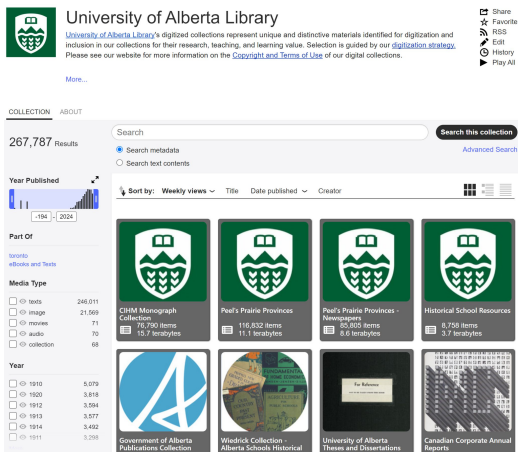
Harmonizing Local Metadata Practices for Library Resources on the Internet Archive

John Huck & Luc Fagnan
University of Alberta Library
DCMI 2024, Toronto, Oct. 20-23, 2024

Hello and good afternoon everyone, welcome to our presentation on Harmonizing Local Metadata Practices for Library Resources on the Internet Archive.

Introduction

- What is the **Internet Archive** (IA)?
- University of Alberta Library and IA
  - Partnership ~**2008**
  - Over **60** collections
  - Nearly **270,000** items



To start off, we'd like to provide you with a bit of background on the Internet Archive and its relationship with the University of Alberta Library. The Internet Archive is an American nonprofit digital library that was founded in 1996, and has since been providing free access to a wide variety of digitized materials. The University of Alberta Library began its partnership with the Internet Archive around 2008, and has been uploading materials directly to the platform since around 2016. The University now has over 60 collections on the platform, which amount to nearly 270,000 individual items altogether.

See https://archive.org/details/university_of_alberta_libraries

Local Workflows

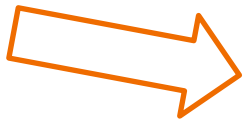
- Several units contributing to our collections in IA
  - Different sources and requirements
- Various domain standards
  - Libraries, archives, etc.

There are several different units within our Library that contribute materials to the platform. These include our University Archives, Special Collections, Digitization units, and various other units.

Though we do our best to coordinate our efforts, it can be challenging at times to make metadata practices consistent as each unit adheres to their local requirements. Of course, with a variety of metadata sources also come various information domain standards from each of them.

Internet Archive Metadata Schema

- Minimal constraints
- Unlimited custom fields
  - 40+ in our collections



Field
citation_use_notes
citation-use-notes
air-date
approval_date
record_created
recording-date
alternate_identifier
identifier-local
polar_pam
ual_identifier
box_id
box_item_number
box_number
folder
folder_id
location

Other challenges stem from the way our units use the Internet Archive's metadata schema. The schema provides flexibility with minimal constraints, which can be ideal for user-friendliness but sometimes detrimental for consistency of metadata. Of particular note is the ability to add seemingly unlimited custom fields, meaning a user can create a new field whenever they upload an item to the platform, intentionally or not. On the slide here, we have a sample of some custom fields that we've observed in our own collections, some of which show unintentional repetition. In other cases, simple typos have resulted in extra fields that go unnoticed.

About 1 year ago, the Metadata team decided to undertake a project to harmonize our metadata practices on the Internet Archive.

Project Goal: Create a unified metadata model for all resources uploaded to the Internet Archive by the library

Our project goal is to create a unified metadata model that we can use for all resources that our library uploads to the Internet Archive.

Metadata Application Profile (AP)

- Adds constraints to a metadata vocabulary for a specific use
- **Singapore Framework** defines AP components

Because of the flexibility of the Internet Archive's Metadata Schema, we naturally thought of Application Profiles as a way of adding constraints and codifying our use of it.

The Singapore Framework defines components of an Application Profile. These include a set of Functional Requirements and a Domain Model.

Functional Requirements

- Findability (IFLA user tasks **Find** and **Identify**)
- Metadata creation from MARC and MODS
- Templates for data entry
- Guidelines for custom fields

Over the past six to nine months we've been consulting with the library units involved with the Internet Archive work to learn about their needs and translate that information into Functional Requirements for an Application Profile.

While the stakeholders expressed broad support for the project, they did not articulate any particular requirements for an Application Profile, perhaps because it was a concept they had not previously encountered.

Stakeholders did have an expectation that metadata should provide a basic level of Findability, which we have taken as a core functional requirement, and have characterized as corresponding to the user tasks Find and Identify from IFLA's Library Reference Model.

Because many of the digitized materials come from our physical collections or have been hosted on library platforms for years, the need to accommodate the creation of metadata from existing records in a variety of formats is another requirement.

Stakeholders asked for spreadsheet templates to help with data entry when creating metadata from scratch, which gives us an excellent reason to use the Tabular Application Profile format.

Lastly, stakeholders requested guidelines for the use of custom fields. Custom fields present an interesting challenge, but there's no reason to think they couldn't be part of an Application Profile.

Domain Model

- Identify local, informal material types
- Align with types from relevant metadata standards and IA schema
- Use **Z39.29 - Bibliographic References** as guide for AP elements

With these requirements in mind, we began work on a Domain Model for the types of resources the Application Profile should cover. Our concept for the Domain Model has three steps.

The first is to identify a set of local, informal types or categories that make sense to our stakeholders and that reflect already uploaded material. For instance, we have a lot of digitized historical postcards, so 'Postcards' is a category our stakeholders recognize.

