National Digital Stewardship Residency - Boston

Project Summaries 2015-16 Residency

Harvard Library – "Preparing for a Trustworthy Repository Certification of Harvard Library's DRS."

Harvard Library's Digital Repository Service (DRS) has been in production for almost 15 years and contains 175 TB of digital material in many different formats including images, text, audio, websites, documents, email and soon video. It is under the management of Harvard Library's Digital Preservation Services and the technical infrastructure is developed and maintained by Harvard University Information System's Library Technical Services (LTS). The DRS is used by approximately 50 libraries, archives and museums across Harvard to provide long-term preservation and access to digitized and born-digital digital material.

After a multi-year project that started in 2009, last year the next-generation version of the DRS with increased preservation capabilities was launched. In 2015 the DRS Improvement Team was formed. This is an opportune time to reassess the DRS in comparison to current digital preservation standards and prepare to seek to certify it as a trusted repository. This would pinpoint additional areas to improve and ultimately demonstrate to the Harvard curators, archivists and collection managers that rely on the DRS for digital preservation that it meets the highest preservation standards.

PROJECT SUMMARY

The Harvard Library (HL) seeks a National Digital Stewardship Resident to prepare the DRS for a trustworthy repository certification based on ISO 16363. The work will be done in 4 phases:

Phase 1: Immersion and Analysis

- Gather ISO 16363 standards documents and related instructions
- Identify, evaluate and select additional tools to help
- Review and become familiar with ISO 16363
- Review and become familiar with publicly available documentation from other institutions and organizations in response to repository certifications
- Locate and become familiar with all DRS documentation
- Create documentation inventory template
- Inventory DRS documentation
- Create a detailed gap analysis of documentation, plans, processes, strategies, etc. needed for repository certification

Phase 2: Implementation and Updates

- Work with DRS repository manager to create an action plan for changes and improvements in preparation for an audit
- Work with Harvard Library and IT staff to make improvements outlined in the action plan

Phase 3: Assessment and Self-audit

- Perform a self-assessment on the DRS
    - Decide in collaboration with DRS repository manager which sections of ISO 16363 will be covered - due to time constraints it may not be possible to cover all of it
- Produce an audit report with the findings

Phase 4: Improvements

- Work with DRS repository manager to determine improvements that can be made to better organize, expose and preserve documentation
- Implement improvements in file system and websites
- Deposit key documentation to the DRS for preservation

Specific Goals/Objectives

The main goal of the project is to prepare for eventual certification of the DRS as a trustworthy repository by an external auditing body so that the Library can demonstrate that the DRS meets the highest preservation standards.

Although that is the main goal there are additional goals with large benefits to Harvard Library:

- Identify key areas of the DRS needing change and improvement including policies, strategies, processes and documentation
- Produce a detailed inventory of DRS documentation which is currently distributed across many file systems, wikis and websites
- Improve the organization, preservation and dissemination of DRS documentation

John F. Kennedy Presidential Library and Museum –"Preservation of a Legacy: Long Term Digital Preservation at the John F. Kennedy Presidential Library."

The Kennedy Library's digitization initiative, known as "Access to a Legacy," is a public-private partnership between the John F. Kennedy Presidential Library and Museum and the John F. Kennedy Library Foundation. The initiative's objectives are to: digitize, index, and permanently retain millions of presidential documents, photographs, and audiovisual recordings; provide worldwide, online access to these materials and facilitate their search and discovery through the

use of metadata; protect historical assets through remote replication; and minimize the deterioration of unique and irreplaceable records and artifacts. Since work began in 2007 we have ingested over 730,000 pages, 23,500 photographs, 1,800 audio files, and approximately 680 moving image files into our digital asset management system, which, along with their derivatives, amount to approximately 70 terabytes of data. We add new content to this system on a daily basis.

The Library has developed proven and efficient processes for systematically digitizing entire collections and series of textual materials, photographs, and sound recordings, as well as for digitizing smaller portions of collections in response to on demand requests from remote researchers. We have also begun ingesting born-digital AV content, most of which documents high-profile events that take place at the Library. We create high quality preservation files along with lower quality access renditions, and we store copies of both on EMC equipment on site and at a disaster recovery location at an Iron Mountain Facility.

While the Kennedy Library may be in a better position than many institutions who are struggling to store their digitized and born digital content, we are far behind where we should be in terms of understanding how to manage and preserve these files. We do not have a formal digital preservation plan, nor do we have a system in place to ensure that our digital objects are not degrading over time. We have yet to fully address the challenges of accessioning and normalizing born digital file formats. We are also dependent on in kind donations for the maintenance of our entire digital asset management system and for the use of the disaster recovery site.

Our hope is that the work of this Resident, who will learn about and document all of the challenges we face and all of the possible solutions we could implement, will enable us to effectively manage and preserve our digital content for years to come.

