

SW Arch Meeting Minutes – February 23, 2012

Agenda

- Announcements and Updates
- Progress on Non-release projects
- Installation of Fedora 3.5
- Discussion of Content and Targets for R7.0

Announcements and Updates

We discussed an issue related to how we are indexing metadata. It appears that descriptive events are not identified in the SOLR schema for indexing. One result is that data objects with certain surnames are not being picked up in a search. We proposed two possible solutions: 1) index all of MODS as a “clump” of text so we would always pick up anything appearing in MODS independently of whether the item was explicitly referenced in the SOLR schema and 2) revisit the SOLR schema to determine what other elements should be included. Given that there is some urgency in correcting this problem, we decided to proceed with #1 above and include it in a proposed new dot release (R6.1.2). The second approach (#2 above) should be forwarded to MDWG for further consideration.

On another issue, we have successfully solved the xml parsing error encountered with special characters in a faculty deposit. The fix required explicit identification of the character encoding (UTF-8) to be used (as opposed to using the default). However, we are still not sure why the default mode worked on the development server and did not work on production. We’ll need to pursue this issue and make sure we have the same configurations and versions installed across all servers.

We noted that R6.1.1 is ready to go to Dave for installation on production. It should be ready for the public on Monday, February 27.

Non-Release Related Updates

The status of each project is summarized below:

- Research data. After we release R6.1.1 on Feb. 27, we should be able to proceed with ingesting new projects including primate teeth, innovation study, and Kantor’s project.
- JPE. No discussion here but we do need shibboleth to stream the JPE videos.
- Analytic. Chad had made some bug fixes and we also need shibboleth on development in order to run the Analytic software. Dave is in contact with Chuck Hedrick on the shibboleth issue.
- XML-1 Datastreams. Jeffery has tested the script on development and will hand it off to Dave next week for testing on the staging server. A second diagnostic stage requiring a new script will have to be developed to address two areas: 1) those objects with ill-formed xml-1

datastreams and 2) any further objects that might have appeared recently and do not have xml-1 datastreams.

- Jpeg Thumbnails. Jeffery indicated that the script to identify those objects that need jpeg thumbnails should be available next week.

Fedora 3.5 and RUcore Release 6.5

Given the discussion above and the proposed second dot release (R6.1.2), we proposed a timeline for installing Fedora 3.5 on the staging server. The timeline is as follows:

1. February 27 - Install R6.1.1 on production
2. Feb. 27 – March 11 – “Soak “ time for R6.1.1
3. Feb. 24 – March 11 – Fix and test SOLR indexing of MODS for R6.1.2
4. March 12 – Install R6.1.2 on production
5. March 12 – March 25 – “Soak” time for R6.1.2
6. March 26 – Install Fedora 3.5 on staging
7. March 26 – Begin testing of the R6.1.2 base on Fedora 3.5
8. April (TBD) - Upon successful testing of R6.1.2 on Fedora 3.5, release to public as RUcore R6.5

Targets and Content for Release R7.0

The R6.5 release will put us in a position to use some of the new features in Fedora and also get a lot of bugs fixed. Our next feature release is R7.0 which is focused on handling large archival masters, complex data objects, and improving faculty deposit. We reaffirmed the following:

- We’ll need to do A/A for user IDs and IPs within the RU domain
- We’ll move ahead with the MP4 container which requires a new player and new file type. Over time, we will move all the MOV datastreams to MP4. Our server architecture should remain the same so this item doesn’t require a lot of development.
- A statistics enhancement will allow faculty to see the number of downloads for a particular article and where the downloads are coming from. Chad and Jie are working on a prototype.
- We added another item to address the UI for the display that provides the “landing page” after hitting the persistent ID. We must make sure the proper metadata fields are displayed. Ultimately, per our recent CISC meeting, we want to be able to link to related events and life cycle of the data.
- A WMS item was added to address file handling and validation. This is especially important for faculty deposit where someone can upload an html file and call it a pdf.

Regarding this capability and enriching our techMD, Isaiah has proposed using the exiftool which extracts technical metadata from a variety of object types. Yang will explore for use in WMS. If all goes well, we'll use the tool in our next release.

With the above additions, a new release summary has been developed for R7.0 that also shows the schedule for reviewing specifications (Attached below).

Agenda Items for Next Meeting

- Video streaming specification - Isaiah
- Complex objects/structure map specification – Chad
- Status of Non-release projects
- Review of R7.0 release content and targets

rcj – 03/12/2012

RUcore Release Summary for R6.x and R7.0

1. **Release 6.1 - January, 2012 – SOLR/Lucene Release**
2. **Release 6.1.1 – February, 2012**
3. **Release 6.1.2 – March, 2012 (indexing of MODs text)**

4. **Architecture and Requirements**

Requirements and architecture only (available TBD)

- Jpeg2000 architecture and page turner
- Messaging and alerting (e.g. alert on ingest complete)
- EAD specification
- UI for data to dynamically create a zip for download
- Context object
- Sakai API interface

5. **Release 6.5 – Upgrade to Fedora 3.5 (Target – April)**

- Testing of R6.1.2 on Fedora 3.5

6. **Release 7.0 – Large Archival Masters (Target – TBD)**

Requirements for development (available for review – TBD)

- Specifications
 - File Configurability – Yang (March 22)
 - Structure Map to represent complex objects – Chad (March 8)
 - Multiple techMDs – Yang (April 5)
 - Fedora checksums – Jeffery and Jie (March 22)
 - Data User Interface – Ryan and Data Working Group (TBD)
- Video streaming – Isaiah (March 8)
- WMS large file support and background ingest (with user notification) - KA - user services for UI spec
- WMS File handling and validation – Yang
- Enhanced UI for landing page (from persistent ID)
- Authentication/Authorization enhancement (user IDs, IP addresses – for RU only)
- MP4 container for presentation (supports progressive and download, prepare for mobile devices, etc)
- Download statistics (at object level, view downloads and where from) – Jie and Chad
- File names using the datastream label field (KA will take to MDWG)

Requirements Complete (or not needed)

- Migration from mss3 to RUcore urls (complete)
- /rarch directory enhancements (complete)
- Djvu applet (esp. for yearbook project) – Testing only

Release 7.x (not scheduled)

Possible Features

- Support for DOIs, opaque suffixes, DOIs to specific datastreams
- Jpeg 2000 and page turner
- Object integrity tool
- Alerting and messaging service
- Read-only system
- Context Object
- Support for mobile devices
- Upgrade of handle server

Release 8.x (EAD release - not scheduled)