

SW Arch Meeting July 21, 2011

Announcements and Updates

Ron mentioned that he has proposed a unit goal of moving production sites from development servers to servers that are supported in a production environment. Initially, we should concentrate on Eagleton, NJEDL, and ETDs. Eagleton is ready for testing on Linux and NJEDL is pretty much ready to go as a collection in RUcore. Moving Eagleton and NJEDL will also remove our two biggest CF sites, bringing us closer to moving away from the CF platform. There will be a SOLR/Lucene presentation on Wednesday, July 27 in lieu of the regular CISC meeting. Chad and Jeffery will present SOLR/Lucene capabilities and the architectural/performance issues that are evident at this juncture. Rhonda, Isaiah, and Ron are in a Yearbook meeting on July 26 to begin to work out the issues of ingesting and presenting yearbooks from RUcore. It was also noted that there is considerable discussion on the Fedora listserv regarding the management of large files (e.g. a 200 TB video collection at the Rock

Release 5.2.1 and 6.x Discussion

The basic change resulting from the CISC review is to merge releases 6.1 and 6.2 in order to deliver SOLR/Lucene later this year. Regarding R5.2.1, in addition to bug fixes there are three updates or features that are needed: 1) updated WMS-ETD interaction, 2) portal assignment to an individual object, and 3) the locating of all archival masters less than 2GB to the /rarch directory. For masters greater than 2GB, Isaiah will continue to handle these manually until we solve this problem in release 6.2. For the /rarch directory, we decided that the xml file (created for files over 2GB) will not be needed for the smaller archival masters. There was also some concern about the scalability of a directory (/rarch) that will ultimately have thousands of files. Kalavani noted that a September release will be tight given vacation schedules. We'll continue to monitor the situation and revise schedules as appropriate. The R6.0 release will be on the new platform (servers, storage, Fedora 3.4.2). Several sanity targets were identified and are shown in the release summary with a target release of R6.0 in November. The major feature of R6.1 will be SOLR/Lucene. We will need requirements from the user interface group in September. Architecture requirements will be developed by Chad and Jeffery. A target for a prototype on lefty64 is set for August 19. The public release date is targeted for December. Regarding the transition to Fedora 3.4.2, we expect some impact on application code related to features such as FeSL authentication, new methods in the REST API, and unexpected side effects. Sho will develop document that identifies the problem areas and possible solutions or work-arounds. Finally, it should be noted that the current release summary calls for releases in September, November, and December. Although there is much parallelism in the work to be done (i.e. different people working on each release), we will need to examine these schedules more closely in the next meeting.

Large Files for the Genome Project

The Genome project consists of several large gzip files ranging from 4 to 6 GB. There were many concerns about the time to download these files, the number of simultaneous downloads that might be initiated, and issues with Google indexing this material. As an initial approach, we decided to do the following: 1) the RUCore object will contain a presentation PDF which directs the user to another location specified by a url, 2) at this url (a web page), a user will be able to select and download the gzip files to their local computer. Initially, these files will be located on lefty64 (Isaiah will specify a location). Ryan will develop the text for the pdf and the webpage. Chad will put the webpage up on lefty64. We all acknowledged that this approach is less than perfect but it is an opportunity to learn how best to handle large files and will help us develop the specification for R6.2.

Other Items

- The PDF server has been unstable and also needs to be upgraded to the most recent version. The instability has not been noticed on the development server. Ron will talk with Tibor about monitoring the production server and the upgrade.
- We will need to migrate legacy master files to /rarch
- There is still an issue with the ghostscript version.

Agenda Items for Next Meeting

- Review of release summary
- Review of the url migration specification
- Getting started on a large file specification for R6.2
- Google Analytics