

Kelly Haydon

December 16th, 2011

Access to Collections

Instructor: Linda Tadic

### **Mapping MARC to Other Metadata Standards**

“MARC is the acronym for Machine-Readable Cataloging,” begins the Frequently Asked Questions section of the MARC website, “It defines a data format that emerged from a Library of Congress-led initiative that began nearly forty years ago.”<sup>i</sup> Early in the history of computers in the workplace, the MARC record was developed using a three-character numerical system that was easily readable by the machines of the day (and rightly assumed to be readable by those of the future). Software development, as you might have guessed, has come a long way since then, and in recent years this system of cataloging has been supplanted by metadata standards that rely on simplified tagged-texts field, readable by both humans and machines. This paper reports on a recent mapping exercise, pitting two other metadata standards against the MARC record: Dublin Core, and SMPTE-RP210.

The MARC record is not intuitive; the numerical codes for title, creator, location, etc. must be memorized or frequently referenced. Nor is it a standard, but a “transmission” of standards recommended by other initiatives<sup>ii</sup>. Increasingly, it is losing its hold on the world of library cataloging and I’ve heard more than one young buck describe it as a dead language. However, where MARC may weigh down cataloging with its hundreds of fields - many near-duplicating content - it excels in a level of itemized detail that saves a cataloguer from consolidating information. Granularity is the defining characteristic of a MARC record. The

Copyright section alone has 22 subfields, with sections as nuanced as the research date by which the copyright was identified and the agency that assigned the copyright.

SMPTE-RP210, developed by the Society of Motion Picture & Television Engineers, is a Metadata Element Dictionary for a number-based transmission standard that meant to be used in a production environment. Expectedly, the standard is highly itemized in physical and technical descriptors, and unlike simplified metadata standards such as Dublin Core, there is not a lot of guesswork or consolidated when mapping MARC physical data to SMPTE-RP210. SMPTE-RP210 supports technical metadata right down to the key frame of a picture element, as well as fields that can be pushed to marketing copy housing the content. While SMPTE anticipates cataloging and archival identifiers, it does not consider data applicable only in the years to come, such as the condition of the element or accrual procedures. However, the data dictionary does have a node specifically for user-define data that can be construed to the needs of the organization.

Unlike MARC or SMPTE-RP210, Dublin Core is not designed for transmission, but a true metadata standard, interoperable with a variety of program languages, and reliant on other standards at an itemized level (such as standards for Formats, Sound, etc). Dublin Core is ideal for publicly accessible metadata (such as World Cat) as it provides only key fields and qualifiers that are most frequently requested by users. The term Contributor applies to all participants in a film (other than Publisher or Creator) and allows for broad interpretations of proper formatting (William Shakespeare, for example, can be written that way, or written Shakespeare, William). In MARC, a performing cast is listed both as a block of text and as individuals, a good example of how the information is duplicated, in DC, the text box is unnecessary and only the individual names are carried over. As its general, non-specific structure is its strength; it is also its

weakness, particular in the area of describing format, which can only be noted in two forms: Format and the qualifier Extent. According to the Dublin Core website, Extent specifically refers to size or running time of the piece, but I'm unclear as to why. Extent can mean a variety of other things in MARC (including number of reels, is that considered a size? I'm not sure).

Mapping these three systems of describing content is often more of an art than a science, and I often made a best guess or chose to leave out the MARC field all together, assuming the idea in bringing one format to the other was too streamline. It is a bit unnerving to know that data is lost in these conversions, perhaps retained in the legacy format over time, but perhaps not. In either case, the best possible - and most helpful - information has been identified through decades of the MARC and in that way influenced the metadata standards of today.

---

<sup>i & ii</sup> Library of Congress. MARC Standards. 12 16 2011 <<http://www.loc.gov/marc/faq.html>>