

Software Architecture Meeting, Thursday October 18th 2007

(Ananthan, Beard, Ellis, Hoover, Jantz, Liew, Marker, Mills, Triggs, Yu)

Agenda

- 1) Metadata Working Group WMS metadata R4.5 specification
- 2) Collection Hierarchy display and searching R4.5 specification follow-up
- 3) Draft proposal for JP2 migration (not covered at this meeting)
- 4) Handle server, single handle pointing to multiple objects (not covered at this meeting)
- 5) Digital Docs/Administrative Space R5.1 Specification Review (not covered at this meeting)

General Announcements

- We will have our main meeting on the third Thursday of each month, with two-week follow-up meetings held on Wednesdays. Chad Mills will send around a proposed schedule of meetings for the upcoming months.

Agenda Items

- 1) Metadata Working Group WMS metadata R4.5 specification

The Metadata Working Group submitted a specification document with proposed changes to several metadata sections (e.g., source, tech, and desc). Most of these changes are necessitated the anticipated inclusion of increasing numbers and kinds of “born digital” objects in the repository. There will no longer be a “Type” element under the Format element in the sourceMD section, but the sourceMD will add a new value to Format, two new elements under the Frame element, and a new element for the dimensions of three-dimensional objects. The descriptive metadata will need a new relatedItem element for information specific to journal citations. This will affect not only the WMS, but also the end user search and display functions.

There was considerable discussion of a proposal to allow for multiple instances of technical metadata for the new “born digital” objects, particularly digital video, which are expected to require a complex array of archival datastreams that require separate technical descriptions. Such functionality might be of immediate use in any open source release for partners, but will require more investigation of certain issues that it raises before it can be implemented here. We need to consider, for instance, such issues as the affect on our use of tar files, functional datastream IDs, and future migration plans. Thus it was decided that Yang Yu will plan on providing the option of multiple technical metadata sections in the Library of Congress release but will not change the underlying architecture for our own 4.5 release. In the meantime, Isaiah Beard will look into ways we might handle DNG and moving images in the future.

- 2) Collection Hierarchy display and searching R4.5 specification follow-up

Chad Mills displayed a model of the proposed changes to the collection hierarchy that will be displayed in various interfaces to end users of the repository. This sparked an animated discussion of the best way to make “invisible” material available to users. The problem occurs when a collection has its own individual objects but also sub-collections of objects. Users selecting the sub-collections but not the parent collection would miss any objects that belong directly to the parent, though these would prove distracting if the selection of sub-collections were to include the parent itself by default.

Some means must be found to allow the parent collection's objects to be selected as if they represented simply another sub-collection. Several solutions were proposed, including changing the collection ownership of objects in such "mixed" collections so that all collections would either be collections of collection objects or sub-collections containing the individual objects themselves, changing the names of mixed collections and creating new sub-collections to which their objects would now point, and altering the checkbox display for such collections based on the information in the collections database, so that a choice for a pseudo sub-collection would appear for any collection with a greater number of objects than sub-collections. There were concerns about quality control for any project that would entail massive editing of existing XML objects, and also concerns that the restructuring involved in changing parent collections and creating new sub-collections would disturb the collection management structures currently in place in the WMS.

It was decided that Chad Mills should go ahead and make changes in the user display based on the information already in the collections database about the current mixed collections. This will have the effect of exposing any "invisible" objects without necessitating structural changes in the WMS. Chad will create and distribute a new spec outlining this approach. At the same time, we will consider coming up with a set of "best practices" that will discourage the licentious creation of mixed collections in the future while still allowing such flexibility when necessary or desirable. We will continue to experiment with collection restructuring and its possible consequences offline.

Next Meeting

The next Software Architecture Meeting is scheduled for Wednesday October 31st at 9:30 A.M. at the SCC. An initial, proposed agenda is below:

- 1) Jie Geng's investigation of Fedora's rel/ext and diringest and how we might use them in the future
- 2) Jeffery Triggs's investigation of different ways of exposing elements of the repository to Google
- 3) Handle server, single handle pointing to multiple objects
- 4) Digital Docs/Administrative Space R5.1 Specification Review