SW Arch Meeting Minutes – March 24, 2016

Agenda

- Announcements and Updates
- 8.1 Update
- 8.0 Update
- 8.1 ABBY file access
- Jobs and Reports
    - RARCH to managed datastreams
    - Adding PDFs to objects with djvu ARCHs or only presentation datastreams.

Announcements and Updates

The DC Area Fedora group is meeting again in late April and is an opportunity to learn what others are doing, specifically in the area of Fedora 4.0. Ron mentioned that we are applying for an NSF grant (Grace is submitting the proposal this week). The library part will involve the creation of a data service that links researchers, datasets, and tools through the use of VIVO and OAI-ORE. Rhonda mentioned that Andy Urban's class is be doing a digital exhibit using materials from the NJDH collection Seabrook Farms.

RUcore 8.1 Update

The code freeze date is April 25 and everyone indicated that they are in pretty good shape to meet the date (we'll have to check with Jie who was on vacation). Kalaivani mentioned that she has not yet heard from Grace regarding whether we can proceed with requested features and bugs in software libraries for 8.1. (In a post-meeting email, Grace noted that we could proceed with CAS-ification activities and any bug fixes, however we should not do any SOAR related updates until she gets further clarification from Kris.)

RUcore 8.0 Update

Dave noted that 8.0 (with Fedora 3.8.1) is running on staging and development servers but has not yet been moved to the testing server. Everyone agreed that he could move 3.8.1 to testing anytime, as soon as he gets the opportunity. We need to discuss when to move 3.8.1 to production – it might be best to make this move when we install the 8.1 release but more discussion is required.

ABBY for OCR/PDFs

Yang reported that ABBY appears to be working quite well for generating PDFs from tiffs and for doing OCR. In fact, OCR elapsed times appear to be about 1/3 of what we were experiencing with AdLib software. However, ABBY still can't handle MS Office documents and URLs. Isaiah will followup to see when we might get the MS Office capability. In the meantime we will continue to use AdLib for both Office documents and URLs. In a post-meeting email, Isaiah reported that software for

creating PDFs from web pages and URLs is available at http://wkhtmltopdf.org/. Yang tried it out and indicated that it appears to work quite well. This should be a discussion in the next sw_arch meeting. It is important to have the Office and URL capability so that we can retire the AdLib software.

Yang reported that the remaining issue with ABBY relates to security and how files are transferred to and from the ABBY server. With restricted IPs we probably don't need passwords. In the meantime, Yang will use FTP and Nick (and the group) will continue to explore the alternatives for providing the best security arrangement

RARCHs to Managed

The work to ingest the RARCH masters has begun. Since the files are quite large, the approach is to ingest files of approximately 150 GB in separate sessions. Dave has run several of sessions that take from 1.5 to 2 hours, and he will continue to ingest the files over the next several weeks, doing one per day for the remaining 12 scripts.

Djvu to PDF

Dave and Jeffery reported that PDFs have been added to all objects with DjVu ARCH's and/or presentation only datastreams. This task is complete.

Other Items

Dave reported that he has not been able to upgrade imagemagick to the latest release and get it to run with jpeg compression. A re-compile of the software will not support this compression on the most recent releases of SUSE. He will continue to work the problem however we may have to consider a different compression option to get it to run.

Ron will send around the most recent release summary.

Agenda Items for Next Meeting

1. Open source software for creating PDFs from URLs
2. Moving Fedora 3.8.1 to production
3. Update on ABBY for Office documents
4. 8.1 update, specifically features and bugs.
5. Additional style sheets for EADs
6. Question re: "source code (text)" and type of item

rcj- 04/04/2016