

Metadata mapping exercise

Class: Access to Moving Image Collections
 Lecturer: Rebecca Guenther
 Student: David Neary

For this metadata mapping exercise I selected three data standards – Dublin Core, PBCore and MARC – which at first appeared to have considerable amounts in common, but the complexities of their compatibility were revealed as several elements of the vocabularies proved difficult, and sometimes impossible to map.

The graph below shows the process of mapping from the relatively simple Dublin Core model, through its expanded AV-centred successor PBCore, into the intimidatingly complex MARC standards. A discussion of my findings follows.

Dublin Core	PBCore	MARC
title	pbcoreTitle	Title - 245.10\$a
subject	pbcoreSubject	Subject Added Entry - 6xx
description	pbcoreDescription	Physical Description - 300.##\$a
type	pbcoreAssetType	Index Term - Genre/Form - 655.##\$a
source	N/A	Source - 337.##\$b\$2
relation	pbcoreRelationType	Additional Physical Form Available Note - 530.##\$a
coverage	pbcoreCoverage	Geographic Coverage Note - 522.##\$a
creator	pbcoreCreator	Main Entry - 1xx
publisher	pbcorePublisher	Name of publisher, distributor, etc. - 260.##\$b
contributor	pbcoreContributor	Added Entry - 7xx
rights	pbcoreRightsSummary	Information Relating to Copyright Status – 542
date	pbcoreInstantiationDate	Date of publication, distribution, etc. - 260.##\$c
format	pbcoreInstantiation	Content Type – 336
medium	pbcoreInstantiationPhysical	Physical Medium - 340.##\$sa
abstract	pbcoreDescriptionType	Summary, etc. - 520.##\$a
identifier	pbcoreIdentifier	Uniform Resource Identifier - 856.##\$u
language	pbcoreInstantiationLanguage	Language note - 546.##\$a
audience	pbcoreAudienceLevel	Audience Term - 385.##\$a
provenance	pbcoreRightsSummary	Ownership and custodial history - 561
rightsHolder	pbcoreRightsSummary	Copyright holder - 542.##\$d

Dublin Core: The simplicity of Dublin Core is entirely relative, with 55 metadata terms available to define catalogue entries. For this exercise, I was certain to select the 15 core metadata elements from which to work, as the easiness or difficulty of crosswalking them into other standards is key to understanding the complexities that can emerge in doing so.

Dublin Core's simplicity is also its weakness, with many terms bursting with information due to the breadth of their definition. For example, the Dublin Core website defines 'coverage' as "The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant."¹ Thus a series of "Coverage=" entries would be required under a DC catalogue entry, whereas PBCore offers a 'coverageType' list, while MARC offers several data sets under which various coverage categories could nest. For simplicity's sake, so as to not fill the MARC crosswalk entry for 'coverage' with a half-dozen entries, I have taken 'coverage' to refer to geographic region. Similar simplifications had to be made so as not to overwhelm the above table.

PBCore: PBCore is in many ways an evolution of DC, designed by public broadcasters in the US to help clarify non-precise DC vocabulary, especially as it applied to audiovisual elements. Instantly some useful clarifications can be seen. 'Type' in DC can describe "general categories, functions, genres, or aggregation levels for content"², which is far too broad a field, especially when describing AV assets, where genre and type are inherently different elements. Thus PBCore

¹ <http://www.dublincore.org/documents/usageguide/elements.shtml>

² Ibid.

offers 'AssetType' and 'Genre' as separate elements, allowing for more precise database searches.

PBCore also offers a wealth of descriptive terms under its 'Instantiation' heading, used for clarifying details of the dozens of formats audiovisual materials can be recorded on. This helps define the size of the item, where in the recording it begins, its length, etc. Many of these specific elements would all fall under 'format' if crosswalked back into DC. As such, for cataloguing moving images, PBCore is a far more suitable standard (as it was designed to be) than DC.

MARC: To the novice, MARC standards are almost incomprehensible, and their numbering codes make it less a filing system than a language unto itself. However, its remarkable breadth of entries allows for extensively detailed cataloguing, so that once the MARC rosetta stone has been cracked makes it is a very useful filing system indeed.

The problem with MARC is not that its scope is too broad, but that the places where certain elements are nested shows the standard to be needlessly complex. For example, the title of an entry in the MARC catalogue is categorised as 'Title', but is only found as a subfield of 'Title Statement', taking the code 245.10\$a. It seems strange that the most pertinent detail about a catalogue entry should fall under such a subheading so many details down.

What this crosswalk revealed is how the breadth of available terms, and their spread throughout the standard, makes MARC an unnecessarily complicated cataloguing tool. Whereas PBCore collected all of its object details under 'Instantiation' in a series of clarified fields, MARC spreads these liberally through its various subheadings. Information from 'format' in Dublin Core could

translate to MARC as 'Physical Description' (300\$a), 'Physical Description of Reproduction' (533\$e), 'Electronic Format Type' (856\$q), or 'Material Base and Configuration' (340\$a). It is not that these clarifications are not welcome and sorely missing in DC, but that spreading them throughout the MARC system as has been done does not make for an easily accessible catalogue. Particularly, MARC's inability to clearly label the roles of contributors to a work sets it behind PBCore in its usefulness to cataloguing moving images, as PBCore's 'contributor' element comes with the sub-element 'contributorRole'.