

Library Lab Proposal: Develop Robust Capability to Generate Catalog Records from Disparate and Varied Data Sources

Proposal Revision Note: The proposal has been revised to highlight the broad applicability of the requested technical functionality for use in ongoing current and future library needs. In line with the Library's foundational strategic initiative of developing an efficient and effective organization, the library needs the tools to capture and utilize existing metadata rather than relying on title-by-title creation of catalog records when data already exists in some form outside our catalog systems. The HFA database is one example of many where existing metadata, irrespective of format and creator, must be leveraged for richer resource discovery. (See Project: rt.berkman #9684)

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 generate catalog records for HOLLIS would facilitate timely discovery of collections and allocate staff resources effectively by repurposing existing metadata.

Test Case Project: The local Harvard Film Archive (HFA) FileMaker database provides 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 would form a proof-of-concept for other similar projects while gaining high-value results: exposing one of the hidden jewels in the Harvard collections by making it discoverable through the HOLLIS Catalog.

Applicability to Future Projects: There are other databases, spreadsheets and other types of descriptions of collections held locally on desktop PCs across campus that the library would benefit from being able to transform into usable catalog metadata. There are also new acquisitions workflows which include obtaining spreadsheets from materials vendors. Donor gifts or antiquarian purchases of large collections are other materials that are sometimes accompanied by spreadsheets. Developing tools to work with existing data, especially tabular data that can be mapped in coherent ways to known catalog fields, has broad applicability.

Options evaluated so far to address HFA needs: Prior to submitting this proposal, and indeed back to late 2009, technical services librarians, Christine Eslao and Linda Takata, met with multiple HFA staff, including the curator Haden Guest and HFA conservator, Liz Coffey (now administratively part of the Weissman Preservation Center) to look at the HFA accessioning workflows and evaluate whether this work could be performed directly in Aleph. Film cataloging specialist, Steffen Pierce, also contributed to the analysis of needs.

After thorough review of the workflows which rely on the FileMaker database, it was unanimously determined that Aleph did not provide functionality that was needed to support the conservation, curation and public services needs of the HFA staff and collection. That is when the record-export option was identified as the best available option with the current FileMaker database.

Background Information on Current Uses of the HFA FileMaker db, as context:

What services is the FileMaker db critical for maintaining?

- Used by all HFA staff at the Carpenter Center and by preservation staff at 625 as the chief source for all information on the core collection including filmographic, physical details, decomposition of the material, location, loans, and rentals. Also has posters and video components.
- Scholarly reference inquiries and institutional loans are searched through FileMaker for researchers worldwide.
- Supports research for the PhD. Program in Film and Visual Studies in the Dept. of Visual and Environmental Studies (VES)
- New fields are easily added and modified by HFA staff to immediately respond to evolving research areas.
- HFA staff track various details that are extremely specific to motion picture film and could not easily be shoehorned into existing databases with stricter data structures.

Using information that HFA staff record in FileMaker, how is it consulted later to help them manage the collection, make materials available for specialized uses, triage or prioritize further conservation or cataloging activities, etc.?

- Quickly retrieve current data to effectively serve curatorial, conservation, & programming needs. Programmers locate 'viewing prints' for screenings and track prints that they borrow (from outside the HFA Collection) or loan.
- HFA and film preservation staff view and edit data at both 625 Mass. Ave. & Carpenter Center
- Using the "Find Mode," able to search on every field as well as search on combined fields. Easily able to identify items in need of preservation or re-housing, or items that would be valuable for exhibition.

FileMaker provides the following for public service functions and internal staff functions:

- Enables searching, data entry & updating by non-Technical Services staff without training in MARC cataloging
- Extensive, separate records for individual prints
- Specialized filmographic data fields pertinent to HFA's materials, workflows, and internal record-keeping. These fields do not correspond to existing MARC fields, and could not easily be shoehorned into existing databases with stricter data structures.
Crucial detailed print condition & usage notes
- Searches by HFA staff that are instantaneous and flexible rather than pre-formulated "reports" that require outside assistance or specialized training in separate reporting software

Technical Specifications

The technical requirements of the project have been analyzed preliminarily, and the data mapping from the FileMaker db to MARC have been identified. For the sustainability of HFA workflows after the data migration to HOLLIS, it would be necessary to develop the initial batch transfer from FileMaker as well as an update function to ensure continued data transfer for new records and edits. In other words, the record export would need to handle a large initial bulk load and then be configured to capture changes at appropriate intervals.

Filemaker to Aleph to HOLLIS/HOLLIS Classic: migration issues or open questions

- Need update function for edits to existing records and entry of new records in Filemaker, such that Aleph data is consistent with Filemaker data after initial record migration.
- FileMaker contains 1 record per print, each with a unique HFA Item number. To implement Archival Moving Image Materials cataloging standard (AMIM) of 1 bib for multiple prints of the same version, would the work of bringing together multiple print records take place before data transfer? Or would single-print records be merged after migration? The former would require preparatory work; either would require cleanup.
- Film Stills Cataloger would be able to enhance FM preliminary records in Aleph rather than starting from scratch as he does now; easier to implement AMIM standard of 1 bib for multiple prints of the same version.
- Need to de-dupe existing records already cataloged manually in Aleph.
 - Prints with full or preliminary records are noted in FM records. (Matching Aleph number is not noted, however.)
 - HFA item numbers are noted in existing Aleph records.
- Important HFA FileMaker fields don't map well to MARC; transfer a limited number of appropriate fields
 - Need an Encl level and 'core' standard for bibliographic records. The following selected FileMaker fields (with MARC/Aleph mapping) were suggested. MARC

mapping needs to maximize indexing and retrieval. Need to decide what record level and 'fullness' needed for HOLLIS/ WorldCat:

ORIGINAL TITLE = 245|a

ALTERNATIVE TITLES = 246|a

DIRECTOR = 245|c

PRODUCER = 508 Note

PRODUCTION COMPANY = 260|b

CAST = 511 Note

DATE OF RELEASE = 008 [7-10] ; 260|c

FORMAT = VM ; LDR [06] (g) ; 008 [33] (m)

ELEMENT = 500 Note

COUNTRY OF RELEASE = 257|a (not indexed) **or** 260|a (does it go with 260|b?)

ITEM NUMBER (generated in FM and unique for each print; needs to be easily searchable) = 562 (HOL) See "Item Number" below

GENRE = 655|a

- Unable to guarantee controlled headings for names in FM (MARS effectiveness would be limited but should be flagged for separate reports to allow authorized headings to be evaluated by Film specialist)
 - For name headings, would we propose that they initially load to Aleph as they appear in FM, but then run through the automated authority services (MARS) to be corrected to authorized forms. From what we see in FM, would we envision MARS would be effective to automate that heading correction, or would most names simply land on reports requiring manual clean-up.
 - Director, Producer, Cast, etc. names are all in direct order. Since names in direct order are rarely cross-referenced, MARS would have no way to identify a name and they would all require manual clean-up. Controlled headings can probably not be machine produced.
- Requires clean up for initial articles in 246 field.
- Terms from the FM Genre table can be exported directly as 653s or mapped to 655 7|2|c|g|f|t or 655 0. Have recently begun adding LCSH (650s) to FM for content. Clean up required.
- Need KW searchable and reportable field for HFA Item number (562 is good, but searchable only in HOLLIS Classic)
- Exportable to WorldCat
- Future benefits to build upon: Access of films through Aeon. Are Item Records required?