

# Software Architecture Meeting, Thursday, January 6th, 2011

Agnew (Guest), Ananthan, Beard, Geng, Hoover, Marker, Mills (recorder), Nakagama, Triggs, Womack (Guest), Yu

## Agenda

---

### 1) Updates

- Google Scholar update (Triggs)
- Status of shibboleth on mss3 (Hoover)
- OVPR Update and additions to lefty64 data prototype for OVPR demonstration (Ananthan)

### 2) R5.2 Specification Compound Data Objects (Mills, Triggs)

- Expected outcome is review, revision and approval of specification.

### 3) Object updates – xml-1 for ETDs and thumbnails for PDFs (All)

- Expected outcome is identifying the issues and assignment of related tasks

### 4) Planning for JPE migration (All)

- Expected outcome is developing a strategy for migration

### 5) DjVu Java Applet (Beard, Triggs)

- Expected outcome understanding the requirements and assignment of related tasks

### 6) Performance of handle server (All)

- Expected outcome is understanding of the issues involved

## Agenda Items

---

### 1) Updates

**Google Scholar Update** Triggs updated the group regarding Google Scholar indexing. As of this meeting 1330 items can be found in Google Scholar. Over the last month the number of items found in Google Scholar has increased from 145 to 1330. Approximately 2000 items are expected. Progress will continue to be monitored. Hoover mentioned some issues that were had with running the scripts to generate the ePrints type Google Scholar pages. Issues include UTF-8 errors and item count mismatches. Triggs will review the issues, adjust the scripts and update the group when a new version is complete. Also a report of the records without a dateCreated value will be generated and shared with the group. Those records without a dateCreated value need correction, or they cannot be included in the ePrints type Google Scholar pages. This will remain a reoccurring topic on the agenda.

**Status of Shibboleth on mss3** Hoover reported the Shibboleth daemon is running on mss3. He was able to complete compilation, although he was not completely satisfied with the process. He will investigate those issues when time permits. The next step is for Hoover to contact Charles Hedrick and have the mss3 machine added to a Shibboleth configuration at OIRT. Once that step has been completed testing of Shibboleth will take place on mss3. This will remain a reoccurring topic on the agenda.

**OVPR Update** Ananthan and a group met with OVPR to demonstrate the progress of the RUCore Data Portal. Some changes will need to be made to the prototype Data Portal such as; metadata elements to display, labels, etc. How the portal might interoperate with the data website OVPR is creating has yet to be determined. This will remain a reoccurring topic on the agenda.

**Other Updates** The group established that minutes will be taken by all members. Order is determined alphabetically, starting with Mills. Nakagama will be responsible for taking minutes at the next meeting.

## **2) R5.2 Specification Compound Data Objects for R5.2**

The specification that was drafted and sent to the group was reviewed. Grace Agnew and Ryan Womack both joined the group for the discussion. While stepping through the specification some major issues that were brought up were the following:

How do we aggregate projects components into an aggregate for a download feature at the view object level? One option is to aggregate the components manually in the first iteration of the feature. In a future version create a feature that aggregate automatically. What do we store in this aggregation? The thought that metadata associated with the components should be included along with a manifest, or map, of the aggregation. Where should the aggregation be stored, in Fedora terms? Two solutions we suggested; store the aggregation as a datastream with the View Object or store the aggregation in a separate object all together and using relationship services associate that “aggregation object” with the View Object it aggregates. For the View Object if we create that object like a collection object, but with datastreams, should we create a new Content Model for that object or modify the collection objects Content Model?

The specification was not completely reviewed. Ananthan expressed that no matter what solution is finally introduced it must be easily understood by the end user. Beard also pointed out that if the aggregation is to be a .zip file a zip format needs to be determined, he will forward his recommendation to the group. Mills will take all comments and proposed solutions and edit the specification and bring a revised draft back to the next meeting for further review.

## **3) Object updates – XML-1 for ETDs and thumbnails for PDFs**

An issue has been found where objects that should have XML-1 OCR text layers do not. Not much was known about the cause of missing XML-1 datastream/OCR text layers for objects with PDF's that do indeed have a text layer. Ananthan and Yu will investigate all possible scenarios and report back the group their findings. Also the creation of the text layer by the PDF will be tested. Triggs will create a script that checks ALL objects in Fedora for a XML-1 and PDF datastreams under the condition that it has a mods:typeOfResource = 'text' in the metadata. The purpose is to determine how many objects need revisiting and correcting. Thumbnails for PDF's were not discussed and will be discussed at the next meeting along with the findings from the reports mentioned above.

## **5) DjVu Java Applet**

Beard reported on a LYRASIS email that circulated reporting the Internet Archives potential discontinuation of DjVu as a supported format. At the request of the Cyberinfrastructure Steering Committee, Beard will be drafting a discontinuation plan and presenting it to CISC for approval. The plan will then be discussed by this group. The major goal is to remove DjVu as a presentation format in RUCore. It was recognized that removal might be simple for documents, where a replacement format is not needed. For maps and large images a viable replacement format will need to be approved and implemented, JPEG2000 is the likely option. In the meanwhile DjVu will still be used for maps and large scale images. To support the continued use Triggs will implement the DjVu Java applet for R5.2. The applet will allow users to view DjVu without installing a browser plugin.

## **Next Meeting – Thursday, January 20<sup>th</sup> 10AM**

---

- 1) Updates; Google Scholar, Shibboleth and the Data Portal
- 2) Continued review of the R5.2 Specification Compound Data Objects
- 3) Continued discussion of Object updates – XML-1 for ETDs and thumbnails for PDFs
- 4) Planning for JPE migration (All)
- 5) Performance of handle server (All)