

Minutes of October 15, 2009 Meeting

Agenda

1. Updates and Announcements
2. Status of R5.1
3. Performance of R5.0 and R5.1
4. Sound disseminator specification
5. Statistics specification for R5.2

Announcements and Quick Updates

It was noted that Grace has approved the use of “video tape” for the VMC videos as opposed to using the suggested new source type of “video recording”. Therefore, the VMC project and the students who are doing metadata can proceed with R5.0. Isaiah has transferred FMG videos to NJEDge and they have been working out some of the Darwin video streaming issues. It appears, from the most recent emails, that Darwin streaming is working. Rhonda will post the specification on RUcore for the PubMed and faculty survey features.

Status of R5.1

Kalaivani reported on the status of testing for R5.1. It appears at this juncture that statistics reporting and signature checking are working. A bug related to marc export and “type” will need to be fixed in WMS. There are some remaining issues with pdfs and the text layer for OCR. Kalaivani will continue to test in this area. Batch ingest at this point looks good, however Vincent will need to do more testing with NJEDL objects. Isaiah will test in the ingest of thumbnails for video objects. Jeffery will test indexing when objects are associated with multiple collections. Kalaivani reported that there are 13 outstanding bugs of which two are critical. At this point, we believe we are on track for a public release in the first week of November.

Performance of R5.0 and R5.1

Ron reported on performance testing that was done earlier by Jeffery, Isaiah, and Ron. For test results on lefty64, the following was found:

- Elapsed times to present results on an initial search still appear to be unacceptable. However, the elapsed times on R5.1 appears to be significantly less than R5.0. We saw numbers in the 40 sec range on R5.0 and the 6 sec range on R5.0 for equivalent searches.
- For the 2nd through the nth searches, elapsed times on R5.1 dropped to less than a second and dropped to the 4 to 6 second range on R5.0
- It was found that we did not get a significant drop in the 2nd through the nth searches on mss2, suggesting that there is a caching effect on lefty64 and more effective use of memory on mss2. Lefty64 has 4 GB of memory and mss2 has 16 GB.
- We still are getting very basic operations on mss2 that are 3 to 4 times slower than on lefty64 (e.g. display a marc record from showfed)

The good news is that it appears that R5.1 performance is significantly better than R5.0. However, it is still possible to get very high numbers on R5.1 (e.g. for an unquoted set of terms which have to be turned into Boolean AND

operations). Dave reported that his tests showed significantly higher numbers than reported above, suggesting that there might be network issues. We decided to do the following to continue performance analysis: 1) trace through inodes to determine if there is amberfish caching occurring – Jeffery and Sho, 2) Dave will install amberfish version 1.4.2 on mss2 so we can determine if there is a difference from the current version (1.6.x), 3) we will conduct the same performance tests off the SCC network - Jeffery, and 4) we will conduct some basic Fedora operations from the command line on both lefty64 and mss2 to try to understand which very basic operations (view metadata) are 3 to 4 times slower on mss2/mss3 than on lefty64. Dave will setup an account on mss2 so we can execute comparable tests on both servers.

Sound disseminator for R5.2

We discussed a preliminary specification from Isaiah that covered three issues related to disseminating sound files: 1) large file ingest, 2) structure map and segmentation of the sound file, and 3) sound files not playing properly in Firefox. We had an extensive discussion on whether or not to treat sound files like videos, i.e. place the archival master in the /rarch directory rather than try to ingest the large file. Given that a significant number of sound files are small enough to be processed to mp3 and ingested, it was not clear that using the video approach was the best one. There are significant tradeoffs and technical issues. It seemed that a consensus was developing around the notion that the sound content model should be flexible enough to treat large files differently from small files. This raised the question of how to determine what is “large” and what is “small”. This type of change would have a significant effect on WMS. As a result, we decided to consider this issue as part of the larger issue of revising our approach to content models (e.g. for ETDs and for datasets). Ron will launch this working group shortly after R5.1 is released. However, it was concluded that we can proceed with solving the second and third issues in order to be able to handle the jazz oral history collection. Jeffery and Isaiah will lead this effort. As part of this discussion, it was pointed out that we need further support for ingesting large files. Isaiah agreed to develop a case study to identify the major issues and Ron will email Fedora people to see if there has been any progress on this issue.

Statistics Reporting for R5.2

Ron mentioned that Jie will start revising the statistics specification in order to incorporate the new portal API architecture. As a result, there will be two documents produced: a statistics API architecture and a revision of the original specification in order to line up with the portal architecture. We will review both of these documents in sw_arch and then review in CISC, per Grace’s request.

Other Items

Kalaivani suggested that we need to investigate special character processing across our entire architecture. We encounter these problems in various domains (WMS, dlr/EDIT) in which representations get converted or re-converted and do not display properly. Ron will include this work as an investigation with fixes scheduled for R5.2 or possibly quick bug fixes for R5.1.

Agenda Items for the Next Meeting

- Status of R5.1
- Statistics update for R5.2
- Software investigations – java bridge and Perl to PHP
- Sound disseminator - update
- Jpeg2000/aware
- Performance issues – follow-up on actions
- R5.2 review

rcj – 10/026/2009