

Minutes of December 16, 2010 Meeting

Agenda

1. Updates and Announcements
 - php script and resource index memory problem
2. Google scholar update
3. OVPR update
4. Specification for submission to NIH
5. R5.2 update

Updates and Announcements

- January meetings are scheduled for the 6th and the 20th. Chad will chair the meetings.
- Ron indicated that we are seeing about 10,000 downloads, on average, per month. This represents a steady increase over the year.

Google Scholar

In the past few weeks, we have seen marked improvement in Google Scholar (GS) results. As of the meeting, there are a little over 1200 results appearing in GS for a search on mss3. Many of the scholarly articles can be found. However, Ron did a sample by just searching for all non-RUL articles and found that he could retrieve about 50% on GS. We decided to continue with two actions: 1) Rhonda will markup several high profile articles (e.g. the Pazzani article) that are not appearing in GS to test the quick fix for added meta-tags and 2) Jeffery and Chad will work to tar up the RUcore Scholarly Resources website and associated cron and send to Dave for installation on mss3. This website uses the successful e-prints model of posting all articles on a web page, organized by publication date. We'll plan to run the cron weekly to update the website with any newly published articles. The Google ad hoc group will continue to monitor activity. From Kalaivani's reporting we should see about 100 articles and around 1900 ETDs in GS. If our success can be sustained, Ron will plan a celebration for sometime in February.

OVPR Update

We have made great progress in providing a data prototype for demonstration. There are six compound objects represented with extensive metadata. A meeting is planned for Friday to review and discuss possible additions. (Note: the list of proposed changes from the Friday meeting is attached). We have to be prepared to offer a minimal level of data service in the R5.2 release.

NIH Specification Update

Rhonda reviewed the updates required for the NIH submission. There were only minor changes, however more discussion will have to take place off-line regarding the more advanced report generation capabilities.

Release R5.2 Update

Based on critical path items in the analytic and WMS development, we decided to move the code complete data to February 28, 2011 and the release target for R5.2 is mid-April. Ron will review the release with Grace.

Resource Index

Sho reported that a bug in the Fedora resource index (actually a Java bug) results in the resource index never releasing memory. There is a planned fix in Fedora 3.5 which is at least a year away. Nevertheless, we have to move ahead with using Fedora relationships in analytics and in the compound object. Chad will produce a simple script that adds and deletes relationships so we can get some additional data on how quickly the problem develops. For now, our only work-around is to periodically re-start Fedora.

Agenda Items for Next Meeting

- Google Scholar update (Jeffery)
- Additions to lefty64 data prototype for OVPR demonstration
- Compound data objects for R5.2 (Chad, Jeffery)
- Status of shibboleth on mss3
- OVPR update (Kalaivani)
- Pending items
 - Object updates – xml-1 for ETDs and thumbnails for PDFs
 - Planning for JPE migration
 - Add xml-1 datastreams for many ETDs
 - Djvu java applet
 - Performance of handle server

rcj – 12/22/2010

Attachment – Proposed Changes to the Data Prototype

Review of Data Prototype on lefty64

1. Indexing of descriptive event. It looks like the mods in these data objects is being indexed but somehow now searched in the portal.
2. Merge tabs – documentation and supp. materials – into a single category called “Project Documentation”. Note that at least one object will have to be changed – Innovation has a document under supp. materials.
3. Add “Back to Search Results” on the portal screens.
4. A handle to go directly to pdfs. I think this is a fairly easy update for WMS/Faculty Deposit. For example, all articles coming from faculty deposit could be assigned an additional handle (call it pdf-handle or article-handle). The handle could be created at ingest time.
5. In the full record, the citation url should not be clickable. Do we need more discussion on this?
6. We will use Project Title rather than Title; Display Project Title before Dataset Title.
7. Regarding the display of “PDF” in showed, we could easily display the content model. So, for articles, one would see “Manuscript” rather than PDF.
8. We should limit search by audience, file type, etc. (I didn’t get all of these).
9. The word “Collection” will be removed from all metadata.
10. Regarding our shibboleth implementation, I believe it would be fairly easy to put up an RUcore page that defaults to Rutgers University. We could also remove the search box. BTW, the search box is totally useless. You have to search on terms like “ccm” for Morris county which no user will figure out.
11. I’ll add an abstract to the Iowa object.
12. Brief record display: add Date Created, Extent (variables, records etc.), Format, file size.
13. Full record display: Display Order: Project Title, Dataset Title, Date Created, File size, Target Audience (Domain), Description, Genre, Citation URL, Media Types, Rights Statement.
14. Full record display: Remove Org. Name, Remove Language; Change Descriptive Event label to Related Publication.

rcj/ka – 12/20/2010