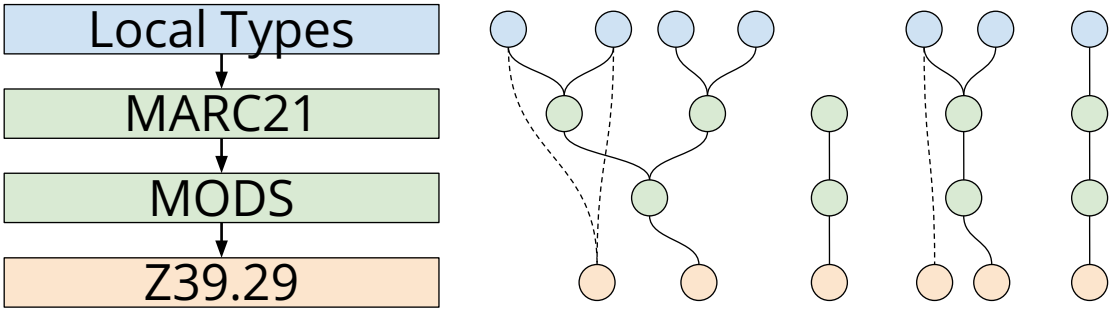
The second part is to align these types with record and material types from the standards of our existing metadata, the Internet Archive schema, and others.

The third part is to make mappings to corresponding material

types from a standard for Bibliographic References, NISO Z39.29.

This standard provides convenient lists of recommended metadata elements for a range of resources, and is intended to support resource citation. We plan to use it as a guide when selecting metadata elements for the application profiles.

Align Material Types

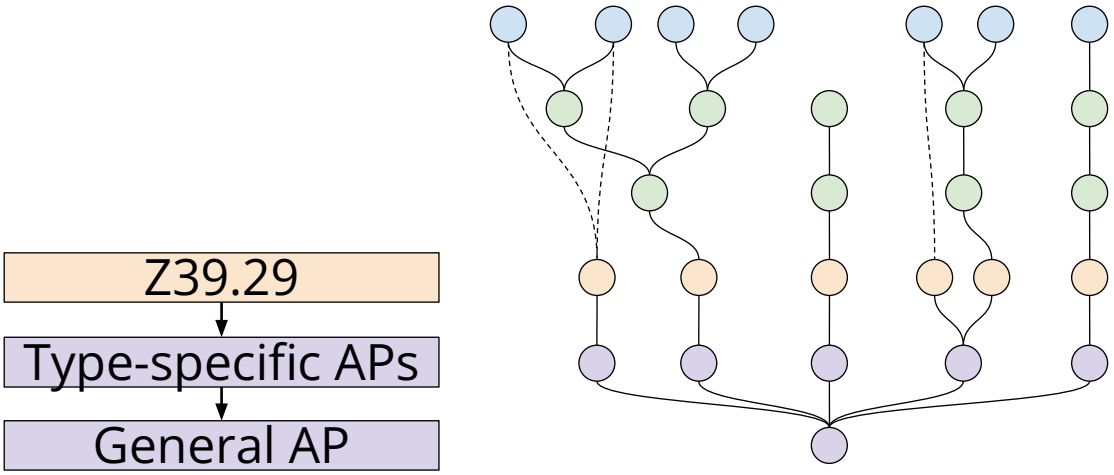


Here you can see how the alignments are arranged like a layer-cake. Each layer represents a set of types from a different standard. Directional mappings funnel down from Local Types through to Z39.29.

On the right is an illustration of the alignment struture concept. Each dot represents a type and lines show alignments between them. There won't always be a one-to-one mapping between types, but that's alright, as long as every type has a path to a Z39.29 type.

The benefit of incorporating all the various types into our domain model is that metadata we pull from existing records will automatically be compatible with the domain model.

Use Z39.29 as Guide for AP Elements



Our plan is to use Z39.29 as the basis for a series of type-specific application profiles. It might be possible to boil them down into a single, general-purpose application profile, but that remains to be seen.

Takeaway

- Leading vs. following in **direction setting**
  - Finding the right balance between **standards** and **reality**

Please get in touch: metadata@ualberta.ca

Thank you!

So to wrap things up, we'd like to end on a takeaway. One key challenge we've encountered in this work is the balancing act between leading and following when it comes to direction setting. This involves determining who makes the decisions on what practices to follow when standardizing, or negotiating with our stakeholders to gain their consent when making such decisions. Ultimately, we're working towards finding the right balance between the implementation of standards, and the reality in which our colleagues carry out their day to day work.

And with that, thank you so much for your attention, and please don't hesitate to get in touch with any questions.

References

Internet Archive. "Metadata Schema." Accessed October 2024. https://archive.org/developers/metadata-schema/.

National Information Standards Organization. *Bibliographic References*. ANSI/NISO Z39.29-2005 (R2010). NISO. Approved June 9, 2005; reaffirmed May 13, 2010. https://www.niso.org/publications/ansiniso-z3929-2005-r2010.

Nilsson, Mikael, Tom Baker, and Pete Johnston. *The Singapore Framework for Dublin Core™ Application Profiles*. Dublin Core Metadata Initiative, 2008. http://www.dublincore.org/specifications/dublin-core/singapore-framework/.

Riva, Pat, Patrick Le Bœuf, and Maja Žumer. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. International Federation of Library Associations and Institutions, 2018. https://repository.ifla.org/handle/123456789/40.

University of Alberta Library. "Digitized Collections." Accessed October 2024. https://www.library.ualberta.ca/digital-initiatives/digital-collections.