

SW Arch Meeting Minutes – June 20, 2013

Agenda

- Announcements and Updates
- View List (Add to Folder) Specification
- Scheduling of sha-256 update
- Status of R7.2

Announcements and Updates

Our next sw_arch meeting will be on July 11 and we will then proceed on our bi-weekly schedule (i.e. July 25, August 8, etc). On one other item, Chad noted that there is one lingering problem with validating our OAI URL for harvesting that is related to the time zone. Chad also noted that the stats package has been turned on for all portals except NJEDL.

View List (Add to Folder) Specification

Joseph reviewed the “Add to Folder” functional specification. This function offers the ability to mark items in search results to be saved, printed, emailed, or exported. Only authorized users can save, however the other functions are available to the public. There were several recommendations for relatively small changes as follows: 1) expand a bit on management of folders indicating a user can save folders and start a new folder, 2) clarify the authentication/authorization language for those user account members who are not at Rutgers, 3) consider adding Zotero as an export function, and 4) provide additional explanation in the Purpose/Background section on the recommendation to use “Add to Folder” as opposed to “View List” for the name of this feature. Joseph will update the document and schedule a review with CISC, possibly at the next CISC meeting on July 10.

Scheduling of the sha-256 Update of Legacy Objects

Jeffery reviewed the progress with the sha-256 update process, noting that he has had to increase PHP memory to accommodate large archival masters. He also proposed that we run the update process in batches where each batch executes the update for a single content model. Given the many anomalies that might be encountered, Jeffery recommended that we retain the original archival masters until we are confident that all updates have been performed properly; after verification that the updates have been completed, we can then purge the original archival masters.

Several additional steps were proposed to be added to the update process. First, the script should also create the techMD for each new archival master and the original techMD will be versioned so we have a record of the change. Secondly, we will need to also create the archival structure map for each object. Finally, at some point after all updates have been made and we have inline techMD records, we will need to launch a separate process to create a managed record (“M”) for techMD.

For going forward, Jeffery will make the above changes and run the process for all content models on the development server. We will review progress in the next sw_arch meeting and decide if we are ready to run the update process on production.

Status of Release R7.2

Kalaivani reviewed the status of R7.2 and the most significant outstanding issues. Several changes were requested in how we handle DOIs. We decided that the uniform title can be used if there is not main title; this will satisfy the DataCite profile requirement that we have a title. Secondly, in dlr/EDIT, a simplification in the user interface will allow the user to make the DOI public in one step rather than first reserving the DOI and, secondly, adding the metadata. Kalaivani noted the potential confusion when we want to purge an object that has been assigned a DOI and re-ingest the object. In this case the DOI for the purged object cannot be deleted and the user (or administrator) will need to mark the DOI for the purged object as “unavailable”.

We decided to make several minor portal related changes. We confirmed that the DOI should be labeled as the persistent URL; we will retain the HDL in the metadata but it will not be displayed. It should be noted that the HDLs in all legacy objects will continue to work. Rhonda raised the issue that identifying all events as life cycle events only works well for research data. Identifying publication events for journal articles as “life cycle” events is not appropriate. We decided on making a change in R7.2 in which only events for research data are broken out as life cycle events; all other events will be shown as part of the regular metadata and will not have a selectable item on the left bar.

We concluded that we will need to conduct a special study on how to handle searches in which the text has special characters (e.g an umlaut about a character). For example, for a surname such as Köhn, a user can enter the text and we can display it. However, a search on Kohn will not pick up the text. This is only one example of the various anomalies we are likely to encounter. Jeffery will conduct this study.

Finally, we agreed to reset the target release date for R7.2 from July 1 to July 15.

Agenda Items for the Next Meeting (July 11)

- User account specification - Rhonda
- Update on sha-256 (update by content model) - Jeffery
- OCR investigation (Isaiah)
- Content models – continued review (Rhonda)