

Minutes of April 12, 2007 SW_ARCH Working Group Meeting

1. R4.1 Status and Schedule

Given busy schedules, ingesting for NJDH, and the mini-conference, we've incurred a number of delays with R4.1 testing. As a result, we are re-setting test and release schedules as follows: a) complete Lefty testing - 4/13/07, b) deliver to Systems for MSS2 installation - 4/16/07, c) complete MSS2 testing - 4/19/07, d) Notice to public - 4/19/07, e) deliver to systems for MSS3 installation - 4/20/07 (no later than 5 pm) and f) public release - 4/26/07. Note that there will be the standard two week period for bug reporting, lasting until 5/10/07. We should also consider moving any new bug fixes forward to R4.2 rather than do additional R4.1.x releases.

2. R4.2 Status and Schedule

- We added two new items for R4.2: a) we'll remove the collection TOC disseminator from "showfed", however we'll continue to ingest the collection object with the disseminator. This disseminator was one of our earliest Fedora disseminators and it's not very useful for most collections (it basically just lists all the items in the collection). At some point, we will want to re-write this disseminator or perhaps begin customizing disseminators for specific collections. b) From an earlier MDWG meeting, we decided to remove physical form (mods:form) because of the significant overlap with genre and object architecture. Yang will need to make the WMS form changes and Chad will need to make some user interface/search changes. At this point, we'll just remove the physical form search capability although we will want to replace with genre when we get an list of frequently used genre items. Chad will need a type-map list, to be supplied by Rhonda. Also, as part of the MDWG recommendations, Yang will update the object architecture list (also to be supplied by Rhonda).
- Policy regarding outdated metadata. As part of this discussion, it was noted that certain types of changes in metadata policy result in many objects in the repository not being updated. Our policy is to update all of these objects at some point using one of the following techniques: a) repair asap of those objects that might result in incorrect behavior, b) in cases where just a few objects are affected, manual metadata changes will be made as time permits, and c) for updates on many objects where a pattern or regular expression can be identified, provide for programmed updates. For example, in this area, we might automatically remove all references to mods:form and also update old datastream naming conventions (e.g. DS1) to the current standard.
- Faculty submissions, LDAP, and collection objects. As a review, we noted that the collection object for faculty will be created automatically. The department collection object will have to be created manually. There was a concern about whether we had sufficient control of faculty submissions. There are two control points. First, by using an LDAP filter, only faculty will be allowed to login to the faculty submissions capability. Secondly, all faculty submissions will be parked in WMS for metadata review. We decided that this scenario does give us sufficient control. Initially, we will do virus checking only on faculty submissions. After we see how it works for faculty submissions, we may want to introduce automatic virus checking for all objects although there is a concern about increasing ingest time. An alternative is to do off-line virus checking.
- Schedule discussion. We're moving the previously advertised release data out several weeks to provide time for more library use of R4.1 and to allow for the installation and

configuration of SSL. We decided that code freeze should be on May 18 and we would provide the public release on June 21. The specific dates are: a) code freeze – 5/18/07, b) ST complete on lefty – 6/8/07, c) deliver to Systems for mss2 installation – 6/11/07 d) ST complete on mss2 – 6/15/07, d) notice to public – 6/15/07, e) deliver to Systems for mss3 installation – 6/18/07, and f) public release – 6/21/07.

3. Marc export

Given some of the complexities of marc export, we decided to implement marc export only for ETDs. Some of the more complex issues involve how to treat collection where we don't want item level records (e.g. photograph collections) and how to determine what objects should be exported. We discussed how to indicate that a record has been exported and concluded that we should treat this as an event and update digiProv metadata to reflect that a marc export has occurred for this object. This update will also be tracked via the Fedora audit trail. We also concluded that we will need a collection manager interface to selectively export a marc record for an object (e.g. in the case where a single record has to be re-exported because of a metadata change). We did not conclude the discussion of how to handle an incremental update to an IRIS record. For example, in the case where we have digitized a book that already has a record in IRIS, we may want to export only the 856 link. Initially, this situation will probably require manual intervention. Jeffery will update the requirements document to reflect the above.

4. Supplemental files for ETDs

ETDs will require supplemental files of varying types (lab notes, media files, etc). There are basically two ways to handle supplemental files: a) created additional datastreams in the main ETD object or b) create separate supplemental file object and link them to the main ETD object. Option a is obviously easier to manage while option b provides more flexibility. The general sense of the discussion was that we should proceed with option b. However, option b requires some considerable architecture work and exploration of Fedora capabilities including the use of the RDF-based relationship services. For R4.2, we decided to use option a with the expectation that we will want to update ETDs objects with whatever new approach is developed. Shaun and Ron will work on the content model specification to provide for separate objects as supplemental files. It is noteworthy that option b has significant impact in deciding how best to manage a virtual object that has multiple simple objects.

5. Other items not discussed

We did not have time to discuss all items on the agenda.

- Cron for doing checksum verification. Dave and Jeffery will work offline to configure and install the cron for doing checksum verification.
- Department names for faculty submissions. We do not have an official list of department names for faculty submissions (at least we are not aware of one).
- Collection IDs. At this point, our method for creating a collection ID is ad hoc and is generally based on an abbreviation of the collection title. We need a process that will guarantee that collection IDs are unique.

Given our recent packed agenda, we decided to schedule meetings more often than once a month. Ron will clear calendars for the next meeting.