

7-1-2013

One Digital Preservation Service, Many Repositories: Medusa at the University of Illinois at Urbana-Champaign

Kyle Rimkus

University of Illinois at Urbana-Champaign

Follow this and additional works at: <https://lib.dr.iastate.edu/macnewsletter>



Part of the [Archival Science Commons](#)

Recommended Citation

Rimkus, Kyle (2013) "One Digital Preservation Service, Many Repositories: Medusa at the University of Illinois at Urbana-Champaign," *MAC Newsletter*: Vol. 41 : No. 1 , Article 6.

Available at: <https://lib.dr.iastate.edu/macnewsletter/vol41/iss1/6>

This Electronic Currents is brought to you for free and open access by Iowa State University Digital Repository. It has been accepted for inclusion in MAC Newsletter by an authorized editor of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

MACers interested in writing a column for “Electronic Currents” are encouraged to contact Joanne Kaczmarek at jkaczmar@illinois.edu.

One Digital Preservation Service, Many Repositories: Medusa at the University of Illinois at Urbana-Champaign

By Kyle Rimkus, University of Illinois at Urbana-Champaign

The Need

The University of Illinois at Urbana-Champaign is developing a digital preservation service called Medusa to provide a storage and management environment for digital collections selected for long-term retention by the library. The library currently deploys, across its numerous departmental collections and repositories of special collections, a variety of platforms for managing access to digital content. Among others, these currently include CONTENTdm, DSpace, the Internet Archive, the HathiTrust Digital Library, Archon, and Olive ActivePaper. While the majority of the master files created for these access systems are stored on file servers managed by the library’s Information Technology group, a significant number of them reside on more precarious media such as hard drives on staff members’ shelves, optical discs in an off-site storage facility, or file servers leased from external companies. Furthermore, each access system generally represents a discrete team of people with its own file management practices and methodological approaches. A data set submitted to the IDEALS institutional repository, a book from the Classics Library digitized by the Digital Content Creation unit, and a hard drive of electronic records accessioned by the University Archives all represent digital objects created under different circumstances and stewarded by the library according to different principles. Each of these digital objects, however, will require a robust, reliable underlying digital preservation layer if it is to stand the test of time.

Meeting this need requires a creative application of traditional digital preservation management concepts. While the applicability of such bulwarks for digital preservation as the Open Archival Information System (OAIS) reference model and the Trusted Digital Repository standard furnish the guiding principles of the Medusa digital preservation repository, the Medusa project team also strives to incorporate ideas from other areas of library and information science into its services and technical architecture. Digital preservation, in fact, furnishes cultural memory institutions like libraries, archives, and museums with a special opportunity to identify commonalities often obscured by the jargon and isolation of their specific subfields, the discovery of which often leads

to unexpected forms of collaboration and innovation. Medusa’s design, for example, has already benefited from exposure to contemporary practices and trends in archival management and preservation assessment.

The Influence of Practices in Archives and Physical Preservation

Archival Practice

At the heart of Medusa is a web-accessible collection registry¹ inspired by archival management systems like Archon and ArchivesSpace. This web application allows preservation managers to identify and create records for

- *Repositories* responsible for general curatorial decisions related to the preservation of, access to, and rights status of collections of digital content;
- Discrete *collections* of related digital objects; and
- *Producers* who create digital files and have oversight over digital collection management, often on behalf of specific repositories.

Medusa’s managers in the Preservation Unit are working closely with colleagues in the library’s special collections to tailor the system’s services to their needs, especially for collections of born-digital materials. Specifically, preservation staff is establishing workflows to integrate computer media migration and file analysis services into nascent practices for the arrangement and description of born-digital electronic records and personal papers in the University Archives. The goal of these efforts is to define a collaborative division of responsibility that integrates specialized hardware and software available to preservation staff into the archival arrangement and description practices of the library’s repositories of special collections, with an eye toward the long-term curation of such records in Medusa.

Physical Preservation Assessment

Just as physical preservation administrators have long cultivated a tradition of systematically mitigating risk to the long-term survival of physical collections by conducting regular collection surveys and assessments, practitioners of digital preservation also acknowledge the important role assessment must play in effective digital preservation

(Continued on page 30)

(Continued from page 29)

management. However, digital preservationists currently possess a rather limited set of tools to aid in assessment. To address this, Medusa allows preservation managers to append assessments on the “collection,” “file group” (that is, batch), or “item” level, and to assign rankings of preservation priority to externally managed content not yet ingested into Medusa. While the format of these assessment records is rudimentary at present, the feature itself is intended to expand over time as digital preservation planning at Illinois becomes more sophisticated. In fact, the Preservation Unit has spent much of the past year conducting a comprehensive survey to identify all digital collections to be managed one day in Medusa, their current storage environment, total number of files, and representative file formats.

Preservation Levels and Technical Details

As part of this pre-ingest assessment process, the Preservation Unit has also been visiting the library’s content producers where they work and asking to observe their file management practices on site. The open nature of these discussions has allowed preservation managers to gain an effective understanding of how content producers manage their respective workflows, including details such as where they store their files and how they name them, as well as what aspects of digital preservation are important to them. Preservation managers then record summaries of these interviews as assessment records in the collection registry and will use them to inform the way that individual collections are ingested and managed in Medusa.

Repository: Rare Books and Manuscripts Library

Carl Sandburg Collection Collection

UUID: 811498e0-e3fb-012f-c5b6-0019b9e633c5-0

Start Date:

End Date:

Contact Person NetID:

Access Systems:

- ContentDM

Type of Resource: still image

Preservation Priority: low

Total Size (GB): 289.688

Description:

The more than 2,700 photographs in this collection are scanned from the Carl Sandburg Collection housed in The Rare Book & Manuscript Library. Spanning the years 1893-1987, these images are part of a collection that includes typescripts and galley proofs of many of Sandburg's works, his correspondence with literary and public figures, recordings and transcriptions of Sandburg's radio broadcasts, and a supporting book collection of approximately 5,000 volumes.

