

Internship Final Report

Daniel Davis

School of Information, San José State University

INFO 294: Professional Experience Internship

Dr. Linda Main

December 5, 2021

Introduction

During the Fall 2021 semester I participated in a non-paid internship at the American Film Institute (AFI), a non-profit institution based in Los Angeles, CA. Fall 2021 is my last full semester of my MLIS program before I move on to my ePortfolio in the Spring 2022 semester. I have been on a loose special librarianship track with the intention of entering the field after graduation, but in my professional career to date, I have not worked in any library, special or otherwise. One of my primary goals for pursuing an internship was thus to get some real-world library and archives experience.

I chose this particular internship for two main reasons. First, it offered the opportunity to work in a special library and with real-world assets in a database management system (DAM). Second, as a film buff, I was attracted to working with material that I have an interest in beyond its value as DAM content.

Site Description

My internship was specifically with AFI's Louis B. Mayer Library. The AFI preserves film and television history and includes a school for aspiring filmmakers. Students of this school as well as other researchers are the library's main patrons.

AFI and the Louis B. Mayer Library are located in Los Angeles, CA, but the internship was completely remote. The internship site supervisor is the institute's archivist. The main project for the internship involved creating and managing metadata for the digital asset management (DAM) system to improve the discoverability and accessibility of the library's filmmaker seminars. These seminars are entered into the system as audio or video recordings,

and the subjects are individuals or panels of people involved in the film or television industry, be they directors, actors, producers, cinematographers, or other professionals.

Learning Outcomes

Learning Outcome 1: Apply methods for processing and organizing an archival collection of print, photographic, and moving image materials in a digital environment.

During my internship, AFI implemented its new DAM system. After that time my primary responsibility for the internship was reviewing audio and video files within the DAM system and filling in certain fields for each record. In this case, each record is a part or the whole of an audio or video recording of an AFI seminar.

Reviewing and tagging the records required understanding the naming structure of the file as it relates to either the digitized version of a physical asset or a born-digital asset. More specifically, the file name includes information about the type of seminar, the guest, the date, the barcode of the physical original (if applicable), and the source of the recording—whether that’s an audio cassette, DV cam, Hi8 cam, etc.

I obtained further information about the seminar from two places: Mayer Library’s LibraryWorld database website, which has entries for digitization notes, barcodes, and source element; and an Excel spreadsheet hosted on Sharepoint and containing columns for (among other things) the seminar moderator, subjects, seminar type, and date.

Additional information about metadata tags is covered in Learning Outcomes 3 and 4.

Learning Outcome 2: Appraise organizational holdings for content with consideration paid toward organizational audience and goals.

The goal of the DAM system and its holdings is to make it easier for AFI students and other researchers to search for and find relevant content. On my part, this required balancing speed with thoroughness when entering metadata for records. Many seminars do not have subject

topics listed in the spreadsheet. I was not required to add subject topics in such cases, but doing so would often improve the user experience and help accomplish the goal of the DAM system as it relates to the serving the audience and improving the search function. However, seminars are typically an hour and a half to over two hours long, so listening to or viewing each one in its entirety would mean adding records to the system at a much slower pace. Instead, I opted for a balanced approach of listening/watching just long enough to determine the main topic or two and adding those as Subject field entries.

Learning Outcome 3: Define and manage metadata electronically as it relates to physical holdings.

Since this was an entirely remote internship, I worked only with digitized versions of physical holdings (and born-digital files). However, it was important to know the nature of the physical original, including what form it takes (cassette, Hi8, etc.), how many sides/parts there are, and which physical copy of the seminar was digitized, as indicated by the barcode number.

This information dictated what metadata I included in the Title, Part, Asset Type, Source Element, Barcode ID, and Audio Channel fields in the DAM system. In particular, the site manager stressed the importance of correctly indicating the part number in the Title and Part fields. For example, the four files for a seminar that covers both sides of two cassettes will have Part field entries of, respectively, “1 of 4” through “4 of 4.” This information was crucial for helping users be sure they were accessing all the parts available to them and determining where to find the information they needed.

Learning Outcome 4: Create descriptive metadata related to digitized and born-digital materials.

The example record below shows the metadata I created for each record. Red asterisks indicate DAM system-required fields, but the site manager also identified other mandatory fields we interns were to fill out based on information found in the file name, the LibraryWorld entry, and the Sharepoint spreadsheet. These additional mandatory fields were Asset Date, Part, Source Element, Series Type, and Program. Other fields were optional because they do not apply to every file. For example, born-digital files do not have barcodes, and most audio recordings are only one channel, rather than having left and right channels.

