Brian Cruz

Metadata for Moving Image Collections

Assignment #2

Metadata Mapping: MARC, PBCore and MODS

For this assignment I created a crosswalk between the MARC 21, PBCore and MODS metadata standards, which can be found in the accompanying spreadsheet. Each standard has its own strengths and weaknesses, often leading to bewilderment as to how to map data between them.

The MARC 21 metadata standard is certainly the more elaborate and convoluted of the three. Given its origins in the 1960s and its harmonization of formats used in different countries, it's rather remarkable that it's as robust as it is. But its esoteric numbering system, along with the use of tags and indicators within its fields, make it much too difficult for a novice to understand, let alone master. Without a guide, one has no hope of figuring out what data goes where. While it is certainly more granular than PBCore or MODS, that very granularity is what makes it ill-suited for use in cataloging moving images. Most fields are intended for print publications, and thus are not relevant for our purposes. There are also many instances in which several fields are used for essentially the same thing -- a perfect example being physical formats -- making it difficult to determine which field is best suited to contain that particular information. If an object is a VHS tape, should that be entered into 007, 300, 337, 338, 340, or 346? An argument can be made for all of them, and that's not even considering the tags and indicators within each one. PBCore and MODS are much more streamlined and understandable at a glance than MARC 21, though they do have problems of their own.

PBCore, now on version 2.0, was created specifically for media cataloging by the public broadcasting community. Built on the foundation of Dublin Core, PBCore's greatest strength is its ability

to provide metadata for multiple instantiations of a single work. A film such as *Casablanca* may have instantiations that include a DVD disc, 35mm film print, and a digital file, each of which shares metadata from the core work in addition to their own unique elements. This gets confusing when one instantiation is composed of several different essences; should one always have to fill in both essenceTrackTimeStart and instantiationTimeStart, even if they're the same value? It's also frustrating that many fields can only be used once; why should every language be entered into one field rather than entering them separately? There is also a lack of granularity for certain fields that seem particularly counter-intuitive, such as all legal data having to fit into pbcoreRightsSummary without separate fields for rights holder, access restrictions, etc. Nevertheless, PBCore is much easier to work with than MARC, and certainly better suited for moving images.

The Metadata Object Description Scheme, conveniently abbreviated as MODS (now on version 3.5), was developed by the Library of Congress as a sort of compromise between the complexity of MARC and the simplicity of Dublin Core (which, as previously mentioned, was the basis for PBCore). A very detailed crosswalk is available on the Library of Congress website for MARC to MODS mapping, though it does skip some valuable MARC fields, such as 347 (digital file characteristics). Unlike PBCore, MODS was not created with a focus on moving images, but like PBCore, it vastly simplifies the fields available in MARC. In practice I found MODS to be very similar to PBCore, and its basis in XML made its use particularly instinctive to me as a former website producer (to be fair, MARC and PBCore can also be expressed in XML, but apparently MODS can *only* be expressed in XML). MODS is far from perfect, however, particularly when it comes to digital video metadata. There are no fields in MODS for bit rate, aspect ratio, or file size, while PBCore covers all those and much more. I also could not find a field in MODS that would be suitable for "version" info, deciding ultimately that it would fit best in a "note" field, along with acquisition and preservation data.

Speaking of preservation data, all three standards were woefully lacking in this regard; in fact, none of them had a specific set of fields to house preservation details. This renders all three unsuitable for exclusive use by an archive; they would have to supplement one of these standards with something like PREMIS.

Ultimately, I found that PBCore to worked best with moving image material. If PBCore allowed certain fields to be used more than once, and added granularity in areas such as IP rights, I would select it as my preferred metadata standard without hesitation.