

## **SW Arch Meeting Minutes – August 29, 2013**

### **Agenda**

- Announcements and Updates
- Multiple copies of articles
- R7.2.1 Release
- Specification of the Yearbook page turner
- User accounts – scenarios
- Mini-spec for updating legacy object to sha-256

### **Announcements and Updates**

We had a full agenda with significant detail in the specifications. There were no specific updates or announcements that were offered.

### **Specification for Multiple Copies (mostly articles)**

Jane reviewed the specification for how we should handle multiple copies of articles in WMS and RUCore. These notes will cover some of the major changes. Jane will update the document and we will finish the review next week. Changes are summarized below:

- The most current version will be accessible to the public and there will only be one set of metadata.
- Articles will be versioned using the Fedora versioning process, thus previous versions of the PDFs and metadata will be retained along with the Fedora audit trail.
- All versions will be encapsulated in a single Fedora object.
- The submission process will be mediated when an author wants to submit a second or succeeding article.
- A descriptive event will be created to relate the current version to previous versions. Note that there will not be an associated relationship (i.e. the Fedora rels-ext) created since the versioned article is not in a separate object.
- All previous versions will be embargoed. We did not discuss the technical details of how this embargo will be implemented.
- All versions will have DOIs. Note that this assumes that we will implement the ability to assign DOIs at the file level. This capability, along with versioning, is currently scheduled in RUCore R7.5.
- In versioned articles, the DOI will be directed to a page that indicates that this article can be requested from an RUCore administrator.

There were many related metadata discussions. Because of time limits, we could not discuss the complete document. Jane will produce another draft with updates highlighted and we will continue the discussion of the specification in the next meeting.

### **R7.2.1 Bug Fix Release**

The proposed bug fixes are complete and ready for test. We should finish testing on Friday, August 30 and move the release to staging on September 3. Testing on staging will be complete by Sept. 6 and our target is to move the release to production on Sept. 9. Everyone who is doing testing is urged to update the status in software libraries.

### **Page Turner for Yearbooks**

Chad reviewed the specification for the page turner and also offered a prototype for review. The page turner uses the open source Internet Archive BookReader application. The primary architectural changes are highlighted below:

- Jpegs will be created from tiff files.
- The Yearbook object will have datastreams for each individual jpeg image. This allows ease of replacement of a single page, e.g. we might want improve the images of some select pages. This approach also offers performance enhancements since jpegs don't have to be recreated in real time.
- In addition to the XML-1 datastream that carries the text for full text indexing, another datastream (OCR-1) will be created that has the coordinates for each term in the text. This approach enables highlighting of the search term on the page. Chad suggested that we might want to merge these two datastreams in the future.
- The landing page for the yearbook will have two links, one that is labeled "read online" and takes the user to directly to the page turner. Another link would allow download of the PDF, however there is still some concern about the time it will take to download these very large PDF files.
- We will not record a download for each individual jpeg that is displayed by the user. There didn't appear to be any statistical value to doing this and would bias the presentation of download statistics for RUcore.
- The page turner will be used on all objects that have type=text and have jpeg datastreams. At this point, the objects that pass this filter in RUcore are primarily yearbooks although there may be other objects that we are not aware of at this time.
- In terms of workflow for yearbooks, all datastreams (i.e. tiffs, jpegs, XML-1, and OCR-1) will be prepared offline. There may be opportunities in the future to automate some of this activity.

We will plan to release the page turner in R7.2.2. Chad indicated that code complete will be accomplished by September 23. We should be able to push the release out to the public several weeks later, i.e. by mid-October.

### **User Accounts and Scenarios**

We did not have time to discuss scenarios and this item is scheduled for the next meeting in which we will discuss scenarios for faculty deposit, the Optimality Archive, and the CKT-e project.

### **Other Items**

Kalaivani noted that the MDWG will be examining issues with entering dates, specifically to address the problem of entering B.C.E. dates, satisfying the DOI criteria, and obtaining proper date ordering.

### **Agenda Items for the Next Meeting (September 12)**

- Multiple versions of articles - continuation (Jane Otto)
- User account specification with scenarios (Rhonda and Aletia) - continued
- Discussion for Grouper Jr - or Guppie or LASY (Yang)
- Mini-spec for updating legacy object to sha-256
- Pending
  - OCR investigation (Isaiah)
  - Repair 4 video signatures

rcj – 09/10/2013