authors find that over 70 percent of records collected by IRs can be considered archival. This includes institutional records, photographs, project reports, newsletters, and scrapbooks, along with faculty papers in universities and colleges (Yakel et al, 2008, p. 335-337). Because of the broad scope in collecting policies and the overlap in goals to document the history of an institution, an IR and the institution's archives can be seen being either complementary or in competitive roles (Crow, 2002, p. 4). By including archives in the development and management of IRs, there is no need for the two to be at odds. Archivists bring experience in describing hierarchies and relationships between records (Burkel, 2008, p. 66), securing the authenticity of archival records within IRs, and can be essential in creating crosswalks between digital and analog records, guiding researchers to related materials.

In taking over the development and management of ALAIR, the ALA Archives has shaped its collecting policies to include archival records along with scholarly research and publications. In collaborating with ALCTS, it also became clear that they were not only seeking to store publications and scholarly works, but archival materials as well. Backlogs of records on websites, social media, desktops, and servers have long been an issue, and an institutional repository seemed to be a partial answer to this problem. In shaping the collecting policy for ALAIR, the IR collects in four broad categories: official records, historical documentation, audiovisual materials, and publications (American Library Association Archives, 2014). By bringing ALAIR under the Archives it is able to take on a complementary role in access and preservation of records.

The development of ALAIR has not replaced the Archives, nor has it even replaced all the digital preservation components of the Archives. The University Archives server is still utilized for

electronic records that might have complex file structures that need to be preserved and cannot be duplicated in ALAIR. Also for records with restrictions on them, such as records that might contain sensitive information, such as case files, personnel records or records with identifying information.

Outcomes

ALAIR has been live since January of 2014 and was announced at the American Library Association's Midwinter Meeting. Efforts were made to meet with the Executive Boards of the ALA Divisions, who are potentially some of the heaviest users of ALAIR. Since then there has been outreach efforts via email, a virtual training session, individual consultations and plans for a 2014 fall training session at the ALA Headquarters. Documentation has also been developed for uploading records and for searching the repository. While the infrastructure and documentation is in place, the challenge has now been getting ALA units to deposit.

While many ALA units have expressed interest in depositing more recent records, there is a large backlog of records stored on websites. Difficulty arises with the amount of digital records stored on websites and challenges behind uploading individual items from ten years' worth of records. The batch upload process is currently not available on the user end and procedures for batch uploading on the administrative side are still being developed.

Lisa Blankenship and Annette Haines point out that by its very nature, an IR requires the collaboration between libraries, in this case archives, and the creators of the content (Blankenship and Haines, 2008, p. 24). In this instance, the ALA Archives has found a strong collaborator with the ALA Headquarters Library. The ALA Headquarters Library has demonstrated interest in the repository and has been working with the Archives in depositing not only electronic born records, but also electronic surrogates that have been stored on computers. The close proximity that the Library has to the rest of the units makes it a good starting point for ALA staff members wanting to use the repository. As of now the Archives and the Library are working with *American Libraries*, the flagship magazine on depositing weekly electronic newsletters that go back to 2006. The collaboration between the Archives and the Library has long been a fruitful one, now with a new element to it in promoting ALAIR.

There has also been interest by affiliate library associations in using the repository. These plans have been put to the side while the Archives focuses on getting the ALA completely setup with the repository. However, the potential of having affiliate library associations deposit into the repository will strengthen the collection and will be of a reflection of the Archives which houses the records of several affiliated and related library associations.

Another outcome from this project is that gaps in the physical archival record are now being filled by electronic records, one example are the records of the ALA elections. The Archives holds the physical records of the elections from 1926 to 2001, but has not received records from more recent elections as many of the records were now digital. A recent reference request illuminated this gap and the lack of accessions of certain crucial records of the association that have mostly gone digital. The ALA Library was able to help fill in the gap in the Archives records by depositing the records of the Election Committee, which cover the years 2002 through 2012. This includes election results, samples of ballots and forms, and voting statistics.

