

SW Arch Meeting Minutes – January 26, 2012

Agenda

- Announcements and Updates
- Status of R6.1
- Non-release related Projects
- Specification for moving from mss3 to rucore URLs
- Large Archival Masters – continued discussion

Announcements and Updates

From the previous CISC meeting, Grace has asked us to proceed with implementation of DOIs. She also suggested that we move ahead with joining the DataCite organization. Grace will charter a team consisting of Chad, Kalaivani, Rhonda, Ryan, Isaiah, and Ron to address implementation and issues related to DOIs. This team will also connect with Caryn for EADs and to Laura Mullen for our published journals – specifically PCSP. (The team acronym is DOIT or do-it). There are some fairly subtle issues that will need to be addressed, e.g. the case where we ingest an article, assign it a DOI and the same article gets another DOI from CrossRef. The DOI federation recommends that the same article not have more than one DOI. In a short meeting after CISC, Rhonda, Ron, and Chad discussed actions to improve Faculty Deposit, given that Rutgers will likely embrace an open access policy later this year. Although there were concerns regarding Google Scholar, we believe most of these have been solved by pushing the personal collections out to the sitemap. We will introduce a feature that will allow faculty to view download statistics for a specific article. Rhonda will investigate reports that the user interface is still to complex.

Release R6.1 Testing Status

We are ready to move R6.1 to production. Kalaivani provided a handout of all the major updates in the release. The document now links to the bugs in software libraries that have been closed. Dave will start the production installation on Sunday evening at 9 pm. Indexing will occur overnight and we hope to bring the system up early on Monday. Developers are asked to check their email Sunday night and to convene Monday morning at TSB. This focus will help us minimize the down time window. After R6.1 is released, Rhonda will put out an email to RUL everyone providing an update on Faculty Deposit new features and Google Scholar indexing performance.

Non-Release Related Updates

We briefly reviewed the five tasks that had been previously identified, i.e. a) update the data project structures, b) move JPE to production from lefty64, c) move the Cranberry project off of lefty64, d) apply thumbnails to all text documents, and e) provide xml-1 datastreams for all text documents –

primarily ETDs. In addition, to support JPE and Analytic (on development), Dave will work with Chuck to make sure we have shibboleth working. Note, also, that handles on the development server will need to be updated. Kalaivani, Chad, Jeffery, and Dave indicated that these tasks can probably be completed by February 15. In addition to providing these much needed features, these activities should also free up lefty64 to be recycled for other purposes.

Moving to RUcore URIs

Chad reviewed the final specification that, once implemented, will guarantee that “rucore.libraries.rutgers.edu” will be a part of the URL rather than “mss3”. There were a few minor updates to the specification (e.g. the addition that sitemap urls will need to be changed). Jeffery will provide additional text to Chad for the more specific areas that he needs to change regarding showfed and the sitemap.

Large Archival Masters

We continued the discussion of how to handle large archival masters for complex objects (i.e. those with directory hierarchy and filenames that should be preserved). From our previous meeting, we had concluded that simple objects will have archival masters with explicit datastreams (i.e. no tars) and the datastream naming convention would be, for example, ARCH-PDF1, ARCH-WAV1, etc. In our discussion, we used the primate tooth project as a case study. After much discussion, we arrived at a consensus for several key issues as follows”

- As in simple objects, there will be explicit datastreams for the archival master (i.e. no tars).
- We will use the structure map to capture file names and the directory structure and to map file names to datastream IDs. For example, when a researcher requests the download of certain file types (e.g. the surface file for primate teeth) rather than the complete object, the structure map will be used to recreate the proper directory structure and restore the file names. This information will be encapsulated in a gzip and download to the user.
- It was noted that we will likely need to add technical metadata for the archival master files and therefore the WMS capability of creating multiple techMDs will be needed.
- In the case where selected files are extracted from the Fedora object (rather than downloading the complete object), we need to insure that appropriate metadata is available. In the primate tooth project, we proposed that all of the descriptive metadata be encapsulated with the gzip file.
- For many of the more complex objects, the archival datastream will be the same as the presentation datastream. If we can avoid duplication of these files, considerable storage space can be saved. The earlier concerns dealt with issues of opening up the archival datastreams, which represent the preservation copy, to the public and incurring some risk that the files might

be somehow compromised. We concluded that software that would package the datastreams in a gzip would avoid presenting the actual archival master file to the user. Everyone felt that this approach was one that we should pursue. However, we did not discuss whether there would be no presentation datastreams or whether the presentation datastreams would take some other form. (In a post-meeting discussion, Chad suggested that the presentation datastreams could take on the form of a rels-int pointer to the respective archival master datastream. So, for a primate tooth surface file of file type .sur, we would have a presentation datastream of ID=SUR1 which would point to the archival master datastream with ID=ARCH-SUR1. This appears to be a very nice way of continuing to include presentation datastreams in the object and would also likely improve performance in delivering the .sur file to the user.)

- We will need to provide a more detailed specification of the above proposals. We want to make sure that this approach can be generalized to other types of projects. Isaiah indicated that the basic approach should work for the FCP videos.

Agenda Items for Next Meeting

- Continue Large Archival Master Discussion
- Release R6.5 (Fedora 3.5.x)
- Streaming server specification
- Jpeg 2000 specification