SW Arch Meeting Minutes – July 12, 2012

Agenda

- Announcements and Updates
- Review of portal statistics specification
- Progress on Non-release projects and projects requiring scripts
- DOI specification
- Fedora 3.5 on Staging and testing of R6.1.2
- Decommissioning lefty64

Announcements and Updates

Ron indicated that he has put together a definition of simple, compound, and complex objects. Although these definitions will not change our work direction, we need to be more precise in order to eliminate confusion in talking with various communities. These definitions will be discussed in the next EAD meeting and will also be reviewed in sw_arch and CISC.

Given the recent server outages, Rhonda has requested that Ashwin be added to the sw_arch list. Everyone supported this so we will add him to the list.

Object Statistics Report for Faculty Collections

Jie reviewed the updates to the statistics specification. The major updates and clarifications are as follows:

- It was clarified that viewing statistics for faculty deposits will be available in both the public RUcore portal and the Faculty/Department portal.
- The "View Impact Statistics" link will appear on both the full record and the Handle resolution (showfed) record view. Note, for implementation purposes, Jeffery and Chad will need to discuss how showfed can determine that it is displaying a faculty deposit resource. The link should not be visible for other resources.
- We decided that the "Visits Per Month" was not really useful for the general public and will not be implemented. It was noted that faculty can still get this data by viewing their statistics on the deposit page.
- For country views, we will show both the table and the pie chart. It was noted that the detailed data showing the downloads should sum properly to the total number of downloads. As a result, it will be necessary to show "Other" as a category.
- For IP address lookups, we will proceed with the pricing option of 2 million queries for $200. However, it was not clear how long the 2M queries will last. As Dave noted, doing a real-time query as opposed to post-processing will consume considerably more queries. It was

recommended that we log the queries, not only to obtain better cost estimates but also as a check on how we are billed.

Jie will revise the specification accordingly and we will consider it final.

Fedora 3.5 on Staging and Testing of R6.1.2

Dave noted that he will reinstall Fedora 3.5 next week to correct some previous problems that may be causing errors with policy statements. Other than the policy issue, testing on of R6.1.2 did not turn up any problems. If we can solve the policy issue, we should be ready to move Fedora 3.5 to development.

Non-release Projects

Dave and Jeffery reported on the status of the non-release projects. The status is as follows:

- The Voorhees rights statements have all been corrected. There was one record that required a manual update. Jeffery and Rhonda will investigate this problem to make sure the script has been properly generalized to catch these anomalies. This project is considered complete.

- The "qualifier" records have all been updated. During the process, there were several script updates required. To properly account for these updates, Jeffery will designate each script update with a version number. The qualifier project is considered complete.

- Over 1000 objects were updated with xml-1 datastreams, however there were 58 errors. Jeffery and Dave will work to resolve these errors and update the script if necessary.

- Assuming we resolve all the errors with the xml-1 datastreams, we will then proceed to updating the jpeg thumbnails.

- For the JPE videos, Dave has to check the url link and will then be able to run the script to ingest the remaining 11 objects.

In addition to the above projects, we decided to proceed with updating the NJDH rights statements. Jeffery reviewed the mini-spec and Rhonda introduced a form for these types of projects. For the NJDH project, Linda will fill out the form and we should then be able to proceed.

DOI Specification

Ron reviewed the specification for implementing DOIs, noting that there are a number of policy decisions that we need to make. After review in sw_arch, the specification will be reviewed in the next CISC meeting. The major points of agreement are summarized below:

- We agreed to pursue an approach in which we would assign DOIs to all objects. However, it was noted that this approach may cause some confusion in the situation where a preprint and the published article both have DOIs. The recommendation is twofold: 1) include a statement in the

metadata so that users know they are getting a preprint and 2) create DOI suffixes that are clearly different from those used by established publishers.

- We proposed to assign DOIs to all existing resources. This approach assures uniformity across the repository and also allows us to cease assigning Handles, perhaps paving the way to ultimately decommission the Handle server.

- We decided that it would not be advisable to assign DOIs to specific file datastreams. The typical use of a DOI is to direct the user to a landing page. We will need to investigate other approaches for providing direct access to a specific file.

- The DOI should become the persistent URL for those objects that have both Handles and DOIs.

- We agreed to further investigate the impact on Google Scholar search results. Two conditions require more investigation: 1) How does GS handle a situation where an article preprint and published article have different DOIs but the same title and much of the same metadata? And 2) are there any issues with having both a DOI and a Handle incorporated in the meta-tags of a resource?

Ron will update the specification to reflect the above points and submit it for an agenda item in CISC.

Other Items

We are on track for public release of the faculty deposit enhancements for July 30. Dave suggested we should investigate what it would mean to become Open URL compliant. We did not have time to address the "decommissioning lefty64" agenda item.

Agenda Items for Next Meeting

- Status of Non-release projects
- Status of R6.1.2/Fedora 3.5 installation and testing
- Complex object implementation specification
- Object definitions - review
- Decommissioning lefty64
- Release numbering for the next dot release
- Pending agenda items
    - RUetd – WMS update
    - Enhanced UI for the landing page (a possible framework) – Chad and Jeffery
    - WMS – creation of relationships (rels-ext) for data projects

rcj – 07/30/2012