

# Making the Pieces Fit: Integrating Preservation into a Digital Material Ecosystem

Jennifer L. Thoegersen  
University of Nebraska-Lincoln  
P.O. Box 884100  
Lincoln, NE USA 68521  
+1 402 472-4558  
jthoegersen2@unl.edu

## ABSTRACT

In 2014, the University of Nebraska-Lincoln (UNL) Libraries implemented the Rosetta preservation software to take a more proactive approach to the preservation of digital materials for which UNL Libraries are stewards. A significant part of this process was determining how to integrate this new software into the Libraries' current workflows and ecosystem for digital materials. This included considering the varied origins of digital materials; determining the purpose of collections and whether preservation was necessary and, if so, a priority; considering deposit strategies; understanding the rights related to the materials, including whether or not they should be accessible and to whom; and implementing policies that outline in which systems digital materials of certain types should reside.

This poster will illustrate UNL Libraries' progress toward implementing Rosetta in its digital material ecosystem. A diagram presenting the relevant technical 'pieces' and accompanying explanatory text will demonstrate how preservation—more specifically, Rosetta—fits into the puzzle of digital material storage, access, and management. In addition, the poster will reflect on the challenges encountered attempting to fit all of the pieces together.

## General Terms

Preservation strategies and workflows

## Keywords

Digital preservation, libraries, implementation

## 1. INTRODUCTION

The University of Nebraska-Lincoln (UNL) Libraries host a wide array of digital materials which come from varied sources and serve diverse purposes. Materials include instructional materials, research data, digitized and born-digital archival and special collections, web-based projects, and multimedia collections. For many of these materials, the Libraries have an interest in—and often mandate to—act as stewards and ensure the content's long-term preservation. While the Libraries' kept multiple backups of all digital materials, there was no active digital preservation initiative in place to ensure the integrity of and continued access to digital materials far into the future.

With this in mind, the Libraries' Data Curation Committee began investigating digital preservation options and selected the Rosetta preservation system by Ex Libris.

## 2. DIGITAL ASSET ECOSYSTEM

### 2.1 Content Streams

The Libraries both generate and curate digital content. The University Archives & Special Collections department collects ephemera related to the University, as well as prominent Nebraskans. Retiring faculty occasionally deposit their research materials with the university on separation from the University, as do University Chancellors and system presidents. The mission of the Archives is "to select, preserve, arrange, describe, provide reference assistance for, and promote the use of rare and unique research materials. The Department maintains these research materials because they are best managed separately from the general collections due to their subject area, rare or unique qualities, source, physical condition or form."<sup>1</sup> In addition to undertaking projects to digitize their unique collections, Archives collects born-digital items from other campus entities.

The Center for Digital Research in the Humanities, a joint program between the University Libraries and the College of Arts & Sciences, generates digital, often web-based research projects with varying rates of continued development. These projects will often have source material, e.g. high resolution images, as well.

UNL Libraries manages one of the largest institutional repositories in the country, DigitalCommons@University of Nebraska-Lincoln<sup>2</sup>, which includes Master's theses and doctoral dissertations, peer-reviewed journals/series, and 'works', both published and unpublished, generated by researchers affiliated with UNL.

UNL Libraries Image & Multimedia Collection<sup>3</sup>, powered by CONTENTdm, includes approximately 100 collections totally over a quarter of a million items spanning over a dozen disciplines. The purpose of collections vary from instructional to archival to historical. Most items are images, with a smaller percentage of text-based and audio formats. Files are generally of web quality, though, depending on the purpose and origin of the collection, items may have inaccessible high quality equivalents, as well. Item deposit is decentralized, occurring at branch libraries and a variety of units across campus.

iPres 2015 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

<sup>1</sup><http://libraries.unl.edu/archives-special-collections-mission-collection-scope>

<sup>2</sup> <http://digitalcommons.unl.edu/>

<sup>3</sup> <http://contentdm.unl.edu/>

UNL Libraries maintains a data repository, where researchers can deposit datasets to be managed by the Libraries and to provide public access to their data.

## 2.2 Fitting in Preservation

A major challenge in implementing a preservation system was determining how it would interact and complement or replace existing systems. Emphasis centered on what content should go into Rosetta, in what order it should be ingested, and how.

The question of what materials to include in the digital preservation system proved more challenging than originally anticipated. Many of the digital materials maintained by the Libraries are not items the Libraries necessarily have much interest in preserving. For example, a large number of the collections in CONTENTdm include low resolution images for which UNL does not hold the copyright. Instead, we are able to provide access to the campus community for instructional purposes. Working through the approximately one hundred collections in CONTENTdm and assessing the rights, the content quality, and whether the collection's stewardship and preservation falls within our mission is essential prior to ingest items into Rosetta.

The order of content to be ingested was largely determined by the following factors:

- **Readiness:** The readiness of both the collection and the relevant unit was a major factor in determining first collections to ingest. Collections needed to be well-organized and have consistent metadata to help ensure quality mapping. Units needed to be able to allocate adequate staff time to be trained on using Rosetta, assist in developing a workflow, ensure the quality of the initial ingest, and continue deposits with minimal intervention
- **Volume & homogeneity:** In order to quickly ingest a large amount of content at once, large collections that were similar enough to use the same workflow were selected. For example, Archives & Special Collections

has many collections of digitized images with consistent metadata. An initial focus was to develop a workflow to allow for regular, continued deposits with little variance.

- **Necessity:** All content selected for preservation needs to be ingested in Rosetta; however, some collections may have a more immediate need, e.g. improved accessibility.

How content would be ingested was the final and most complex hurdle to getting content into Rosetta. While the variety of content complicated the issue, the variety of ingest options was most challenging. Exploring the various setup options was time consuming especially when attempting to identify the implications for these choices, e.g. how much staff time in each department will be required both upfront and ongoing, what training will be necessary, how expansion will occur as more units begin using the system, what method makes the most sense for legacy and future data, what technical issues may hamper the workflow.

The variety of options has allowed for tailored solutions for units:

- CONTENTdm export converted to Rosetta METS packages for automatic ingest
- Ongoing single-item manual deposits
- Manual bulk deposits
- ZIP deposit with CSV metadata

## 3. DISCUSSION

Integrating a new system into an existing setup can be a challenge, especially when attempting to create new workflows to transfer data seamlessly. Pieces will not fit together perfectly, and bridges must be built to span the gaps. This step, from creation to Rosetta ingest, is the first of many along the road of long term preservation.

## 4. ACKNOWLEDGMENTS

My thanks to DeeAnn Allison, Michael Litwa, Jason Bougger, Mary Ellen Ducey, Joshua Caster, and the various individuals and units at UNL Libraries that assisted in Rosetta implementation.