Eddy Colloton

10/16/2014

Metadata for Moving Image Collections

Assignment 1

Metadata for Searching Moving Image Collections Comparison

The way in which an archive organizes its metadata is paramount to providing access to the archive's collection. The amount and specificity of metadata about a resource determines the ability of a search engine to properly identify it and link it to a user's query. That being said, every archive is different, and the type of collection, type of user, and context for the archive must all be considered when evaluating the effectivity of an archive's metadata practices. For example, the UCLA Film and Television Archive's collection and user-base is very different from the Internet Archive's collection of moving images, and therefore searching the two collections online yields a very different experience. Each archive's searching tools and granularity of metadata description will be discussed in detail, and then compared.

The UCLA Film and Television Archive's online catalog has a robust, and academic approach to presenting their catalog online. One can both browse or search the archive but only through separate portals from their main website (cinema.ucla.edu). Browsing is organized in archival order, the user first selects Motion Picture or Television collections, and then chooses from a list of several collections within the archive. Many of the collections have their own websites, for example the Mayme A. Clayton Collection (claytonmuseum.org), but the UCLA site provides a list of the films from the collection in a pdf or excel document as well. However, some of the lists provided by UCLA are

incomplete, such as the Dorothy Arzner Collection (https://www.cinema.ucla.edu/collections/dorothy-arzner).

When using UCLA Film and Television Archive's online catalog's search functionality, one can immediately narrow his or her search by Title, Credits, Topic or genre/form, Keyword, Inventory number, Release date, Broadcast date, Collection name, Title variants, Credit variants, Pre-existing works, or Holdings. Each field is explained thoroughly below the banner, with examples of queries. It's relatively easy to use, and follow, but is also bare-bones. The design is not aesthetically pleasing. For a even more refunded search, there is an "advanced" tab which provides three fields for searching, each with 2 drop-down menus. The first designates wether the user's search should include "all of these" terms, "any of these" or "as a phrase." The second drop down menu, preceded by the word "within", lists the aforementioned categories (title search, credit search, etc.). The three search fields are separated by yet another drop down menu with the modifiers "and," "or," and "not," allowing users to include or exclude related search queries.

The search results I received from the UCLA Film and Television's Archive were generally what I was expecting to find, although I occasionally needed to refine my search methods. Using the keyword search with "Lillian Schwartz" was too broad. I got works by other filmmakers with the last name Schwartz, who had nothing to do with Lillian Schwartz. The "edit search" button conveniently led me back to the search page, and by selecting "credits search" button, I got more helpful results. The first result was a short film titled "Enigma" (1971) that Lillian Schwartz created with Ken Knowlton. The other result is "Some of my best friends are—" a 1971 film which had a wardrobe assistant of the same name as the artist. The broad nature of results is understandable, given the breadth of the

archive. Similarly, searching for "Computer animation" in the "genre/form" field gave me many results I was interested in, such as *Cybernetic 5.3* by John Stehura and *Permutations* by John Whitney, but also *A Bug's Life* and *Allosaurus: a Walking with Dinosaurs Special*, which I was not so interested in. The "happy medium" solution given the extent of UCLA's collection was searching for a more well-known and prolific, yet individual filmmaker, like Stan Brakhage, which results in almost 40 of his films, and no works which do not directly pertain to him in one way or another.

The metadata which allows such searches to be possible is organized using MARC21, there are "view" options in the upper right corner of the website that allow for "brief view" or "Staff MARC view." Further complying with information science standards, the "genre" subject field (MARC tag 655) uses the Library of Congress Subject Headings. This generates data which is quite granular, providing all of the information one would need to "map" the system's MARC entry on to a EN15744 entry (except country of origin in the case of the Lillian Schwartz film, puzzlingly), and exceeding this minimum set by providing thorough content description and synopsis. However, not enough information is provided to map to the more granular EN15907 standard, possibly due to insufficient data. For example, while it is possible Lillian Schwartz's "Enigma" has won awards, it certainly did not win an Academy Award or Golden Globe, so this information would be relatively obscure. Regardless, this level of granular metadata strongly supported finding material that matched my queries, and the search engine presented the results in a logical and easy to understand way, alphabetical by title.

No system is perfect, however. For instance, UCLA's online catalog's website did combine subfields of the "Added Entry Personal" data (MARC tag 700) together, conflating

them. The MARC view separates subfield $d, “dates associated with a name,” and subfield $e “Relator term,” to read:

“7001_ |a Schwartz, Lillian F. |q (Lillian Feldman), |d 1927- |e animation.”

