ArchivesSpace-Archivematica-DSpace Workflow Integration

Michael Shallcross
Bentley Historical Library
University of Michigan
Ann Arbor, Michigan 48109-2113 U.S.A.
1-734-936-1344
shallcro@umich.edu

Max Eckard
Bentley Historical Library
University of Michigan
Ann Arbor, Michigan 48109-2113 U.S.A.
1-734-763-7518
eckardm@umich.edu

Abstract

In recent years, ArchivesSpace and Archivematica have emerged as two of the most exciting open source platforms for working with digital archives. The former manages accessions and collections and provides a framework for entering descriptive, administrative, rights, and other metadata. The latter ingests digital content and prepares information packages for long-term preservation and access. In April 2014, the Bentley Historical Library received a $355,000 grant from the Andrew W. Mellon Foundation to partner with the University of Michigan Library on the integration of these two systems in an end-to-end workflow that will include the automated deposit of content into a DSpace repository. This poster will introduce the "ArchivesSpace-Archivematica-DSpace Workflow Integration" project and its goals, strategies, and development roadmap.

General Terms

Preservation strategies and workflows; Innovative practice.

Keywords

ArchivesSpace, Archivematica, DSpace, Digital archives, Workflow development, Appraisal, Arrangement and description.

1. INSTITUTIONAL CONTEXT

The Bentley Historical Library collects and preserves unique materials related to the University of Michigan and the state as a whole. 80 years after its founding, the library has amassed 8,000 research collections that range from the papers of governors, to the records of student and faculty, to the entire historical record of intercollegiate athletics at Michigan. These holdings include more than 20 TB of digital content, with extensive web archives, born-digital archives, and digitized collections of print, photographic, and audio-visual materials. As part of its mission, the Bentley Historical Library is committed to ensuring the preservation and accessibility of this content over the long-term by implementing professional best practices and standards in its workflows and infrastructure.

The Bentley has actively managed large collections of born-digital content since the 1997 accession of former University of Michigan President James Duderstadt's personal computer. The 2010-2011 MeMail Project (funded by a generous grant from the Andrew W. Mellon Foundation) helped the Bentley develop more robust and uniform preservation procedures as staff explored strategies to collect and preserve the email of key university administrators. This work led staff to develop AutoPro, an ingest and processing tool comprised of 30 Windows CMD.EXE scripts that guides archivists through a standardized workflow and creates a full audit trail. Since moving into production in 2012, AutoPro has been used to prepare more than 230 accessions of digital archives (approximately 1.5 TB) that are accessible in Michigan's Deep Blue DSpace repository.

2. PROJECT GOALS

While an effective tool, AutoPro's command line interface and limited error handling capability create inefficiencies and the amount of time it takes to maintain and update scripts and software on individual workstations has significant implications for sustainability. To address these issues, the Bentley Historical Library sought funding from the Andrew W. Mellon Foundation to integrate ArchivesSpace, Archivematica, and DSpace into an end-to-end digital archives workflow. The unique strengths and affordances of the different systems lend themselves naturally to specific archival functions: ArchivesSpace for accessions, description, and tracking rights and administrative metadata; Archivematica for ingest and creation of Archival Information Packages (AIPs); and DSpace for preservation storage and access. In bringing these platforms together, project staff hope to achieve three main goals:

1. Streamline a digital archives workflow from ingest through the deposit of fully processed materials into DSpace. Manual interaction and intervention with digital archives will thereby be reduced to essential procedures to create greater efficiencies and remove possibilities for human error.

2. Facilitate the creation and reuse of metadata among platforms, including archival description, administrative information, and PREMIS rights. These metadata will be recorded in ArchivesSpace, associated with information packages in Archivematica (in addition to essential technical metadata), and displayed and/or acted upon by DSpace.

3. Improve reporting functionality and provide better tools in Archivematica so archivists may (a) review and

appraise files in a more thorough manner and (b) logically arrange content with archival description from ArchivesSpace (and thereby associate digital objects with archival object records).

All project deliverables, including modifications to source code, plugins, and documentation will be contributed back to the appropriate open source code bases or otherwise be made freely accessible to the archives and digital preservation communities. The Bentley will also ensure that new features and functionality are modular so that other institutions may adopt some or all of the project features (for instance, only the appraisal and arrangement tab and integration with ArchivesSpace) and/or modify code to meet local needs.

