

Minutes of the March 8, 2007 SW_ARCH Working Group Meeting

Prior to starting the meeting, we had a brief discussion on the new user interface and how we are representing the three types of collections: static, hybrid, and dynamic. We concluded that there might be some confusion on the part of the user when he/she selects only "NJDH" and none of the underlying subcollections. Jeffery and Chad will explore this further and suggest possible improvements.

1. R4.1 Discussion

Our discussion of R4.1 clarified several important aspects of the faculty submissions capability. Regarding collection object setup, Kalaivani will set up the single collection object under which all department collection objects will reside. After this setup, the WMS will automatically create the department collection object and the personal faculty collection object. We also agreed to leave the collection hierarchy level set at three for the UI which will mean that users will see the following hierarchy: RU->Dept/Centers->Specific Departments. Therefore the actual personal faculty collection will not display on the UI for RUcore. Of course, in the department custom portal, they may want to change this hierarchy to show the personal faculty collections.

We resolved several technical issues. First, we will use the CNRI registry of "1782.2" for all RU (and not RUL) materials. Secondly, for the collection ID for faculty collections, we will use the netid. Since netid won't be in place with LDAP until R4.2, WMS will automatically name these for the R4.1 duration. For the collection id for departments, we will have only two for R4.1 (library and oirt). For R4.2 we hope to be able to use a university wide department ID. Dave will investigate the LDAP directory and talk with Chuch H to see if this is possible.

Critical dates for R4.1 include the following: code complete - done, ST complete on lefty - March 16, install on mss2 - March 19, completion of ST on mss2 - March 22, notice of downtime to partners - March 22, delivery of R4.1 to systems - March 23, delivery of R4.1 to the public - March 29.

2. R4.2 Release

The code complete date for R4.2 is April 20 and the target release date for R4.2 to the public is May 23. The other intermediate dates are in the Feb. 8 minutes. We added a few items to R4.2 including the following: a) we will drop shn compression from audio, b) implement department IDs as collection IDs, c) convert PDF text layer to a separate datastream for FT searching - Jeffery will provide the script to Yang, and d) Rhonda will collect additional requirements coming from the faculty deposit mockup discussed at the RUL faculty meeting.

3. File Formats

We had an extended discussion on file formats which concluded with the following:

- For ETDs. The presentation format will be PDF and the archival master will also be PDF. We will extract the text layer (assuming it exists) automatically for FT searching.
- For Faculty submissions. Faculty will submit both the original file format (Word,

PowerPoint, or other text formats). These original submittals will be captured as the archival master. The presentation format will be PDF, and where possible, we will extract the text layer for FT searching. Rhonda pointed out that there will be procedural difficulties with asking faculty to convert a Word document to PDF. Although this can be done with a simple "print" command, the tool would have to be installed on the client. It should be noted that we do not have a good solution to this problem and it also exists for ETDs. We have explored purchasing a server product that would convert to PDFs automatically, however these products are expensive and we are not sure they will handle all the different font conditions and other variables that are likely to crop up.

4. Related discussions

- Maximum packet size for mysql. Dave has set this to 8MB however some of the current coding practice is not sustainable since the size of an sql statement scales with the number of objects in a collection. Dave has requested the programmers need to rethink their approach so that it is more sustainable.
- We noted that it is unclear how to get IJS oral histories started. We believe that the capabilities are in RUcore and we have much of the digital material. Kalaivani will talk with Jane and also pursue setting up a collection object.