SW Arch Meeting Minutes – April 23, 2015

Agenda

- Announcements and Updates
    - techMD conversion to "managed"
- REF-1 dataststreams
- ABBYY FineReader
- HDL Audit – update
- Code freeze status

Announcements and Updates

Ron indicated that the five pilots for migrating to Fedora 4.x are complete with results available on the Fedora Wiki. Several of the custom pilots appear to have object models similar to our (e.g. inline MODS metadata) so we should be able to learn from these pilots. Kalaivani has indicated that there have been major updates to objects (e.g. for the Tombstone collection) to provide dates for these resources. We will need to run the script again soon to move the DOIs from reserved to public. Jeffery continues to pursue how we might get the Wilpur script to operate on our inline metadata in order to convert to "managed". He has contacted Chris via email.

Adding REF-1 Datastreams

Jeffery indicated that there are 402 objects in the SOAR portal. Of these, 30 do not have coversheets; Jeffery will forward the IDs to Rhonda and Kalaivani. Of the remaining 372, Dave's analysis indicates that 79 have REF-1 datastreams and do not need any update. The remaining 293 fall into two classes: 1) those that have a cover sheet and an earlier PDF from which the REF-1 datastream can be constructed and 2) those that have a coversheet but no earlier versioned PDF. Jeffery and Dave will work to run the script to repair those in the first class. The second class will need to be updated manually. Jeffery will update the reports section in software libraries.

ABBYY Recognition Server

Isaiah reviewed the specification for updating to the ABBY Recognition Server. Suggestions for updating the document included the following: 1) provide an A:B cost comparison with the current AdLib software to demonstrate the cost savings, and 2) add wording to demonstrate the urgency of moving to this new software package. There are a variety of pricing options; the group did not voice a preference for any specific option. Isaiah will present at the next CISC meeting and we will then proceed to next steps.

Code Freeze For R7.6

All developers can meet the code freeze date for R7.6 of April 27. Kalaivani urged everyone to review the some 35 outstanding R7.6 active bugs and reports in software libraries. The bugs need to be fixed by April 27 or moved to another status. After discussing vacation conflicts, we decided to move the public release target from June 15 to June 22.

HDL Redirect Update

In an earlier email, Dave reported the following: 1) the CNRI handle server running at hdl.rutgers.edu:8080has been shutdown, 2) in its place the new handle.server software was put in place. and the Apache configuration file (hdl.rutgers.edu.conf) has been adjusted to point to the new handle.server web directory, 3) the redirect to hdl.rutgers.edu:8080 was commented out, 4) the AllowOverride was changed from limit to ALL so that the .htaccess directives could be processed, 5) the previous hdl directory was renamed to hdl_old and the Apache Rewrite module had to be activated.

He will leave this configuration in place until next week to see if there is any need for adjustments, before moving it over to rep-prod and having the DNS entry switched. Dave also discussed several options for continuing to run the OCR software until we can replace it with new software. For example, he might try moving the binary files to another server.

Other Items

Anomalies for Page Count Analysis. Chad reported that there are a number of anomalies that surfaced from the page count analysis. Some PDFs don't work, there are still some objects with only a djvu file for presentation, and there is a large group of objects that apparently have an incorrect MODs relatedItem. Also, the collection identifier is missing the "rucore" prefix in certain instances. Kalaivani and Chad will investigate further.

WMS and Versions. Yang described three issues that need discussion and will not be fixed in R7.6. The issues include the following: 1) submitting a version to RUcore that is already in the repository, 2) deleting a genre term that is tied to a version, and 3) how do we handle file names for resources that originate as a URL. We decided to discuss these issues in the next meeting and we will position them in either R7.6.1 or R7.7.

Updating to SP3. Dave proposed that we not update to SP3 when we release R7.6. Everyone agreed and Dave will propose a date for the SP3 update.

Development Process. After some discussion on the process for releasing software to Dave (in light of the addition of the development server), we decided that the process can continue as before. In particular, developers would need to manage the synchronization of the development and test severs. We do need a document that describes this process.

Agenda for Next Meeting (May 7, 2015)

- Converting techMD to "managed" – update
- REF-1 update status
- WMS discussion
    - Resubmitting the same version of an article
    - Genre
    - Filename for a URL resource (where to store the URL)
- Page count anomalies (e.g. MODs relatedItem)
- OCR options until we purchase new 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.
    - Specification for managing software on development and test servers.

rcj- 05/04/2015