As noted above, while the Subject field is not mandatory, and whether or not a certain file had any subject topics listed in the spreadsheet, I sometimes added additional Subject field tags based on the audio/video file content. These might be existing tags that I chose from the DAM system pulldown menu for that field, or I might create new tags as appropriate. I also sometimes added other tags that the original person filling out the LibraryWorld and spreadsheet entries perhaps didn't listen to or watch the file closely or long enough to catch, such as the names of additional guests or the seminar moderator. Whenever I added tags not found in the LibraryWorld and spreadsheet entries, I informed the site manager so that she could update those records to match what is in the DAM system.

Library / Archives / Audio Video Archive / Seminars / Seminars / BallLucille /

HLMS - Lucille Ball (1/18/74) - Part 3 Audio Archives

OverviewDetailsTags

Library / Archives / Audio Video Archive / Seminars / Seminars / BallLucille / HLMS - Lucille Ball (1/18/74) - Part 3 (Audio Archives)

RECORD TYPE

Subtype
Audio Archives


0:00 / 0:00

Unique identifier *AI1B05
Legacy identifier
HLMS_011874_Ball_Lucille_A
P15832_S1_2of2_R

FILE INFORMATION

BASIC INFORMATION

Parent folder
BallLucille (AI1AZB)

Title
HLMS - Lucille Ball (1/18/74) - Part 3

Guest
Lucille Ball

Moderator
James Powers

Asset Date
Single Date
01 18 1974

Part
3 of 3

Asset Type
Audio Recording

Source Element
Cassette

Barcode ID
AP15832

Audio Channel
Right

Series Type
Conservatory Seminars
HLMS

Program
Conservatory Seminar

Description

Lead document

Lead document for:
Location
Add a value

Tags

Subjects
Acting
Career
Comedy

Department
AFI Archives

Archival Collection
Add a value

Staff Editor

Edit Date
mm dd yyyy

Project
Add a value

ADMINISTRATIVE INFORMATION

Purpose
Ready for Review

Editor's rating
★★★ (F)

Notes

Source

Copyright
©American Film Institute

Usage rights

Storage group
Default storage group

Display template

Status: Not started

Edit Statuses

ACTIONS

Download

Share

Orders

Created by Emily Wittenberg on 09/24/2021 12:01:39 AM (UTC)
Edited by Daniel Davis on 10/13/2021 10:07:14 PM (UTC)

Duplicate

Edit

Save

Delete

Cancel

New

Communication

The primary channels I and the other interns used for communicating with the site supervisor were email for conveying tasks and asking questions, plus Zoom for weekly check-ins with the entire team (the site supervisor and six interns). Zoom is an excellent tool for casual conversations and quick catch-ups, and for sharing ideas and experiences. Email is an effective means of communication in that the site supervisor is quick to respond to questions and provide further work when needed. However, a common spreadsheet or document (e.g., on Google Docs) of previous questions from all interns might have been easier than email for searching or browsing through to find answers from the site supervisor.

Conclusion

Management Style

My site supervisor's management style was an ideal mix of involvement without micromanagement. She encouraged me and the other interns to ask her questions and to communicate with each other so that we might learn from one another's experiences with the content and working in the DAM. She was always quick to answer questions, or to at least say when she needed more time to look into a particular issue. She respected our time by keeping our meetings to one 30-minute gathering once a week. There are no changes in her approach that I would recommend.

Technology

The main piece of technology for completing work in the internship was the DAM system. This software was an iteration of Cortex by the company Orange Logic. The web-based interface required logging into the system, at which time we the interns could access a My

Collections page that had the individual files (each being one part of a seminar recording) that the site supervisor had assigned to us. Each intern had their own My Collections page with their own set of assigned files to avoid duplication of effort or confusion of ownership. The page for a specific file had an Overview tab where we could listen to or view the seminar recording, and a Details file where we would fill out the metadata fields.

The other two web-based tools we used were the LibraryWorld database and the Sharepoint-hosted Excel spreadsheet, both of which provided information necessary for coding the files. We also used email for communicating with the site supervisor and Zoom meetings for our weekly check-ins. One intern set up a group on the Slack chat website for interns to communicate with one another, but we used this only on a couple occasions.

