

Software Architecture Meeting Minutes - February 17, 2011

Attending: Kalaivani Ananthan, Isaiah Beard, Jie Geng, Dave Hoover, Rhonda Marker, Chad Mills, Sho Nakagama, Jeffery Triggs, Yang Yu (recorder).

Agenda:

- 1) Updates
 - Google Scholar (Triggs)
 - Data Portal update (Mills)
 - Patches (Triggs, Hoover)
 - R5.2 Development Status (All)
 - PDF Server, related to OCR of PDF files (Mills)
 - FFMPEG compilation
- 2) Object updates, XML datastream report from development server
- 3) Planning JPE Migration (All)
- 4) Handle server (Nakagama, all)
- 5) Handling modification requests (Ananthan, Mills)

1. Updates

- Google Scholar

Google scholar has 60 new objects and appears to be downloading the site maps frequently. The browse site is not yet linked but CM and RM will put the link in place today (2/17). We are approaching the 6-8 week window during which Google said they will completely refresh our Google Scholar indexed items.
- Data Portal update

DH applied the SPSSPOR patch. All objects have been moved over from lefty64 to mss3. KA has updated file policies in WMS on production to handle SPSSPOR files.
- Patches

SN will do some final testing of the book disseminator patch on lefty64 and on staging machine.
- R5.2 Development Status

The code freeze date has been finalized. It is March 21 (Monday).
CM is working on exhibit and annotation (due for NJVid by March) and progressing well.
RM mentioned that we will no longer be developers for NJVid after March 31st.
JT has completed new ShowFed application which resembles the RUcore search interface. It is not renamed yet, but ready for testing.

JG is doing well and the modules she is working on for release 5.2 will be delivered on time. YY proposed that before the testing for 5.2 starts, he will offer an information session for testers to introduce the changes made in WMS.

- PDF Server, related to OCR of PDF files

The problem with production PDF server is that it is not generating OCR text from PDF file. This has not been resolved. The testing PDF server, on the other hand, is functioning correctly. Nick Gonzaga has issued a ticket to adlib software to fix the problem. SN will work with NG to make sure that production and testing PDF servers are synchronized, including windows OS. As an interim measure, the production environment will use the testing PDF server for now to allow OCR text be generated. CM will provide instruction to DH on how to make this happen.

- FFMPEG compilation

DH reported that ffmpeg will not compile on solaris. DH and SN believe that this is probably hardware related issue - the software may be more linux oriented and solaris lacks the support for the specific hardware needed. SH suggested to disable most of the features that are not used and see if it will compile. CM indicated that this is a crucial piece of software to make the annotation application function. If the compilation still fails, we may have to look for alternative solutions.

- Other updates

DH reported that Tibor Purger is working on the RFP actively for purchasing linux hardware to replace solaris for RUcore infrastructure. Hopefully, it is close to going forward.

Releasing files on production server. DH indicated that when the system hits a high watermark of storage, it begins to move certain files to offline storage with its own algorithm. 22,266 files from workarea with 350GBs of data, have been released. What would happen when application calls the file stored offline is unknown. Reloading may be required, but we don't have room to stage everything back. Cleaning up workarea is needed. The problem is we don't know which projects need cleaning up. KA will look into this, but she also indicated that better coordination of the project status between SCC and TAS is needed. DH will do a report on the files that have been released.

RM reported that there will be some end-of-year money left, potential for some hardware/software upgrade or maintenance.

2. Object updates, XML datastream report from development server

For objects with PDF but have no OCR xml datastream, JT has a script to report them on lefty64. He also had a website setup for this purpose. A rough estimate is there are about 1000 objects having this problem. JT will send the link to the report website to sw_arch. The impact of reindexing when conversion script is run is unknown, but needs to be evaluated. The conversion script needs to be delivered to DH with detailed instructions. The report script needs to be run before conversion script.

3. Planning JPE Migration

DH needs to call Chuck about the Shibboleth installation and configuration. Migration document needs update, possibly JT's document about data migration could be edited and reused for this purpose.

4. Handle server

Some major issues with our current handle server are: slowness due to its use of file based database (prelease), master passphrases are unknown which causes a number of related problems, total dependency on the communication with UVA, handle phrase is too long, and possibility to expand the handle use to individual datastreams. SN proposed to install a new handle server, which uses mySQL database. There is also some discussion about whether we really need handle server. This topic needs more careful thoughts and further discussion.