A Digital Preservation Repository for Duke University Libraries

*David Chandek-Stark and Jim Coble, Digital Repository Developers, Duke University Libraries*

In 2012, the Duke University Libraries initiated a project to develop the infrastructure for a preservation repository for the digital content produced or acquired by the Libraries. The first phase of the project has focused on the output of the Libraries’ Digital Production Center (DPC), which digitizes materials for preservation and access. The preservation repository currently under development aims to pull together in one administrative system the digitized content along with its descriptive, technical, and administrative metadata. Currently, these data and metadata are spread across a number of different systems, with sometimes tenuous connections among them based on file-naming conventions and, as one consultant put it, “tribal knowledge.”

For this project, the Libraries chose Fedora Commons Repository as the leading open-source repository software believed to be best suited to meet the intended goals. Early on, the project team became interested in the Hydra Project as a means of facilitating the development of the desired repository solution. The two developers assigned to the project attended Hydra Camp at Penn State in October 2012 and subsequently began developing the first phase of the repository. A pilot project was completed in February 2013 and, as of this writing, the initial production implementation of the preservation repository is targeted for June 2013.

As noted above, one of the goals of the project is to bring together the digital content files and various pieces of metadata about them into one system. This includes pulling descriptive metadata from sources such as CONTENTdm and the library catalog (Aleph from Ex Libris) and storing it in the repository connected to the object it describes. Technical and administrative metadata are also collected from the DPC and stored in the repository in relation to the objects to which they pertain.

We are using the Hydra framework (which leverages Solr and Blacklight in addition to Fedora) to build a staff-facing administrative interface as well as a batch ingest process, a recurring fixity check process, and an image file export application intended for use by DPC to create zip files of archival image master files for delivery to patrons upon request.

The presentation that we propose will discuss our project to date, with particular attention to topics such as the following:

* Coming up to speed on the Hydra framework for two developers with no prior Ruby or Rails experience
* Modeling our content and metadata as ActiveFedora objects
* Developing a process for ingesting batches of existing digitized content and associated descriptive, technical, and administrative metadata
* Using PREMIS Events to record ingestion, validation, and periodic fixity checking
* Preserving external calibration targets for digitized images
* Developing a prototype image file export application
* Provisions for including derivatives in the repository
* Future plans