

SW Arch Meeting Minutes – April 9, 2015

Agenda

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

Announcements and Updates

Ron provided an update on discussions with Manish regarding our support for big data and specifically the Ocean Observatories Initiative (OOI). We will conduct a prototype experiment in April to create a resource that points to raw sensor data that is housed in the CORE building on Busch. This prototype will demonstrate the ability to assign DOIs to large datasets that are stored externally to RUcore. In addition, we will begin exploring how we can create and assign DOIs to dynamic subsets of data that are extracted from these underlying very large datastreams. This work represents a major new initiative for the Library and RUcore to provide services for very large datasets.

Kalaivani has indicated that most of the incorrect date formats have been corrected, including all dates for NJEDL. There is ongoing work to update those objects (typically collection objects) with no dates. We will run another report in May after these updates are complete.

Due to an illness, the agenda items for migrating techMD to “managed” and adding REF-1 datastreams will be discussed in the next meeting.

HDL Redirect – Next Steps

Per Chad’s earlier request, there has been some basic sanity testing of HDLs on the test server and no apparent problems have been detected. We agreed to move ahead with packaging the redirect software and installing it on production. Chad will package it and prepare a README file for Dave. Dave will check it out with basic testing and arrange for the appropriate domain name changes. We will also take the existing HDL server off line. (In a subsequent email, Grace has provided the approval to actually decommission the HDL server).

ABBYY FineReader - Update

Isaiah provided an update on obtaining price quotes for purchasing the ABBYY software, noting that there will be an increase for creating PDF files from tiff files. Given the maintenance costs for the current AdLib software, we can probably save money by upgrading to ABBYY.

We have not yet determined which OS to use for the ABBYY software, although Windows is certainly a candidate. On a related issue, Kalaivani noted that the current PDF software is running on Windows 2003 which is no longer supported. Our plan is to not upgrade this software and just move it to

Windows 2012. Our plan is that we can continue to use this software with the expectation that we will hopefully soon move to ABBYY. Isaiah will prepare a report for the next CISC meeting.

Release Summary and Targets

We briefly discussed release targets and the addition of two migration studies – to Fedora 3.8.x and to Fedora 4.x. Per our last meeting, we also added an item into R7.7 for additional SOAR features to support the RU open access initiative. Ron will make minor updates and plan to review the release summary in the next CISC meeting.

Fedora 4.x Discussion

We discussed experiences from the recent DC area Fedora Users conference. Ron noted that the Portland Common Data Model seems to be gathering momentum and is a model that we should examine for RUCore. Both the University of Maryland and the University of Virginia are approaching the migration to 4.x in different ways. The UMD will actually begin production on 4.x with new content and are not yet attempting to migrate from their Fedora 2.x installation. However, Virginia is providing a federated search capability in 4.x to enable searching across multiple 3.x installations. Everyone seems to be dealing with many of the same problems (PIDs, data model, migration, etc.). Of note, Fedora 4.1.1 has been released and there is a FOXML migration utility available with this release.

The group noted that moving ahead with 4.x makes sense for RUCore in that we remain part of the Fedora community. In addition, 4.x offers new capabilities (linked data support) and considerable more flexibility. However, we all acknowledged that there will be significant work ahead to develop the 4.x data model and redesign our software to accommodate this model. Regarding the data model, we will need to enlist MDWG to provide guidance on how or if we preserve the METS approach that we have used in RUCore.

Agenda for Next Meeting (April 23, 2015)

- Converting techMD to “managed” – update
- Adding REF-1 datastreams - update
- Update on ABBY FineReader software
- 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.