

Software Architecture Meeting - Minutes, Thursday, January 10, 2008

Agenda

1. Quick updates
 - a. PDF server
 - b. Google indexing
 - c. Statistics reporting
 - d. Audit of handles
 - e. Audit of signatures
2. Further discussion of video server (postponed to next meeting)
3. Other outstanding R4.5 issues
4. Release planning (next meeting)
5. WMS open-source availability for NUL

General Announcements

- Yale is organizing a regional Fedora users conference to be held in July. Ron is on the program committee.
- By request, Ron will consult with Northwestern on issues of Fedora and PREMIS. Trip to NWU will occur in late February.

Updates

- PDF server.
 - Chad reported. Development server is working OK. We need installation of production server by 1/18. Ron will talk with Ann and Sho will contact Nick.
 - We will not accept html as a faculty submittal format. More thought needed as to how and why one would want to do this.
 - We will convert multiple Word documents into multiple pdfs, taking them in the order as they are submitted and presentation IDs will be PDF1, PDF2, etc. If only one PDF, then display will drop the numeric, i.e. it will just show PDF.
 - A question whether an Excel file was being converted properly. (Subsequent tests by Chad show that conversion is OK).
 - We need to bring back into the sw_arch the issue of the archival format which was raised in reference to Excel files. The same problem exists with Word documents.
 - Although we concluded in the meeting that we would convert submitted pdfs to pdf/a, commentary on the list suggests we just accept the pdf as submitted. Given that we need to revisit this issue with ETDs as well, this seems like a sensible approach.
- Google indexing
 - The software for submitted the site map appears to be ready.
 - We will run the cron job nightly. Jeffery will include this in the release notes.
- Statistics reporting
 - Jie has fixed the load and scalability issue with the stats “full report” mode.
 - We will include a “my stats” link on the faculty submittal page, under netid authorization so faculty will see only their stats.
 - For the next release, we need to plan for stats on the custom portals, display the handle, and do a bit of re-architecting of the stats page to break out the alerting function. Fedora 3.x will provide access to Java messaging services (JMS) for alerting.
- Handles
 - Handles on lefty and mss2 will now be prefixed with their server name to eliminate confusion between handles on development servers and mss3.

- We'll run handle audits on mss2 and mss3 to make sure they are all working.
- Signatures
 - We have a legacy problem with signatures that were part of the Realiti collection. The problem being that the techMD points to a DS1 datastream (a pdf for display) rather than the ARCH1 (tar file). This is a messy problem to edit. Discussion subsequent to the meeting suggests that the best way to correct the problem is to pull the 70 some objects from mss3, filter (i.e. correct techMD so it points to ARCH1) and re-ingest. We want to do this in the test interval so we can start clean with signatures with R4.5.
 - There was much discussion about how to make sure the signature process is scalable. The agreement was as follows: a) we would estimate, conservatively, how many signatures can be run nightly, b) divide this number into the total number of objects to determine number of nights to make a full pass, c) implement a cron to do the checking.
 - Notification of failed signatures will go to the listserv "bad_sigs" with initially Kalaivani, Dave, Jeffery, Isaiah, and Ron to be on the listserv. Others are welcome. (Post-meeting, Isaiah has set up the listserv.)
- Video server.
 - We have some bandwidth issues with serving QuickTime with the Helix server. We're OK for release 4.5 but will need to address this issue in the next meeting. Isaiah has proposed that we consider moving to the open-source Darwin server.
 - We also have an architecture issue as to how to connect to an external video server. A post-meeting with Isaiah indicates that we will probably need to implement NFS. As a result, for R4.5, we will provide downloads of Flash or progressive streaming of QuickTime.
 - We should consider this architectural upgrade as part of R5.0
- Other R4.5 items
 - Authorization for dlr/EDIT has been set up with three levels: repository administration, metadata manager, and content provider.
 - User interface – Rhonda will review the default rights statement.
 - User interface – we will add links to display MARC and the full record to the full record display.
 - Pipeline virus scan will be done in the next feature release.
 - Shaun reported that journal export will not be available in R4.5. Shaun and Rhonda will work on the metadata needed and the plugin required for OJS.
 - Mysterious jpeg thumbnails disappearing – Kalaivani will follow up.
 - Ron will forward RLG consulting report on Northwestern, which has several very positive comments about RUcore work.
 - Jeffery proposed that we consider a handle audit as utility in dlr/EDIT
- Topics for next meeting (January 24)
 - Release planning, specifically with respect to future Fedora releases
 - Video server discussion
 - OJS re-scheduling
 - Digital documents (Mary Beth)