

SW Arch Meeting Minutes – March 12, 2015

Agenda

- Announcements and Updates
- DOI Update
- techMD to “managed”
- HDL Audit
- Pending Updates
 - REF-1 datastreams
 - Unpacking tar archival masters
- Review of R7.6 Targets

Announcements and Updates

Ron mentioned that the Library Cabinet has approved the Seminar room to be the future Digital Humanities Lab. Implementation has already begun and signs on the Seminar room have been replaced. As a first step, five iMAC workstations have been ordered. The Cabinet has also approved the Data Group’s recommendation that we proceed with the research data service. Regarding our attendance at the Fedora User Group meeting, Ron will distribute notes and organize a meeting for those who are attending and for any other sw_arch members who want to attend.

DOI Update

Some 7000 DOIs have been moved from the reserved state to the public state. The most recent script updated all of the VMC DOIs that just had copyright dates. We still have some 250 objects that have either incorrect dates or no dates that remain in the reserved state. In order to update these objects, Jeffery will product four lists as follows: 1) resource objects with no dates, 2) resource objects with incorrect dates, 3) collection objects with no dates, and 4) collection objects with bad dates (probably very few of these, if any). Kalaivani will take these lists to MDWG to determine how best to proceed.

Although not a DOI problem, Ron also mentioned that we need to update data objects that have no creator. Although these DOIs work, Thomsen-Reuters will not index any of our data without a creator. In a brief analysis, Ron reported that the DataCite metadata store indicates that they have some 27,000 of our objects of which some 14,000 do not have the creator field filled in. Many of these records are legitimate (e.g. a Roman coin record does not have a creator). We decided to move ahead by just fixing the data records by manually refreshing the metadata which will populate the creator field by virtue of the recent xslt mapping update to handle “researcher” and “principal investigator”. Ron will work with Jeffery to proceed with this update.

TechMD to “managed” - Update

Jeffery will continue to investigate the use of the Wilpur script to accomplish this conversion and report back in our next meeting.

HDL Audit and Redirecting HDLs

Chad reported the results of the HDL audit on staging and production. Of most interest to our group were the results on production which indicated 31,496 resource objects with the following results: 1) 13 with a MODs loading problem, 2) 4609 with no HDL, only a DOI, and 3) 26,874 with good HDLs. Of the 26,874 with good HDLs, the following was reported: 1) 26,746 resolved properly, 2) 95 - HDLs not found or HDL redirecting to wrong resource and 3) 33 had resource/URL issues of which 31 are redirecting to another resource (i.e. the HDL was used in more than one record). Kalaivani and Chad will work to repair (or purge) the 13 with MODs problems and the 31 with duplicate HDLs. After this is done, we should be able to proceed with the HDL redirect.

Once the redirect is in place, we should be able to restrict access to the HDL server since the only remaining application is OCR. Isaiah is investigating the use of ABBY software for both OCR and PDF generation and should be able to inform us soon as to the price of the software. If possible, we should try to purchase this software before June, assuming we can find the funds. After we have pricing information, we should take the case to CISC.

Pending Updates

Ron noted that we still have two pending updates: 1) the addition of the REF-1 datastream to article deposits that don't have this file and 2) unpacking the tar archival masters. The group decided to move ahead with the REF-1 addition. Jeffery will create the script for this task.

Review of R7.6 Targets

Our original R7.6 targets need to be amended due to the extended time to get the server configuration established and also since developers are spending more time on responding to user requests and bug fixes. As a result, we decided to create dot release R7.6.1 that would contain the ROA submission and the IP/NetID restriction features. The R7.6.1 release is targeted for August, 2015 with a code freeze date of mid-June. With this change, the R7.6 code freeze will be in late April and the public release is targeted for mid-June. We still should be able to push out the next major release (R7.7) this year with a code freeze date of October and a public release date in the November/December timeframe. Ron will update the release summary with specific dates and we can review it again in the next meeting.

Other Items

Kalaivani raised the issue of conflicting policies for SPSS .por files which are affecting one dataset (Iowa Caucasus). We decided that the policy should reflect an ID of POR (rather than SPSSPOR) to be consistent with the other SPSS file type which has an ID of SAS. Aletia and Ron will work to

replace the SPSS dataset for the Iowa project with one that has the POR ID and Jie will then need to update the stats database in order to preserve the stats that were registered under the old ID.

Dave noted that the new development server is running on SP3 and the other servers are all on SP2. He will address the update by moving the test server to SP3 in the next several weeks.

Dave also addressed the issue of configuration consistency across all servers and requested developers to not make explicit references to /mellon/htdocs in their configuration files. References to /mellon can be accomplished through the use of symlinks. This approach provides flexibility in disk space management on all servers. Yang raised the issue of having different configuration files on production and development servers. After prolonged discussion, it appears that we still do not have consensus on this policy.

Agenda for Next Meeting (March 26, 2015)

- Converting techMD to “managed” – update
- Adding REF-1 datastreams
- Update on ABBY FineReader software (take to CISC?)
- Review of revised R7.6 Targets
- Fedora Users Meeting – Focus on Fedora 4.x
- Pending
 - Updating xacml policies (per Chad’s investigation)
 - Revisit/clarify our quick fix policy (do quick fixes get a dot release?)
 - Review all PDFs in RUcore to determine if they have restrictions.

rcj- 03/23/2015