A Permanent Home for Ephemeral Policies: Integrating DSpace with an Enterprise Content Management System in an Institutional Archive

Nicholas Webb
Digital Archivist
Icahn School of Medicine at Mount Sinai, New York, NY
nicholas.webb@mssm.edu

Session Type (select one)
- ☐ Panel
- ✓ Presentation

Abstract

This presentation will discuss the successful integration of a DSpace instance with an enterprise content management system, examining some of the challenges faced in connecting two disparate information systems for the purpose of long-term preservation.

In 2014, IT staff at the Mount Sinai Medical Center implemented a content management system (CMS), based on the open source Liferay portal, to centralize the management of policies and procedures across all Medical Center departments. As the permanent home for expired policies, the Mount Sinai Archives was a stakeholder in this project. Archives staff worked with members of the CMS team to develop tools and workflow for automatically exporting policies from the CMS into the Archives' existing DSpace digital repository, which had previously been used primarily for the storage and dissemination of manually cataloged digital objects. The integration process required addressing a number of technical and descriptive challenges, including crosswalking descriptive metadata, automating the creation of preservation copies, and dealing with inaccuracies in user-submitted metadata. The end result was a bespoke combination of automated data transfer and manual staff intervention by which changes to the CMS are regularly ingested into the repository.

Conference Themes
Select the conference theme(s) your proposal best addresses:

- ☐ Supporting Open Scholarship, Open Science, and Cultural Heritage
- ☐ Managing Research (and Open) Data
- ✓ Integrating with External Systems
- ☐ Re-using Repository Content
- ☐ Exploring Metrics and Assessment

- ☐ Managing Rights
- ☐ Developing and Training Staff
- ✓ Building the Perfect Repository

Keywords

- Content Management Systems
- DSpace
- Institutional Archives

Audience

The primary audiences for this presentation will be repository managers working with institutional archival records, who may find the presentation relevant to their own electronic records challenges, and the DSpace user and developer community, for whom it will be an interesting case of a nonstandard repurposing of the DSpace platform.

Background

The proposal most directly addresses this year's conference theme of "integrating with external systems" – in this case, integrating a repository platform designed for scholarly use with a content management system designed for an enterprise intranet. The presentation also touches on the theme of "building the perfect repository" by exploring how a scholarly repository was used to preserve an active electronic records series with ongoing accretions – a successful repurposing by which a small institutional archive with limited resources was able to avoid the need for additional software or hardware.

Presentation content

I plan to begin with a brief description of the Mount Sinai Archives' digital repository, placing it in the context of the use of institutional repository software to manage archival records. From there I will discuss the CMS project, emphasizing the Archives' role as a stakeholder. I will discuss the descriptive (crosswalking of CMS internal metadata to qualified Dublin Core) and technical (export in the DSpace Simple Archive Format) aspects of the integration process, including reflections on how the DSpace Community-Collection-Item data model compares to other forms of hierarchically structured data. I will conclude with a discussion of lessons learned and future plans, including examination of the workflow bottlenecks (creation of preservation copies and correction of user-submitted metadata) which have an ongoing impact on the system.

Conclusion

Institutional repositories are sometimes called upon to manage and preserve digital objects that do not fall neatly into the categories of scholarly output or research data but are nonetheless institutionally and historically valuable. This presentation will demonstrate that successful integration of an institutional repository with nonstandard external data sources is possible even in the context of a small archive with limited resources, demonstrating the ways in which digital archivists working in such a context can be part of the broader cross-disciplinary conversation around institutional repositories.