Private Description:

Access URL:
<http://images.library.illinois.edu/projects/sandburg/>

Notes:

[Back](#) [Edit](#) [Delete Collection](#)

Assessments

[Add Assessment](#)

File Manager

Bit root directory

File Groups

[Add File Group](#)

Show entries Search:

						Total file size (GB)	Total files	
52	File group 52	Digital Content Creation	Mixed Content	\\libsysdigi2\Workspace \RBML\Sandburg_RBML	PDF, DOC, JPG, DB, DS_STORE, BRIDGESORT	2.58	512	View Edit Delete
175	File group 175	Digital Services and Development	Master Mixed Content	OSLF				View Edit Delete
214	File group 214	Digital Content Creation	Master Mixed Content	Maxtor200G\Sandburg	WAV, MP3, XLS	170.978	688	View Edit Delete
215	File group 215	Digital Content Creation	Master Content	Maxtor300G\Sandburg_Materials_-_Shipment 2	TIF,JPG,DS_STORE,DB	39.72	3764	View Edit Delete
216	File group 216	Digital Content Creation	Master Content	Maxtor300G\Sandburg_Materials_-_Shipment_One	TIF, DB, DS_STORE	76.41	2523	View Edit Delete

Showing 1 to 5 of 5 entries [Previous](#) [Next](#)

Figure 1. Survey record for a digital collection queued for ingest, indicating the location of master files on a file server, the Oak Street Library remote storage Facility (OSLF), and an external hard drive.

Medusa’s collections include but are not limited to text, image, audio, and video preservation master files created for library units by its digital reformatting operations, as well as born-digital electronic records from its special collections. Newly acquired content may be ingested into a bit-level or object-level preservation store. In bit-level preservation, basic metadata records are used to keep track of what files are associated with what collections; however, the system is not aware of how those files are associated with one another to comprise objects, and all file formats are allowed. The bit-level preservation store also enables content managers to perform actions on their collections such as virus scans and the generation of technical metadata produced by the File Information Tool Set, or FITS. It is anticipated that these services will be of great use in the management of packages of heterogeneous content such as are commonly found in institutional repositories and accessions of electronic records.

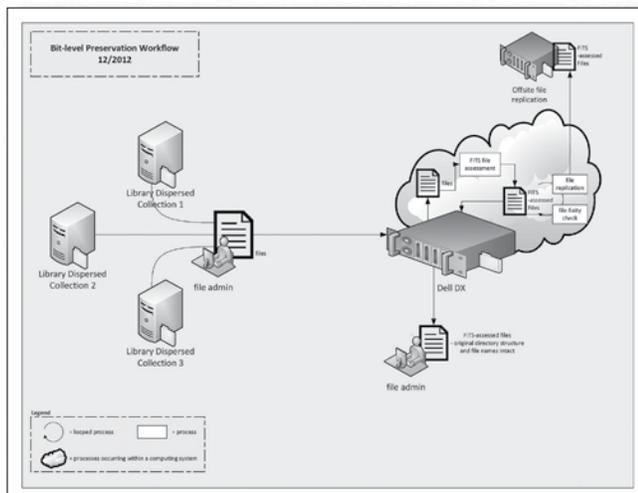


Figure 2. Bit-level ingest workflow.

In object-level preservation, on the other hand, the system is concerned with objects, or bundles of associated files that conform to a specific submission information package, or “SIP,” profile containing detailed file format requirements. The Preservation Unit is working to develop these package profiles in collaboration with digital content managers to determine how best to structure directories of content for ingest into Medusa. As these profiles are completed, Medusa’s software developers adapt their ingest tools to meet their specific needs.

On the technical side, Medusa relies on the PREMIS metadata standard to track events and rights related to digital content and MODS as its descriptive metadata stan-

dard. It is being built using the Hydra application stack of Fedora, SOLR, Blacklight, and Ruby on Rails. Hydra, so named because of its ability to support multiple “heads,” or customizable services, on top of a single Fedora repository, is seen as a potential long-term solution to the challenge described at the beginning of this article—namely, the proliferation of access repository services at Illinois without a central hub at which to exercise institutional control over the life cycle of digital library objects.

As of May 2013, the Medusa development team had implemented a functioning collection registry, a bit-level ingest feature, and an object-level PREMIS packager. All project code is available in a Github repository.² Development is ongoing, following the “scrum” methodology of agile project management, with representatives of the library’s Preservation Unit providing user specifications to programmers who work in weekly sprints to add and enhance features. The project team maintains repository policies and specifications in a regularly updated wiki.³ Medusa will play a crucial role in the infrastructure for digital preservation shared across the University of Illinois at Urbana-Champaign Library’s many departmental units and special collections repositories and is slated to go live as a service in the fall of 2013.

Notes

1. Although much administrative information is limited to authenticated system users, public collection entries may be viewed at the Medusa website, <https://medusa.library.illinois.edu/>.
2. The Github repository website is used to manage the project code, <https://github.com/medusa-project>.
3. The Medusa Wiki is used by the project team for keeping track of project policies and specifications, <https://wiki.cites.uiuc.edu/wiki/display/LibraryDigitalPreservation/Medusa+Digital+Preservation+Service>.

There is no Preservation Essentials article for this issue of the *MAC Newsletter*. Please check back for the October issue when Preservation Essentials returns!