

SW Architecture - Minutes  
7/30/15

Present: Kalaivani Ananthan, Isaiah Beard, Jie Geng, Dave Hoover, Rhonda Marker, Chad Mills, Aletia Morgan (recorder), Jeffrey Triggs, Yang Yu

1) Announcements/Updates

- a. OIT Inventory of non-“CASified” web sites. Chad updated the group regarding the OIT LDAP/CAS request that all web apps providing access to university data or using LDAP authentication must be registered to OIT by 7/31/15 and sites migrated to CAS/Shibboleth by 1/31/16. It’s not clear yet what happens if this cannot be done, but we should keep migrating apps to RUcore SSO. Noted that the RUcore SSO doesn’t seem to explicitly adhere to OIT specs, but that will have to be clarified. In some apps we don’t use LDAP for authentication, but rather to categorize users and retrieve user information, so we don’t know if this falls within the OIT request.

Kalaivani reminded us that most of the data is by definition “public” vs. university – restricted, so she is working with Tibor Purger to decide what apps to register, and verify whether LDAP query is still going to be permitted. It’s unclear what OIT is defining as “local databases” that might still be allowed to be configured with independent authentication. She is also investigating whether we use any “internal data” for our RUcore implementation. Note that this OIT request only relates to web-accessible applications, not drive mappings or other internal applications.

- b. TechMD changes. Jeffrey reported that he and Dave Hoover have not yet started to work on these changes, that have been discussed previously. Dave said he did work on an object on production, and all appeared to work properly. He and Jeffrey will review whether changes in restricted permissions will affect our processing, and will develop any needed changes, starting with the development server.
- c. Ref1 data streamer. Jeffrey also noted that he needs to test it with Dave. Current changes worked on DEV so far, but he’s seeing date errors on production. He and Dave plan to work in it 7/30.
- d. Kalaivani asked about how best to monitor jobs/reports, to keep the momentum going. Rhonda suggested that Jeffrey might monitor, and offer a “Jobs & Reports” update as a regular SW\_Arch agenda item. We’ll see how this works.

2) RUcore Release 7.6.1 Update

- a. Chad has trained Rhonda, Marty and Kalaivani on creating access restrictions.
- b. Bug Fixes - Yang found some new issues – generating the PDF for the Presentation Master is creating Monochrome output, and he is working on a fix for that. This, plus he needs to test

his fix to the bug that generated duplicate collections. He has a total of 7 fixes that need to be finalized and tested. Should be ready for testing by 7/31.

- c. There have been some problems running statistics, part of notifications. Following some discussion about implementation of DLR/EDIT and IGNORE commands to resolve, Kalaivani will test by end of day 7/30.
- d. Planned 7.6.1 Rollout Schedule:
  - August 13 – Final Application Packaging
  - August 17 – Install on Staging
  - August 18 – Testing
  - August 24 – Goes into Production
- e. Dave reported that OIT security scans required him to update SSL, Apache, PHP. He's done the upgrades on test/staging with no problems. Will update on Production without any formal downtime notice. Also, DNS changes are required for shift from the old handle server, awaiting OIT DNS changes. At that point, he will put IP restrictions on the apache OCR server to limit access to repository machines. Otherwise, all server software is up to date.
- f. Kalaivani asked how we plan to test access restrictions. The group reviewed the path of requests via the Rutgers Library API. There are some concerns about how OIT/TD is structuring the network, and whether OIT will limit the ability to configure public IP addresses vs. private address space. She has also asked Ed about getting public address ranges for IJS – but Ed said we couldn't have a public IP address range. She will discuss with Ed further. Dave noted that restricting access by IP isn't generally the best approach, since addresses change.

The question arose – is there any reason to restrict access to Thumbnails? General agreement was that they should remain publically accessible.

### 3) ABBYY OCR Server

Isaiah Beard reported that we finally got the license details on Tuesday, 7/28. We are licensed for 1 Million pages with OCR creation, but without the Office document extension (since not released yet). We now need to define the plan for migrating to ABBYY.

At this time, the migration is included in the 7.8 development plan, set for a Spring 2016 rollout. Isaiah suggested that we recommend to CISC that it be moved up to 7.7, and swap with something else that could be pushed back to 7.8. EAD was suggested as one feature to delay with the agreement of CISC. The MS Office extension might then be available, and if so could be incorporated into the rollout. As a reminder, right now we are still using the Adlib tool for OCR of MS Office documents. Given that it does not conform to OIT CAS requirements, it was suggested that we should try to do all outstanding OCR jobs promptly to avoid any issues. This will be added to the 8/12 CISC agenda.

Kalaivani reminded the group to ensure that Nick is involved with this process. Dave and Kalaivani agreed to use the Windows Test Server for the ABBYY testing, although Isaiah noted that we are not

licensed by server, but rather by the page count, so we can install the program on multiple servers as long as they are pointed to the same page count.

#### 4) RUCore 7.7 Specifications & Scheduling

EADs and WMS – Chad is still not clear about how the requirements for this feature. He will talk with Ron, but might need to involve Caryn Radick and Grace Agnew. Chad also described the requirements for the OOI project, and noted that Yu Hung Lin is working on a metadata model that he will review with Grace in August. There needs to be a conversation at the 8/12 CISC as to whether the OOI project that is currently considered a high priority and will require a great deal of effort in the near term, needs to be prioritized ahead of the EAD initiatives.

NJDH – Linda Langschied is working on a statistics report, to be labeled “file counting report”.

Fedora 3.8 – is currently shown on 7.6.1, and has obviously been deferred. Yang would like CISC involvement in the assessment of upgrading to 3.8 or to 4 or not at all. After discussion, there was no consensus about the relative advantages or benefits of these options. Within the group, there is a need for more understanding of these issues, so it was agreed to create an informal study group to investigate further. Dave also suggested that we should look at participating with the Fedora community. A key concern of the group is whether we are wasting development time implementing features that may be available to us in 4.0 or beyond?

Chad is setting up an initial lunch meeting to begin the process **[now set for Tuesday, 8/18/15 at noon in the Heyer Room]**.

Next Steps:

Specs for Page Count in 7.7 – Rhonda will do by 9/13/15

Spec for ABBYY rollout – Isaiah & Chad will do by 9/13/15

By middle/end of August, we should be able to set firm dates for 7.7.

- 5) Additional conversation about SMAP1 Datastreams: The question came up about whether they should be deleted. Since they’re not being used for anything significant now, Yang thought they were just “clutter”. At the same time, Rhonda offered that they could be an important component of moving us towards being a “trusted” repository, in terms of documenting the movement of data and validate the archival structure. However, at this time, the SMAP1 does not capture any changes made to the object after the initial ingest, so it would need modification to become a source of archival validation. Again, we need to understand whether this kind of functionality might be already (or soon) available in Fedora 4, since it would be a lot of work to make this change before seeing any benefit. One suggestion was to change the datastream name to SMAP\_Arch as a possible first step, but there was no final decision yet.

-A Morgan