However, the “brief view” (presumably what most users will view), provides a link to the subject “1927-animation,” which has nothing to do with a film made in 1971 by a woman who was born in 1927.

This level of scrutiny would be exhausting and unfair to the Internet Archive’s collection of moving images. The Internet Archive’s moving images are uploaded by Archive users, many of which are in the public domain, or have a Creative Commons license. Due to the nature of the collection it was a bit difficult to find films or video of the same type as the ones I had been searching for on the UCLA Film and Television Archive. For instance a keyword search for “Stan Brakhage” results in a list of short videos inspired by the avant grade filmmaker’s work, but very little actually featuring Brakhage, and none of his work. This makes sense, of course, because none of his work is in the public domain or Prelinger Archive.

The specific nature of the type of media the Internet Archive collects makes searching more challenging. Broad searches for terms like “senate” or “Beethoven” often provide results which do indeed pertain to the US Senate or Ludwig van Beethoven (respectively). However searching for “Beethoven fifth symphony” does not provide an audio recording of Beethoven’s 5th, but rather two album covers of recordings of the symphony, a video of a lecture on the symphony, and several other less relevant results which have “Beethoven,” or “symphony” in their description. One may be satisfied with these results if they were simply hoping for information about the Fifth Symphony, but

these results are not particularly intuitive when compared to the results of an archive like the UCLA Film and Television Archive.

There are multiple way of searching the Internet Archive. The “simple” search provides a drop down menu for media type, further delineated by genre below the media type. The advanced search provides multiple fields that allow the user to search by Title, Creator, Description, Collection, Mediatype, Date or Date range, and three custom fields. The custom fields allow for numerous options, some of which apply to specific media types, such as “album-cover”, while others are more general, like “copyright-expirydate.”

The metadata search fields can become quite granular when applying the “custom field” function, depending on how it is used. However, this also depends on the uploader to provide granular metadata about the resource. For instance, when searching in the “AnimalCollective” collection, one gets 25 audio recordings of live concerts (bootlegs), but if the search were more specific, searching in the “AnimalCollective” collection with “audio” in the mediatype field, the search would have no results. Even though there are resources that match the more accurate description, it is more beneficial to search broadly.

The use of optional, user input metadata makes searching the Internet Archive more akin to searching Google than a library’s card catalog. A word which has some relationship to a title will get many more results than the specific title itself. Because of this, it could be argued the metadata does not support finding resources, however, it can lead to broader discoveries.

The Internet Archive has a good deal of documentation about the open source

search engine and syntax the website uses[1], but it is unclear (to me) if they use a metadata standard for video. Different entries provide a varying amount of information about the specified resource, filled out by whomever posted the material, as opposed to a cataloger or other trained professional. Many text resources have much more specific metadata, often including an ISBN number, and some even including a link to a MARCXML record.

The search results for the Internet Archive are then, by default, ordered by "relevance" as determined by their search engine. The results can also be displayed by average rating, download count, date, or date added. A user can also choose to group results by relevance, mediatype or collection.

The Internet Archive is very browsing friendly, sorting videos first by broad genre terms, and then further dividing the results into more granular genres. Some of these granular genres, such as the "Brick Films" (although none appear to be shot on film) can then be sorted by average user rating, download count, date (presumably date of creation), and date added. Unfortunately not all of these sub-genres allow for this search, as is the case for Animation Shorts, which only allows browsing by Subject/Keyword. This is particularly disappointing as the "Subject/Keyword" sorting is far too granular, with many subjects only containing one video.

How an archive's collection is organized and the extent to which the collection's metadata helps users search says a lot about both the archive and its users. The academic, standardized search functionality of the UCLA Film and Television Archive reflects the institution's notoriety and active role in the archive community, as well as the scholarly and professional nature of its user-base. In contrast the Google-esque search functionality

[1] The search engine the Internet Archive uses is called Apache Lucene and it uses the SolrQuery Syntax, a set of rules for how the search engine interprets user inputs.

and user-generated metadata of the Internet Archive reflect its populist and crowd-sourced workflow and mentality, not to mention the broad user community. Neither approach is wrong or right, each merely reflects the distinct role the archive fulfills. There is no, nor should there be, a one-size fits all moving image archive, rather, communities must come together to create the archive which best serves their goals, as exemplified here by the UCLA Film and Television Archive, and the Internet Archive.