

Technical Services Planning Committee

Minutes

13 December 2001

Respectfully submitted by Sue Wartzok, FIU, and Susan Heron, USF

Present: Saiyed A. Ahmad, Donna Alsbury, Cecilia Botero, Mae Clark, Janice Donahue, Becky Donlan, John Hein, Susan Heron, Nancy Hershoff, Dot Hope, Allison Howard, Martha Hruska, Lydia Motyka, Dan North, Mary Ann O'Daniel, Lawan Orser, Janice Ross, Sally Somers, Linda Sutton, Sue Wartzok, and Nancy L. Williams.

The meeting was chaired by Susan Heron.

1. System Migration: Update on the cleanup work at UF

Dot Hope and Nancy Williams reported on the cleanup work instituted at UF in preparation for migration to the new system. They met with the functional units (Monograph Acquisitions, Serials, Contributed Cataloging, etc.) to make an inventory of all non-standard practices. They then put together a list of LMS Migration Projects. The document may be viewed at:

<http://www.uflib.ufl.edu/rs/rsd/migpro.html>

The following searches were found to be useful to find problem records:

find no=badoclc (Watch for records in which the OCLC number is preceded by "ocm." The "ocm" prefix is being added in loads from YBP and it must be removed before migration. Watch also for invalid text in the field and for numbers which are not 8 digits long. Most problems with length can be fixed by adding or removing leading zeros. These bad OCLC numbers are not being indexed properly in NOTIS. Thus, cleanup is not just something which needs to be done for the migration.)

find no=source (and the beginning of the code for vendors from whom you have received records, e.g., find no=source bn, find no=source nois)

find fmt=series (in order to find records for which the series information might better be recorded on an authority record)

find fmt=null (in order to find records with null characters in them)

Another problem which is more difficult to cleanup is when a diacritical mark has dropped out of a word leaving it searchable only as two separate words. Example: fran cais, fran cois, bogat a. Many of these records also have extra spaces at the beginning and end of subfields. Most of these records are from batch loads from BNA and YBP.

Incorrect language codes can be discovered by searches which look for disjunction between the place of publication and the language: k260=madrid and lng=eng.

Lydia Motyka has set up a webpage entitled Creation of Test Data. (<http://www.fcla.edu/FCLAinfo/lmsimp/creation.html>) She recommended that all libraries begin flagging records for inclusion in the test file. Libraries should try to anticipate migration problems and include records from all the identified problem areas, such as records with diacritical marks or records from a recon project or a data load. Approximately 100,000 records will be tested from each library. In addition to the flagged records, FCLA will ask libraries to select runs of records—e.g., 10,000 very old records, 10,000 in a particular subject area, etc. Dot was asked to share with the committee the various groups of records which are eventually selected by UF.

There was also discussion of the difficulty of testing acquisitions records because they are so volatile. It may not be possible to test acquisitions records now; instead people should try to have an idea of the kinds of things that need to be tested. Fund codes can't be set up in Aleph until version 15.2 is out. At this time no one knows what's similar or different between the current version and 15.2.

John Hein raised the question of how much historical acquisitions data should be left behind. He said that if we knew how long we would continue to have access to NOTIS, we might feel better about not migrating historical data. Donna Alsbury said that they could keep the DB2 data indefinitely. They are planning to keep Arrow going as an archive for the auditors.

2. Data Cleanup from the FCLA Perspective

Mary Ann O'Daniel had prepared a handout entitled "SUS-Wide Data Cleanup Tasks and Cleanup Tools." (This handout and the ones mentioned in sections 4 and 6 below are available on the FCLA webpage at: <http://www.fcla.edu/FCLAinfo/jcom/jcompg.html>.) In the handout she listed the types of old records which we must decide whether to migrate. She asked the committee to let her know if they think of additional types of records to add to the list. She pointed out that in leaving behind historical records the status code on the holdings record as well as the bibliographic record must be correct. She also pointed out that FCLA will not be able to do elaborate parsing of data at the time of migration because that would take too much time and thus be too expensive. Once we know more about Aleph tools for data cleanup, better decisions can be made on whether to cleanup certain data ahead of time or after migration.

Mary Ann also urged the group to clean out obsolete location codes and to consolidate whenever

possible. She also mentioned that FCLA will clean up obsolete indicators in our databases. She said that using the "fi no=source" is particularly useful so that patterns of funny data can be found. She recommended searching in Catalyst because it gives a better overview and it's not limited to 5,000 hits. Catalyst 2.0 works with Windows2000.

It was pointed out that unlinked items migrated to Aleph will create brief bib records. Libraries may wish to link up the unlinked items that they wish to migrate and leave behind those they haven't been linked. Course reserves are going to present a problem because they are unlinked items that we may want to migrate.

There was also discussion of the problem of duplicate authority records and of whether authority records should be migrated. UF thinks it is going to migrate series authority records only.

3. Ex Libris Authority Files

Martha Hruska and John Hein reported that Aleph can use a master LC authority file. This is the file which is updated weekly. It cannot be edited. However, each library can have a sub-authority file for local exceptions. The system knows that when there is a record in the local file, it takes priority over the master file. For instance, if our series usage deviates, that information will go into the record in our local file.

Nancy Williams said that if there is no 010 field in an authority record, the NOTIS system considers the record to be provisional. The library then receives reports of records which may have been superseded by an LC record. Thus it is a good idea to enter 4XX and 5XX fields in your provisional records so that the system has more options to match against.

Martha suggested doing an inventory of any local subject thesauri being used at your library—e.g., map subject headings at UF. John lamented the loss of the NOTIS merged headings index: it is the best solution ever invented and we're leaving it behind. If a name-title heading needs to be controlled, Aleph will do nothing to control the name portion alone.

The Aleph requirements for collection codes were also briefly discussed. The need for the codes to be unique across the system was questioned. The codes will translate into public displays of up to 80 characters. Mary Ann expects to get some examples from SUNY and she will share the information when she gets it.

FCLA and CCLA may become what is termed *premiere partners* with Ex Libris. Adjacency searching is on our list for a possible future joint development project. However, we will not have access to the source code and thus our relationship will be different from what we had with NOTIS.

4. Update on Status of Data Loads

Mary Ann had prepared a handout updating the committee on the status of data loads. It was entitled "FCLA Dataload Update." She asked that she be informed if there was anything missing from the list. YBP loads are being tested in Aleph; we may have to switch the data format to the one Innovative Interfaces uses. Marcive uses NOTIS-specific codes; we may have to re-do our profile with Marcive.

5. Election of Chair and Vice Chair

Sally Somers was elected Chair; Allison Howard was elected Vice Chair.

Nancy Williams is beginning phased retirement in May 2002. Consequently, she is resigning as co-chair of the authority control committee. Linda Smith from North Florida is taking over as co-chair; Cecilia Botero will remain as the other co-chair.

6. CAGER update

A CAGER update had been requested. Jim Michael, Chair, was unable to be present. He had provided a written report entitled [CAGER Committee Report to the TSPC](#).

The meeting adjourned at 11:40 a.m.