3. DEVELOPMENT ROADMAP

Artefactual Systems Inc., the developers of Archivematica, joined the project as programmers in late 2014 (having previously served as technical consultants); since then, the company has assisted with an in-depth review of the Bentley's digital collections and workflows; an analysis of existing features and functionality in ArchivesSpace and Archivematica (with additional exploration of areas for future development and integration); and the articulation of functional requirements and development priorities. System integration formally began in April 2015 with the commencement of agile development sprints. The project will proceed through the following seven phases through its deadline in April 2016:

Phase 1: Creation of a new appraisal and arrangement dashboard tab in Archivematica. Initial development will focus on creating a new Archivematica dashboard tab and user interface to characterize and appraise files before intellectually (and logically) arranging them with ArchivesSpace. This new tab will provide faceted searching within transfers and generate reports for one or more transfers (or components thereof), with information on file format and media type distribution, duplicate files, size on disk, sensitive data, etc. The Bentley also seeks improved tools for archivists to view or render content within Archivematica to gain a better understanding of intellectual content and value, confirm the presence of sensitive data, or deaccession materials.

Phase 2: ArchivesSpace integration. Once a basic user interface is established for the appraisal and arrangement tab, project staff will focus on the nuts and bolts of ArchivesSpace integration. The most prominent feature will be an ArchivesSpace pane in the appraisal and arrangement tab that will permit archivists to match files/folders from the transfer backlog with an ArchivesSpace archival object, thereby creating a Submission Information Package (SIP) that will correspond to a digital object record in ArchivesSpace and form a single 'item' in DSpace. To accomplish this, archivists will load the appropriate resource record in the ArchivesSpace pane and then navigate through the intellectual hierarchy to an appropriate level of description, at which point content will be dragged and dropped from the backlog transfer pane onto a specific archival object. In navigating the intellectual arrangement, users may create new (or edit existing) archival object records at any point, operations that will be limited to the title, date, 'level,' and a note, with other fields to be modified directly in ArchivesSpace. Once this arrangement has been finalized, the archivist will click a "Create SIP" button to initiate Archivematica ingest procedures and use the ArchivesSpace API

to create or edit an archival object records and generate associated digital object records.

Phase 3: AIP repackaging. The Bentley currently stores AIPs in its DSpace repository and plans to continue using these for both preservation and access to avoid the redundancy of creating and storing separate Dissemination Information Packages (DIPs). As part of this strategy, the library packages large, multi-file digital objects in .zip files to simplify archival management and user access. Once content is moved from the Archivematica transfer backlog to an ArchivesSpace archival object, archivists will have the opportunity to select materials (within a single SIP) to be packaged together in a .zip file, with the ability to create multiple .zip files per SIP. In cases where a SIP contains only a few files, these will be deposited without packaging. After arrangement and packaging decisions are complete and a SIP has been created, Archivematica will continue with its ingest workflow, generating the .zip file(s) at the conclusion of this process. In another departure from current workflows, the METS file, metadata, and log files produced by Archivematica will be placed in a .zip file and deposited as a bitstream alongside the data in the appropriate DSpace item. With this 'chipped dog' approach, original and preservation copies of content will be available to researchers while access to the metadata and logs is restricted to archivists and repository staff.

Phase 4: Refinement of the appraisal and arrangement dashboard tab. Based upon user testing, feedback from other institutions, and additional development work, Artefactual Systems will refine the appraisal and arrangement tab to ensure that its features and layout best meet the needs of the user community at large.

Phase 5: External tools integration. Once changes to the Archivematica dashboard and integration with ArchivesSpace have been successfully implemented, Artefactual Systems will explore the integration of external tools to permit viewing and rendering a wider variety of file formats and mime types.

Phase 6: DSpace integration. Development related to DSpace will involve system-agnostic technologies such as SWORD and ResourceSync to ensure that the ArchivesSpace-Archivematica integration could be modified to function with other repositories (such as Fedora or CONTENTdm). Major requirements include the automated deposit of content to an existing DSpace collection, the crosswalk of descriptive and administrative metadata to Dublin Core elements in DSpace, and the ability to return unique 'handles' to the ArchivesSpace digital object record so that <dao> elements will include direct links to content.

Phase 7: Bug fixing / completion. The final phase of the project will involve the resolution of any bugs and final development tasks, taking into account additional user testing and feedback.

4. MORE INFORMATION

Updates on development efforts will be posted to the Bentley Historical Library's project blog, where questions and comments are welcome (see http://archival-integration.blogspot.com/).