SW Arch Meeting Minutes – March 28, 2013

Agenda

- Announcements and Updates
- Status of R7.0.1
- Virus scanning specification
- User interface and landing page
- DOIs – implementation issues

Announcements and Updates

The Reaper exploit involved a number of us in restoring services. It appears that all services are back and we have an updated "Reaper" with only production applications. We will need to review the remaining directories and decide to retain or delete as appropriate. Chad and Jie proposed to aggregate the reporting for download of jpegs to just one generic jpeg type rather than reporting downloads for small and medium jpegs separately. All agreed that we should do this; Ron will talk with Linda who might have had an interest in this more refined reporting. If no issues, we will put the fix in R7.2

Status of R7.0.1

There is a little more testing to be done for this release (e.g. WMS upload of large PDFs). We proposed to move the release to staging on April 3 and to production on April 8. Given the minor updates, we decided that no announcements are necessary.

Virus Scanning

Isaiah reviewed a draft specification for implementing virus scanning. The proposal is to use clamAV and its associated daemon, clamd. We agreed to initially do a scan of our storage systems for development, staging, and production. This step provides a starting point which gives us reasonable assurance that there are no viruses lurking in the current file system. We proposed that this process be executed in the R7.3 timeframe (target release date of August 15, 2013).

We also agreed that clam should be used to do a virus check when uploading to temporary or work areas in WMS and FD. Isaiah will do some timing to determine what the cpu requirements are for large files. Clam provides capabilities to insert a virus into a file for testing purposes. Isaiah will update the specification to cover timing and installation issues and we will review the specification again in the next sw_arch meeting. Note that the Clam software requires daily updates to keep current on checking of all recent virus attacks. Virus scanning for upload was positioned in R7.4 (release target is December 15, 2013).

User Interface for the Landing Page

Chad reviewed the draft specification with the objective of establishing a common format (as much as possible) between the landing page (from the HDL or DOI) and the detailed result for the same resource in a search. This commonality will apply to the major thematic portals, RUetd, Faculty collections and root portals. Other specialized portals (e.g. NJDH, VMC) will continue to operate as they currently do. The "bottom line" in the implementation strategy is make the URL for a full record result and the persistent URL the same. Chad also proposed to add an additional restriction to special resources and collections by prohibiting the viewing of even

the bibliographic data. This feature might be used for the Administrative Documents collection or perhaps data resources that need to be restricted for various reasons. This feature will be included with the new landing page implementation in R7.2

DOIs – Implementation Issues

Ron and Chad revisited some of the implementation issues regarding DOIs with decisions as follow:

Declaring DOIs Unavailable. The EZID service provides a capability to display a "tombstone" page indicating that the resource is not available. There is also a capability to prevent or allow harvesting. We decided to combine these two states and will always assume "unavailability" also implies that the resource cannot be harvested. We also decided that posing the unavailability question to WMS users is likely to be confusing. Thus, the proposal is to allow the WMS metadata preparation and ingest process to continue as usual. A new function will be implemented in both WMS and dlr/EDIT that uses the DOI API to provide a "manage DOI" capability. After ingest, a user or administrator who understands the implications of making a DOI unavailable, can change the status of the DOI using the manage capability.

Using Test or Production DOIs on the Development Server. The EZID service provides a test prefix to be used during development to create temporary DOIs.   However, these test DOIs will not resolve after two weeks. The EZID service recommends that we not create production DOIs for test purposes. We proposed that the default mode on development servers will be to use the test prefix. All test DOIs will not resolve after two weeks which did not seem to cause any problems. However, developers and testers will likely want to create a few production DOIs to make sure the production capability works.

Meta-tags and Google sitemap. The DOI should replace the HDL as a meta-tag in our resources. Also, since we are moving the rucore-based urls, the Google sitemap will need to be updated to reflect this change.

Assigning DOIs to Legacy Objects. We had previously agreed to replace all the HDLs in legacy objects with DOIs. This process can be run independently of our regular software release process and is now targeted to be completed immediately after our R7.2 release. Note that update of checksums to SHA-256 is scheduled to be done right after the R7.1 release.

Other Items

In our previous meeting, Chad and Ron agreed to confirm with Grace that we will use opaque suffixes in the DOI string. Grace has, in fact, confirmed this approach.

Agenda Items for the Next Meeting (April 11)

• Virus scanning (Isaiah)

• Content models – next steps - WMS configuration, updating FD, stats, techMD (Rhonda)

• FD features to be added to R7.2 (Rhonda)

rcj – 04/08/2013