SW Arch Meeting Minutes – March 14, 2013

Agenda

- Announcements and Updates

- Release summary

- Outstanding R7.0 issues

- R7.1 (Fedora 3.6.2 upgrade)

- DOIs at the file level

Announcements and Updates

We enjoyed refreshments and toasted to a successful R7.0 release. From the review with Grace of the release summary and the recent CISC meeting, we will be adding user accounts as another dot release (see next section).

Release Summary and Dot Releases

The release summary as reviewed in CISC on March 12 and in SW_ARCH on March 13 is now final and will be posted soon on the RUcore website. By way of summary we have created an R7.3 dot release that will focus on the implementation of user accounts (code freeze June 14 and to the public on August 15). The other dot releases remain basically the same. We have also used a guideline for all of these releases to position the public release at two months after the code freeze. Given the smaller release content, these intervals should be achievable. There was some concern about pushing R7.4 (faceted browsing, jpgeg2000, wowza) out to the end of the year. However, throughout the remainder of the year, it looks like there will be opportunities for parallel development (developers working on several releases simultaneously). For example, we may be able to complete the jpeg2000 and wowza capabilities sooner and perhaps release these features prior to the user accounts release. We also want to move some of the EAD display capabilities earlier than the R8.0 if possible.

Outstanding R7.0 Issues

We decided that a quick fix (R7.0.1) is needed to address some critical problems in the R7.0 release. The accepted quick fixes are summarized below:

- ETD import – handling of a temporary directory if a process is interrupted (Yang)

- Multi-byte conversion – do conversion page by page to address the PHP "out of memory"" issue for large OCR xml files (Yang)

- Ordering of file types in showfed (Jeffery)

- File packaging fix (Chad)

Ron also discussed another possible quick fix to statistics regarding the percentage calculation for unique objects downloaded. The objective is to report on the breadth of usage of a collection by dividing the unique objects downloaded by the total number of objects in the collection. Although it was agreed that

the "breadth" statistic is a good one, there remain some semantic and labeling issues. Jie and Ron will work these out and present at the next meeting.

With the above four quick fixes, we decided to move ahead with R7.0.1. The developers have completed the code and done unit testing. We will begin development testing on Monday, March 18 and should be able to push this release out in a week or so.

Release R7.1 (Fedora 3.6.2 upgrade)

Dave summarized the steps to handle the upgrade to Fedora and associated service pack, PHP, and firmware upgrades. The basic steps area as follows:

1. After one week of soak time for R7.0.1, Dave will install the service pack 2.0 upgrade on all three servers (this should occur some time during the week of April 1).
2. He will then install Fedora 3.6.2 on staging and we will do basic sanity testing with the R7.0.1 release.
3. Assuming no major difficulties with 3.6.2 on staging, Dave will then install Fedora 3.6.2 on the development server. Developers may need some time to deal with new features in Fedora at this point.
4. Once we have reached a stable point with Fedora 3.6.2, Dave will upgrade PHP to version 5.4. Again, we may need some time to address PHP incompatibilities.
5. As a final step, we will request Ashwin to do the Iselon firmware upgrade.

DOIs at the File Level and Versioning

Ron reviewed the draft specification for assigning DOIs at the presentation file level. Two scenarios were reviewed: a) the article content model and b) the data project. For the article content model, it was proposed that we not assign a DOI to the PDF since we already have a DOI that takes one to the landing page for the object. The reasons for this were three-fold: 1) it is contrary to what journal publishers do, 2) it is likely to be confusing, e.g. we could end up with three DOIs for a single article – the preprint, the publisher version and the DOI for the PDF and 3) it ignores the DOI federation recommendation that we have only one DOI for a resource. Still, there were opinions that we might want to assign a DOI to the PDF to provide for those cases where we don't want an intervening landing page and for architectural consistency.   For the data project, there appeared to be a consensus that we can assign DOIs to the individual presentation file levels. The benefit is that researchers can now have a much more fine-grained citation capability, a capability that is not typically provided in research data portals. Secondly, we can offer a landing page for the file level that presents more detailed technical metadata.

We also reviewed one versioning scenario in which a researcher has given us a replacement excel file and would like the previous version not to be accessible. In this case, we would add a new datastream for the replacement file with a new DOI and mark the previous DOI as unavailable. We would also need to add an event statement in the metadata and possibly block the earlier version with an xacml policy.

Ron will update the draft with the above comments. The draft will be discussed in the CISC and in an upcoming meeting of the Research Data group.

Other Items

We will proceed to register the RUcore OAI harvesting URL with the central registration agency, as approved by CISC. Also, Chad and Ron will confirm with Grace that we will stick with opaque suffixes in our DOI implementation.

Agenda Items for the Next Meeting (March 28)

- Virus scanning (Isaiah)

- User interface and revision of the landing page (Chad)

- Statistics update (Jie and Ron)

- Content models – next steps - WMS configuration, updating FD, stats, techMD (Rhonda) (4/10 mtg)

- FD features to be added to R7.2 (Rhonda) (4/10 mtg)

rcj – 03/25/2013