Library Lab Project Final report:
Generating Catalog Records from Disparate and Varied Data Sources (Prototype: HFA FileMaker Db to Aleph)

Project summary

Vendors, publishers and content authors provide the library with metadata for library resources; sources of descriptive metadata are often available to the library in the form of databases, spreadsheets, or other kinds of tabular data. Developing robust capability to ingest this data and automatically generate catalog records for HOLLIS facilitates timely discovery of collections and allocates staff resources effectively by repurposing existing metadata.

The local Harvard Film Archive (HFA) FileMaker database has provided a ready example with which to develop and test the desired data-export and record-creation functionality within Aleph / HOLLIS. The results of the project form a proof-of-concept for other similar projects while gaining high-value results: exposing one of the hidden jewels in the Harvard collections, the Harvard Film Archive's moving image holdings, by making it discoverable through the HOLLIS Catalog.

Accomplishments

We created and loaded over 20,000 bibliographic records with rich keyword access and linked holdings records into Aleph, through a process that involved Excel, MarcEdit, and the Aleph record loader, as well as some changes to the FileMaker database itself. With the FileMaker database continuing to be a critical accessioning, conservation and circulation system for the HFA, we needed to ensure that the Aleph records continue to reflect the HFA collection as it continually grows in FileMaker so we also designed a workflow for quarterly updates.

In developing the workflow for the HFA FileMaker project, we have created local documentation that will be generalizable to other projects involving data from desktop databases, spreadsheets or other forms of tabular data.

A description of the process and all necessary files are available to the Harvard community on the Harvard Library Information & Technical Services wiki: HFA FileMaker to Aleph project.

Challenges – anything we couldn't do

Due to the non-standard data entry and variable presence of data in certain fields, we were limited in the number of controlled access points that we generate for the MARC records. For instance, the personal name data we had did not lend itself to the creation of controlled name headings (1XX or 7XX fields). Not all language or country data translated neatly into fixed field codes, leaving a number of blanks and placeholder values.

Next steps

The first batch of records loaded into Aleph requires some corrections. Erroneously-exported records will be deleted. Adjustments may be made to HFA item numbers to remove leading zeroes, which will require some cleanup work in FileMaker first. Additional cleanup, including the removal of duplicate item numbers, will be required in the FileMaker database in anticipation of the first quarterly record update, scheduled for late January.

We are exploring under which Harvard OCLC symbol we would like to export holdings to OCLC. Once these records are exported to OCLC, the visibility of these records will increase measurably and requests for access to the collection is anticipated to grow substantially.

Presentations that involved the project

Library Lab lightning round in Lamont Library
Library Lab video
Library Lab showcase

The project was also discussed spontaneously at a number of Harvard Library meetings when examples of similar needs surfaced in the community. From the beginning, a goal of the project has been to raise awareness of the work being done so that anyone else who can benefit may do so and that no effort is duplicated unnecessarily.

Project sponsor, Michelle Durocher; Project Manager, Christine Eslao