MIT Libraries – "Archival Storage: Bringing Holistic Decision-making into Action."

The NDSR Boston Resident at MIT Libraries in 2015-2016 will be an active participant in an Archival Storage project, working to ensure that digital preservation standards are adhered to as a multi-stage approach is devised and implemented. The project will address both organizational and technological storage requirements and expectations to acquire, preserve, and provide our digital collections as the range and scale of our digital content expands. The Resident will experience a balance of planning to doing, research to practice, organizational to technical issues, and governance to day-to-day operations within a standards-based framework for demonstrating good digital preservation practice.

State Library of Massachusetts – "Assessment and Workflow Analysis for the Preservation of Born-Digital and Digitized Massachusetts State Publications."

The State Library is mandated by law to preserve and make available a complete collection of Massachusetts state publications. As such, the library has an extensive collection of state documents dating back to the late 18th century. In the past 10 to 15 years, agencies have increasingly published materials electronically, sometimes no longer producing a paper copy. Although agencies are, by Massachusetts law, supposed to send eight copies of their printed publications to the State Library, often they are only posted on their websites. It is common for items to disappear from websites after a period of time, so it is imperative that the library capture and preserve these materials.

The library has had a DSpace digital repository since 2006 for these documents. The staff actively adds materials located and has worked with a company to develop a crawler management system that identifies potential documents from URLs known to contain documents and tracks whether such found documents are considered "state publications". Staff downloads these reports, catalogs them and adds to the library's repository. The library staff actively seeks out born-digital documents and records the URLs found for tracking. However, there has not been a comprehensive review of Massachusetts state government websites completed.

Finally, the ability for agencies to publish materials online has blurred the meaning of what a state document is and who is charged with preserving it. Printed newsletters can become newsletters in pdf form, or in straight HTML format, or the content could now be published as a blog. This project will further refine and prioritize what the library should be preserving.

University of Massachusetts at Boston, University Archives & Special Collections, Joseph P. Healey Library - "Digital Preservation Planning and Implementation in University Archives and Special Collections at UMass Boston."

University Archives and Special Collections in the Joseph P. Healey Library at the University of Massachusetts Boston seeks a National Digital Stewardship Resident to develop policies and workflows for the strategic, long-term preservation of digital assets in the department's custody and care.

The Resident will begin by reviewing existing University Archives and Special Collection (UASC) practices regarding digitization, description, and preservation – experience that, along with hands-on work with our existing systems and a review of the digital stewardship landscape at the University, will help the Resident understand departmental processes, limitations, and future needs. Through work with the legacy digital content (data, image and video files) from the Mass. Memories Road Show, the Resident will research best practices and begin to develop workflows and policies that align with and refine existing procedures and that effectively address

issues that arise at all steps of a digital object's lifecycle: digitization practices, the generation of technical, descriptive and preservation metadata, and, finally, processes of providing access to digital assets and ensuring the long-term, cloud-based digital preservation of those assets. Additionally, during this first phase of the project, the Resident will begin to address other issues, such as the relationship in UASC between local storage and access (hard drives, CONTENTdm repository) and cloud storage and preservation (UASC's hosted, cloud-based preservation platform, DuraCloud).

Once the Resident has completed this first Research and Practice phase, the Resident will submit a preliminary report on the digital stewardship landscape and needs of UASC that will be shared with UASC staff, Library leadership, and other University stakeholders, such as Information Technology. At this point, in a Review and Testing phase, the Resident will work collaboratively with UASC staff, Library leadership, and other stakeholders to further develop workflows that are thorough and tested, and which prepare UASC for continuing internal digitization projects and existing and ongoing community-engaged digital history initiatives such as the Mass. Memories Road Show.

Finally, the Resident will, from this experience, develop policies and procedures in collaboration with UASC staff, Library leadership, and University stakeholders that will help the department ensure and benefit from the long-term, cloud-based digital preservation of UASC's digital holdings.

Once policies and workflows have been reviewed and tested, the Resident will begin to work on the Implementation and Final Report phase of the project. During this phase, and to further test and refine policies and workflows, the Resident will oversee, with the assistance of UASC staff and student employees, the assigning of preservation and technical metadata and the eventual ingest of more than 6,000 previously digitized legacy images, videos files, and associated metadata from the Mass. Memories Road Show into the department's cloud-based, hosted digital preservation platform, DuraCloud. An important step for inclusion in the workflow will be determining processes and schedules for checksums and regularly verifying and documenting the integrity of files stored through DuraCloud.

Throughout this last phase, the Resident will prepare a final report that will include documentation of the project, as well as recommendations and, ideally, procedures for managing a range of digital file formats archival across collections. The Resident will also have the opportunity to prepare blog posts for the UASC and Library blogs about this work. There is of course the possibility that a summary of the report may be revised into an article for the Society of American Archivists, the New England Archivists, or another professional publication or journal.