Deconstructing and Reconstructing a Repository – Strategies for Nimble Rebuilding

The University of Cincinnati Libraries has been relying for five years on a DSpace institutional repository (http://drc.libraries.uc.edu, now at over 560,000 records, and maintained by OH-Tech and OhioLINK.

While in the OhioLINK environment we have experienced four migrations – staying in the same DSpace version but migrating to the Amazon Cloud when server space was not available, migrating to OH-Tech shared infrastructure and an upgraded version of DSpace; migrating from PostGreSQL and DSpace 1.6 to Oracle and DSpace 1.7; merging two DRCs into one DSpace 1.8 installation. During the course of those migrations I came to understand DSpace export and import capabilities, how our handles worked, and what we could expect when large amounts of records were physically moved to different disk storage, and databases and collections were exported and imported.

The OhioLINK community determined in January of 2013 that maintaining over thirty institutional instances of DSpace was not sustainable for its future; at the same time the University of Cincinnati Libraries determined that we had larger goals for data management and e-science, for the digital humanities and for our university born-digital records, all of which we hoped a broader repository service could encompass. OhioLINK asked it's member libraries to migrate content away from the existing DRCs by December 2013. So in February and March of 2013 the University of Cincinnati Libraries began to put together a development team and a plan, and began to evaluate our choices for a next generation open source repository.

At the time that this proposal is being written we face the following unanswered questions:

Can we build, while assembling a new development team, within the OhioLINK requested timeline, a production Fedora instance with either a Hydra (projecthydra.org) or Drupal (Islandora) front end, capable of supporting all our existing use cases (images, text, audio, video, 'small data', dissertations, direct faculty submission, cultural heritage and institutional repository content) and sustaining the size of our repository?

Can we build a DSpace instance that has the potential to develop to meet our broader repository service goals?

Should we even settle on one platform, or instead replace our current repository platform with something similar, at the same time as we develop something new, more flexible and capable?

Can we maintain our domain name and handle server prefix (now registered with handle.net and doi.org) , with record identifiers maintained, so that the persistent URLs which now point to content are indeed persistent post-migration?

Can we build a system to replace a successful large instance of DSpace, minimizing service disruptions for our users by maintaining most existing functionality, but at the same time design a system that is does offer a path to new functionality and service models?

At the time of presenting this 24/7 talk I will be able to present our initial answers to these questions and our plan and detail, as the project manager for this development, the challenges we have met and challenges we will have yet to overcome.

_____

Linda Newman
Head, Digital Repositories and Collections
University of Cincinnati Libraries
Linda.Newman@uc.edu
513-556-1555
http://digitalprojects.libraries.uc.edu and http://drc.libraries.uc.edu