

SW Arch Meeting Minutes – March 26, 2015

Agenda

- Announcements and Updates
- ABBYY FineReader
- REF-1 datastreams
- HDL Audit – next steps
- Review of R7.6 Targets
- Fedora 4.x discussion

Announcements and Updates

Ron mentioned that Grace along with other Co-PIs has submitted an NSF grant to provide an RU data network with a discovery layer based on linked data. Regarding a previous discussion on updating the “creator” field in DataCite, he has refreshed the DOI metadata for the research data projects (collection objects) and, after some 48 hours, the changes are reflected in the DataCite metadata store. The group did not want to use the available script for updating the creator metadata so Chad volunteered to do the update manually. Ron will ask Jeffery to send the list of IDs to Chad. Regarding the migration of techMD to “managed” datastreams, Jeffery has posted a note to the list but no response as yet.

ABBYY FineReader

Isaiah reported on his investigation for using ABBYY for OCR and generation of PDFs. There are two possible licensing options: 1) a page-per-year license with a maintenance fee and 2) a total page count license. With either license, we can put the software on as many servers as need be (some of these might our MAC workstations). Isaiah is waiting on a response from the vendor regarding PDF generation. In addition, for our next meeting, Isaiah will prepare two prices options, a low price (maybe for 150K pages per year and a higher priced option that would give more pages or more flexibility. Kalaivani will inform us regarding the maintenance costs of the current AdLib software which will figure into the final cost calculations. We also decided that there does not appear to be any open source solutions that are worth pursuing.

We will need to take the result to CISC with a clear explanation of why we need to move ahead with an improved OCR software solution (basically the current server is on its last legs and we can't update the OCR software).

Adding REF-1 Datastreams

Jeffery reviewed the script for adding REF-1 datastreams to article deposits in order to facilitate the generation of cover sheets. The REF-1 datastream is a version of the PDF without the cover sheet. Jeffery noted that there are two classes of objects that need to be updated: 1) a small set of objects that only have a PDF with a cover sheet and 2) a larger set that has an earlier versioned PDF without the cover sheet. The script is designed to add the REF-1 datastream to the second class of objects whereas the first class will have to be updated manually. Jeffery will proceed with two actions: 1) test the script on development and 2) produce two lists that identify the set of objects in both classes.

HDL Redirect – Next Steps

Chad has the script ready to go and will send a link to sw_arch so we can do some random testing. If no problems, we will ask Dave to put the software in place on production.

Review of Release Summary and Targets

The updated release summary now reflects the addition of the dot release (R7.6.1) which includes the ROA submission and netID/IP restriction capability, both of which were moved out of the R7.6 release. Specific target dates for R7.6 and R7.6.1 were also added. In addition, we agreed to add a general “SOAR Support” item in anticipation of more feature requests for the OA initiative.

In the discussion of Fedora 4.x, the group proposed that we add two additional activities to the release summary. We proposed adding a study to determine if we should migrate to Fedora 3.8.x. This activity was designated with the dot release R7.6.2 (although it should be noted that there is no software delivery for this dot release). We also added an activity in R7.7.1 for a pilot study to examine the possibility of migrating to Fedora 4.x.

These changes will need to be reviewed in CISC.

Fedora 4.x Discussion

In preparation for the DC area Fedora Conference, we began the discussion on issues related to migrating to Fedora 4.x. We all agreed that Fedora 4.x is totally new and it is not clear as yet what the advantages will be for RUcore. The data model will be quite different requiring significant software changes. Also, we will require software utilities to migrate existing content to the Fedora 4.x platform. We will continue this discussion in the next meeting after everyone has returned from the conference.

Agenda for Next Meeting (April 9, 2015)

- Converting techMD to “managed” – update
- Adding REF-1 datastreams - update
- Update on ABBY FineReader software
- Move ahead on the HDL redirect
- Fedora Users Meeting – Feedback from the meeting
- Pending
 - Unpacking tar files
 - Updating xacml policies (per Chad’s investigation)
 - Revisit/clarify our quick fix policy (do quick fixes get a dot release?)
 - Review all PDFs in RUcore to determine if they have restrictions.