As noted above, there was no formal repository for questions posed by interns and answered by the site supervisor. I mentioned a common document as a potential solution; alternative technologies that might be an even better solution are project management software or a ticketing system where the interns would submit questions and comments. These tools might be more expensive and difficult to implement compared to a common document, but they also might better facilitate finding answers to previously asked questions, saving time for the interns when seeking a best practice, and also for the site supervisor by minimizing her having to re-answer questions an intern asks again (whether they or another intern asked it originally) because they couldn't find or didn't want to search their emails for her previous answer.

Responsibilities

Before the site supervisor implemented the library's new DAM system, my main task was reading seminar transcripts and giving them appropriate metadata tags based on the content. My

job was to read through an entire transcript—typically 40 to 60 pages—and list any person, movie, or subject discussed in enough detail as to be of interest to a researcher. I then emailed the metadata—the name of the seminar interviewee, the date of the seminar, the subject tags, and the moderator—to the internship site supervisor.

The other task I performed at the beginning of the internship was for the internship site supervisor's controlled vocabulary project. This involved taking seminar metadata entered by previous interns and dividing blocks of tags into individual cells in an Excel spreadsheet, while also changing terms to title caps (leaving film and TV show titles in all caps) and fixing errors.

As described in previous sections, my responsibilities once the DAM system was up and running was to evaluate the digitized audio and video recordings of seminars with film industry professionals dating back to the 1970s, and then create descriptive metadata for the files in the DAM system. The goal of this project is to make it faster and easier for researchers to find the content they need.

With all of these tasks, the site supervisor's main concern was that we interns tag the content thoroughly and consistently so that it could be easily discovered by researchers.

Challenges

These goals of thoroughness and consistency led to a couple of challenges. First, there were inconsistencies in information as found in the LibraryWorld entries, the spreadsheet entries, and the file names in the DAM system. For example, one seminar might be listed as "HLMS" (Harold Lloyd Master Seminar) in LibraryWorld and titled such in the DAM system, but then the spreadsheet might indicate a different type of seminar (e.g., ASC Cinematographer Seminar).

The site supervisor would then have to advise on whether this particular seminar should be coded with HLMS, ASC, or both.

Another challenge was trying to make sure all the files for a seminar were accounted for. A file name with “S1_2of2” means that it’s the first side of the second asset (usually an audio cassette) for a seminar. If there is no information on side 2 of that second cassette, the file name for side 1 should instead just have “2of2” without the “S1_.” However, while naming the side 1 file, the person digitizing it might not have yet known if there is content on side 2 and thus out of caution included “S1_” in the file name; but then upon discovering there is no content on side 2, they neglected to go back and delete “S1_” from the side 1 file name. It would then be up to the intern to determine if the seminar concludes on side 1 or if there’s a side 2 in existence but missing from the intern’s collection in the DAM or from the DAM entirely. It was almost always the former; however, while rare, there were a couple instances of missing files (either in this scenario or others), in which case we interns would notify the site supervisor and she would have to find the physical asset, digitize it, and get it in the system.

There were also some challenges stemming from technical problems with the files. An audio file might have too much static to be understandable. A video file might be lacking audio, or be all black. Whether or not to code the file was not always obvious; for example, perhaps somebody with a better audio system would be able to hear and decipher a low-volume audio file that I could not. If these files were one part of a multi-part seminar, whether or not to code them would affect how the Part fields of the other files for the seminar should be coded.

Valuable Courses

INFO 200 – Information Communities was a useful course in that it provided a good background on information-seeking behaviors of particular communities and cohorts.

Understanding how individuals within these communities interact with information shaped my work for the internship, especially my thinking around how to tag content for students and researchers who might be interested in particular seminars or topics discussed within those seminars. For example, a researcher would probably find it a waste of time to access and watch or listen to a seminar for a particular subject and find it mentioned only briefly without further context or elaboration.

INFO 202 – Information Retrieval System Design and INFO 284 – Archive Records Management were two other very useful courses in that they gave me a basic understanding of information retrieval system design and organization in general and DAM systems in particular. This helped smooth my introduction to the AFI DAM system by giving me familiarity with aspects such as information structure, metadata hierarchies, and controlled vocabularies.

Additional Courses

While I felt very well equipped to handle the responsibilities of this internship, two courses that I would have liked to have taken beforehand are INFO 240 – Information Technology Tools and Applications and INFO 242 – Database Management. In INFO 240 I would have obtained a background on modular web design and strategies for presenting information on websites. In INFO 242 I would have learned more about the structure and design of database management systems, including organization and querying. Together these would, I believe, have imparted some interesting insights into the strategy behind and decisions that went into designing the AFI DAM system as well as the AFI database setup on LibraryWorld.