SW Arch Meeting Minutes – October 10, 2013

Agenda

- Announcements and Updates
- R7.2.2 - Yearbooks
- Journal article upload (looping) - bugfix
- User Account scenario - data
- OCR, JRA publish – proposal
- Mini-spec for metadata updates

Announcements and Updates

Ron indicated that the DataCite metadata extensions are under review by MDWG. These relatively minor changes will enable Thomsen Reuters to index all of our data that has DOIs. The EAD specification will be reviewed in the working group next week and, assuming agreement, will be brought into sw_arch. The Roman Coins portal is now on production along with the 35 coins that were transferred from development. Kalaivani indicated that she created a task in software libraries (assigned to Isaiah although many will participate) to track and update all of the various software package we use with RUcore.

R7.2.2 Release – Yearbooks

We are pretty much ready to go with Yearbooks. Testing on staging will be complete on Friday and Dave will install the release on production Monday morning (Oct. 14). Peter will need to ingest some 15 Yearbooks once 7.2.2 is available. This should put us in good shape for the Cabinet demonstration on Tuesday (note, it appears that the Cabinet meeting was subsequently canceled).

Journal Upload Looping

We have an intermittent bug in which a journal article deposit appears not to complete properly and therefore does not send out notifications that a deposit has occurred. Rhonda estimates that of some 20 deposits it appears that this problem occurred on with maybe 15 deposits. However, it appears that all files have been uploaded and subsequent ingests have been successful. Various tests with different content, different wired/wireless connections, different platforms, and network connections within and external to RU have not yielded any clues as to the problem. We agreed upon a quick fix release (R7.2.3) in which we will post a comment to the user that we are working on this problem and they need not re-submit the article. Jie and Rhonda will work out the language of the comment. Also, Ron will deposit one of his older articles to see if the error re-occurs.

We will also add one more fix to R7.2.3 which involves a change so that the 035 tag is included on an ETD MARC export. Jeffery has already completed the fix and will include it in R7.2.3. Since we will allow R7.2.2 to "soak" for two weeks, the code freeze date for R7.2.3 should be October 28.

User Account Scenario - Data

Aletia reviewed the scenario for the CKT-E project. A passworded site has temporarily been set up to stream videos using Wowza. Going forward, the CKT-E project will require the ability to annotate videos using the Analytic and to provide proper authentication and authorization for those researchers that are not at Rutgers. From an earlier review of our A/A approach, the CKT-E security requirements should be handled in our release R7.3 (Fall, 2013). The RUcore release R7.4 will include integration with the Wowza server and the update to Analytic to work with Wowza streams. It appears that the release R7.4, available in the Spring, 2014, should meet the needs of the CKT-E project. The process for determining access permissions for these resources would provide the input for our A/A system to properly authorize those users who are granted permission.

OCR, JRA Software for OCR

Isaiah reviewed a proposal to address lingering OCR problems. Our current situation involves the use of client software that is no longer compatible with newer operating systems. In addition, the pipeline OCR has both performance and timeout problems. Yang indicated that the OCR software is now run in the background and he also has updated PHP timeout values. Yang and Jeffery will run a test on a 200 page document. For going forward, Isaiah proposed the purchase of 3 licenses for ABBYY FineReader which is compatible with multiple platforms and handles OCR for many different fonts. The licenses can be variously configured to serve both client and server needs. Isaiah will update the proposal to include a configuration and we will proceed to purchase (funds required, about $1200).

Mini-Spec for Updating Metadata

Jeffery reviewed the mini-spec to do metadata updates that had been requested through the software libraries website. The updates are required to remove the "uniform" attribute from titles and, for photographs, to change the genre from "NJDHFormat" to "AAT". Jeffery outlined the three steps to be taken each update and he will notify Dave when the scripts are ready to run.

Agenda Items for the Next Meeting (October 24)

- DataCite Metadata Extensions (Ron)
- Retroactively assigning DOIs to Objects (Jeffery)
- Review of version specification to identify components that can be delivered prior to R7.5
- Integrity Report/Validate Object (Jeffery)
- Decommissioning Djvu
- Content model updates – status.
- Checksum failures on four videos

rcj – 10/21/2013