SW Arch Meeting Minutes – January 31, 2012

Agenda

- Announcements and Updates

- Status of R7.0 testing

- Review of the mss3 to RUcore URL migration

- Mini-specification for migrating NJEDL objects to production

- Process for updating legacy objects to sha-256

Announcements and Updates

Rhonda reported on a very interesting session from ALA in which Chris M. (a former SCC part-timer) discussed the Library of Congress World Digital Library. It is always good to see our alumni succeeding in the larger digital library world. Ron highlighted two items from the DC area Fedora conference that are worth pursuing – the use of Fedora journaling to synchronize content across multiple Fedora installations and the use of Amazon Glacier (cloud storage) for additional copies of our archival masters.   The cost of one month of 1 TB of storage using Glacier would be about $10. There has been progress with a VM server (aka the "sandbox"), however most of us in sw_arch have not received any communication on the status or plans. The VM server has significant potential for us to reduce development delays in moving to the next version of Fedora by always having the next version running in a VM partition. Rhonda will pursue the status with Tibor. Ron noted that the DOI implementation for R7.1 has been approved and he will consider the specification complete, with minor updates.

Status of R7.0 Testing

Kalaivani reported that there are several remaining bugs in WMS, ETD import, and batch import that still must be fixed. We also had significant discussion on how WMS should handle the ingest transaction. Yang indicated that after Fedora ingest there are four remaining tasks that must be executed including Handle assignment, SOLR indexing, and statistics update. A failure in any of these tasks can render the status of the ingest uncertain. Yang has implemented a fix that deletes the object from Fedora if there is a failure in any of these tasks. If these failures happen frequently, it will create significant overhead for people to do a re-ingest. We decided to take two actions: 1) Yang will force a failure to see what we can learn about the problem and a possible recovery strategy and 2) Jeffery will work with Peter B. to re-ingest on development an NJEDL object that experienced a SOLR indexing failure. There is a hypothesis that the indexing failure occurred because the rather large XML datastream had not been ingested although Fedora had given notice that ingest was complete.

Unless there are blockages from the testing and remaining bug fix activity, we will proceed with R7.0 according to the following schedule: February 8 – testing on development complete, February 15 –

testing on staging complete, and February 25 – release of R7.0 to the public. Isaiah will send out release notices on February 18.

Review of the MSS3 to RUcore URL Migration

Chad reviewed the specification that had been approved in February of last year. Several minor issues were raised that will need to be addressed as we undertake this task as part of R7.1 as follows:

- In the process of updating Handles to DOIs, we should also point these DOIs to the new RUcore URLs.
- We will need to redirect bookmarked "mss3" URLs to a document that explains to the user that they should use the persistent ID to cite RUcore documents.
- At some point in the future, we will set up aliases for the development and staging servers (i.e. dev-rucore and staging-rucore).

Chad will make these minor updates to the specification also note the update to RUcore URLs will be done in the R7.1 release.

Mini-Specification for Migrating NJEDL Objects to Production

Jeffery reviewed the mini-specification to migrate NJEDL objects from development to production. Dave raised the question about whether the proposed objects were possibly already in the production repository. We decided to ask Peter B. to compare the objects on development to the ones that are already on the production server but not publicly accessible. The assumption is that the objects on development might have improved metadata but we need to confirm this. Rhonda will work with Chad and Kalaivani to provide portal access to the NJEDL collection on production to facilitate the comparison process.

Process for Updating Legacy Objects to SHA-256 Checksums

Jeffery reviewed the process for updating legacy objects to sha-256 checksums. The basic procedure is to unpack the ARCH tar file and create individual archival masters for each of the files, assigning new IDs (i.e. ARCH1, ARCH2, etc) and creating the new checksum. The following special issues were noted:

- All RARCHs will need to be recreated as managed datastreams in the object with appropriate ARCH IDs.
- The filenames should be inserted in the ALT-ID attribute of the datastream.
- The new ID terminology should be used (DARCH, SARCH) for derivatives and structure maps.
- Directory trees will need a structure map.

- As part of the update process, the techMD datastream will become managed.

As an example, a book with 100 tiff files encapsulated in a tar file will result in 100 archival masters with sha-256 checksums. The techMD datastream will contain 101 blocks of metadata. All unique data in the master block (originally created in WMS) will be replicated in each block for the tiff files. In addition, size and checksum data will no longer appear in the technical metadata. This data will appear only in the corresponding attributes of the datastream. Jeffery will update the specification with the above notations and will run tests on development using various content models in order to capture the variety of structures embedded in tar files.   We were not sure whether it is better to do all the processing directly on production or copy archival masters to development.   We will review the results of Jeffery's tests before making a final recommendation as to how to proceed.

Agenda Items for the Next Meeting

- R7.0 Status
- Specification for configuring SOLR to add metadata fields without a code release
- RUetd – WMS update
- Update on DOIs

rcj – 02/11/2013

RUcore Release Summary – Specifications and Target Dates (R6.x and R7.x)

1. Release 6.2 – Faculty Deposit Enhancements – Released September 17, 2012
2. Release 6.3 (Analytic) – Released September 28, 2012
3. Releases 6.3.1 (October) and 6.3.2 (released Nov. 7)

4. Release 7.0 – Large Files, Complex Objects, Faculty Deposit (Release Target – February, 2013)
Specifications
- File Configurability - Complete
- File hierarch/structural map - Complete
- Fedora checksums – Complete
- Background ingest with alerting - Complete
- Download statistics for the Faculty Portal - Complete
- Faculty Deposit User Interface (more enhancements) - Complete
- Complex object User Interface – download zip for object or select files from a directory display(see structural map specification) – Complete
- API for OAI-PMH

Development (Code complete – November 21, 2012)

5. Release 7.1 – URL/DOI release (Release Target – TBD)
Specifications
- mss3 to RUcore migration – complete
- DOI implementation specification
- Dynamic field indexing with SOLR
- UI for landing page

6. Release 7.2 – Upgrade to 3.6.x (Release Target – TBD)
- Install Fedora 3.6.x on staging
- Testing of R7.0 on staging
- PHP upgrade
- Post-release update of "legacy" objects to sha-256 checksums

7. Release 7.5 – Faculty Services, Video, Security (Release Target – TBD)
Specifications
- Authentication/Authorization (Scenarios available)
- WMS File handling, validation and metadata extraction (with exif, media tools)
- WMS – support for external relationships
- Versionioning (file, object, and project)
- Faceted browsing
- Faculty service enhancements
  - Upload an html file in faculty deposit
  - Dynamic bibliography
  - Statistics enhancements
- MP4 container and support for Wowza
- Jpeg 2000 and page turner
- Schema for technical, source, and rights metadata

8. Release 8.0 – EAD Support (Release Target – TBD)
rcj – 01/18/2013