Emma Heck
SI 676
Metadata Application Profile
November 20, 2023

Introduction

For my Omeka site, I have chosen to use the "Cats" collection from LOC's Free-to-Use libraries. Below is a metadata crosswalk that explains the crosswalking of metadata for the Cats collection to Dublin Core (and one MODS) elements. I chose to use Dublin Core for most of the target elements to allow for some flexibility when adapting the source elements. The metadata crosswalk guidelines are meant to provide instructions for creating metadata for digital records.

Metadata Crosswalk

Collection: https://www.loc.gov/free-to-use/cats/

Source field name	Source field path/dict name	Target	Target namespace	Notes
call_number	item[call_number]	dc:identifier	DC Element	Physical location of item
creator	item['contributor_names]'	dc:creator	DC Element	Contributor names
date	item['date']	dc:date	DC Element	Creation date
description	item['description']	mods:physicaldescription	MODS	Format type and sometimes a brief description of what the image contains
format	item['format']	dc:format	DC Element	E.g. photo, book, etc.
language	item['language']	dc:language	DC Element	
mime_type	item['mime_type']	dc:type	DC Element	
partof	item['partof'][0]	dc:publisher	DC Element	Entity/collection of which the item is part
subject	item['subject']	dc:subject	DC Element	At least one term to describe what the image is about
title	item['title']	dc:title	DC Element	

Reflection

Once I had a better understanding of the process, I found the assignment to be very doable. It was definitely a challenge in that I haven't done anything similar before, but I feel confident about the fact that I was able to (mostly) successfully generate the csv for Omeka. A few problems arose, the first being that I was unable to extract the correct information for the contributor_names field. All of the rows returned 'not found,' except for two where the value was just 'creator,' and one with 'contributor_names.' I also had trouble with the header name. I first named the field 'creator,' but then decided to change it to 'contributor_names.' However, when trying to write the updated csv to reflect those changes, the headers would not update. I was able to debug these issues by just deleting the original csv file that I had rewritten over, and running the cell again with my updated changes. This step fixed the majority of the problems and returned the correct data for the contributor_names field, including listing 'not found' when a contributor name was not present. For the next part of the process, I plan to clean up the data to remove unnecessary brackets and quotation marks, as well as change the capitalization of some of the values (like the collection name in 'partof'). I also plan to remove unnecessary metadata. For example, the 'partof' column provides metadata for the collection to which the item belongs. The first item's partof value was extracted as follows: {'count': 16870, 'title': 'genthe collection', 'url': 'https://www.loc.gov/collections/genthe/?fo=json'}. However, the only piece I actually need is 'title': 'genthe collection.'; the count and url are not relevant to the 'partof' tag. Overall, the process of extracting the metadata was a lot more straightforward than I imagined and now that I am on the other side I definitely see how this process will make future work much faster and easier.