

SW Arch Meeting Minutes – October 23, 2014

Agenda

- Announcements and Updates
- R7.5 Testing Status
- Specification for adding the REF-1 datastream
- Updates required for resources with cover sheets (label and alt-id)
- IP/netid restriction scenarios
- Content model for dark archives
- Pending
 - Process and policy for managing crons
 - Updating xacml policies (per Chad's investigation)

Announcements and Updates

We briefly discussed the issues regarding migration to Fedora 4.x. Although F4.1 will provide some migration tools, we will probably have to develop many of our own tools. We also need to discuss the possibility of moving to F3.7 or 3.8 for bug fixes and possible performance improvements. This will be a significant discussion item after the release of R7.5

R7.5 Testing Status

Kalaivani reported that testing is going quite well with 51 bugs fixed, 16 to test, and 19 still active. She also reported on the outstanding bugs in the various subsystems (SSO, dlr/EDIT, SOLR, WMS, RUFID, portal, etc.). The ongoing testing plan and schedule is as follows:

- Oct. 27-29 (Monday – Wednesday) will be a temporary halt to testing to allow for an intensive bug fix period.
- Oct. 30 – Another testing session with the expectation of completing testing on development.
- Oct. 31 – package the code and send to Dave for installation on staging.
- Nov. 4 – begin testing on staging
- Nov. 11 – one final test on staging if necessary
- Nov. 17 – release R7.5 to the public

Adding REF-1 Datastreams

Jeffery reviewed the specification for adding the REF-1 datastreams to the some 300 resources that have cover sheets. The purpose of the REF-1 datastream is to always provide a PDF that can be used to create a presentation PDF with cover sheet. After some discussion regarding the mechanics of how the script should be executed (on production or development), we accepted the specification as written. It should be noted that there will be a few resources that do not have a “clean” PDF to be used for generation of the REF-1 datastream. These resources will be updated manually with the REF-1 in the event that the presentation PDF must be edited for some reason.

Updates – Labels and Alt-id

In the process of adding cover sheets to PDFs files, we have lost the label and alt-id attributes on the datastreams. In an earlier discussion, Chad noted that he could temporarily fix the label issue in configuring the portal. It appears that this fix is in place. We will need to provide a script to automatically update the label and alt-id attributes. This led to a discussion of pending scripts to be run and the priority of each. We decided on the following in priority order:

- Add DOIs to all legacy objects with HDLs. We will run the script on all objects on development, do a test on staging and then run the script on production. Jeffery and Dave will work to do this update before R7.5 is released (by late October or early November).
- Service pack 2.0 update will be done on production as part of the R7.5 public release (Nov. 17)
- A script will be run to unpack all tar archival masters in early December (after the two week soak period for R7.5).
- Develop and run the script for updating label and alt-id attributes on datastreams (target this script for mid-December).

IP and NetID Restriction Scenarios

Ron presented the various scenarios that would involve restricting access to resources based on either IP or netid. There was lots of discussion on these scenarios and suggestions on how to update the specification. Rhonda suggested a typology of restrictions as follows: 1) by IP, 2) by netid, 3) a community restriction, and 4) search engine restrictions (i.e. Google, Bing, etc.). Also, there will be policy issues about what types of restrictions we want to support. Ron will update the document for further review in sw_arch and in CISC. At some point these scenarios should serve as requirements for implementation in R7.6.

Content Model for Dark Archives

Ron presented the proposal for a new content model for dark archives. The group accepted the proposal with the following changes: 1) we decided that a presentation file would not be necessary, 2) the readme instructions for how to restore a resource should appear in both the technical metadata and as an archival file, and 3) we need criteria for what types of resources would be considered for a dark archive. One possibility includes resources that are part of a non-RUcore project such as OJS published journals and the Eagleton polls. The possible criteria and a review of the dark archive content model are planned for the CISC meeting of November 12.

Other Items

Yang raised the issue of how WMS should retain a record of all versions of a journal article. He proposed that WMS retain a small portion of metadata for each version in order to expedite the display to the user (i.e. retrieving versions from Fedora would create performance issues). We will bring this discussion back to sw_arch after R7.5 is released.

Agenda for Next Meeting (August 14)

- Specification for unpacking archival master tar files (Jeffery)
- Process and policy for handling crons (all)
- Updating xacml policies (per Chad's investigation)
- Specification for updating labels and alt-id attributes on "cover page" resources