Dorothea Salo, a digital repository librarian, warned that, "The notion that faculty members will actually push buttons and type metadata in order to deposit materials into IRs is an article of faith among repository-software developers. In practice, however, most deposits are third-party mediated, many by librarians, some by support staff or IT personnel" (Salo, 2008, p. 112). True to Salo's statement, the initial uploading of backlogged materials from ALA unit websites will need assistance from Archives staff. The process has been slowed as batch uploading is limited to administrators from DSpaceDirect, making it difficult for ALA units and Archives staff to

upload backlogs of electronic records on their own. There are also problems with ALA units being able to identify which collections to deposit into. Within their user login, a dropdown menu showed all available collections, but did not display the hierarchy, making depositing confusing especially when there are "Publications" and "Minutes and Agendas" collections for almost every community. An update in the software now shows the hierarchy in the menu, however the complex structure of the ALA does not lend itself well to this function.

Despite the concerns with overcoming the initial uploading efforts, ALA units and members have expressed strong enthusiasm for the project. There are hopes that the ALAIR will grow into a substantial resource that will connect the analog records on the shelves of the Archives to the downloadable document in the institutional repository.

Bibliography

American Library Association Archives (2014). "ALAIR Policies."
        <http://archives.library.illinois.edu/ala/alair/alair-policies/>. [Accessed: 8/1/2014].
Bicknese, Douglas (2003-2004). "Institutional Repositories and the Institution's Repository:
        What is the Role of University Archives with an Institution's On-Line Digital Repository?"
        Archival Issues, v. 28, n. 2, p. 81-93.
Blankenship, Lisa and Annette Haines (2008). "The Institutional Repository: Storing and Sharing
        Digital Scholarship." *Art Documentation: Journal of the Art Libraries Society of North
        America*, v. 27, no. 2, p. 22-25.
Brichford, Maynard (1979). *Guide to the American Library Association Archives.* Chicago:
        American Library Association.
Burckel, Nicholas (2008). "Academic Archives: Retrospect and Prospect." In: Prom, Christopher
        and Ellen Swain (ed.). *College and University Archives: Readings in Theory and
        Practice.* Chicago: Society of American Archivists, p. 3-26.
Crow, Raym (2002). "The Case for Institutional Repositories: A SPARC Position Paper." In: *ARL
        Bimonthly Report 223.*
        <http://www.sparc.arl.org/sites/default/files/media_files/instrepo.pdf>. [Accessed:
        8/10/2014].
DuraSpace (2014). "DSpaceDirect Features." <http://dspacedirect.org/features>. [Accessed:
        8/10/2014].
Lynch, Clifford (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the
        Digital Age." In: *ARL Bimonthly Report 226.*
        <http://www.arl.org/storage/documents/publications/arl-br-226.pdf>. [Accessed:
        8/1/2014].
Salo, Dororthea (2008). "Innkeeper at the Roach Motel." *Library Trends,* v. 57, n. 2, p. 98-123.
Shreeves, Sarah L. and Melissa H. Cragin (2008). "Institutional Repositories: Current State and
        Future." *Library Trends,* v. 57, n. 2, p. 89-97.
Spindler, Robert (2008). "Electronic Publishing and Institutional Memory." In: Prom, Christopher
        and Ellen Swain (ed.). *College and University Archives: Readings in Theory and
        Practice.* Chicago: Society of American Archivists, p. 53-68.
University of Illinois Archives. "Digital Workflow." < http://archives.library.illinois.edu/staff-
        resources/digital-workflow/>. [Accessed: 8/29/2014].
Yakel, Elizabeth, Soo Young Rieh, Beth St. Jean, Karen Markey, and Jihyun Kim (2008).
        "Institutional Repositories and the Institutional Repository: College and University
        Archives and Special Collections in an Era of Change." *The American Archivist*, v. 71, n.
        2, p. 323-349.