DCPAPERS

Harmonizing Local Metadata Practices for Library Resources on the Internet Archive

John Huck[1,*,†] and Luc Fagnan[1,†]

[1] *University of Alberta Library, Cameron Library, Edmonton, Canada*

Abstract

The University of Alberta Library began a project in 2023 to consolidate and harmonize the library's metadata practices for material uploaded to the Internet Archive (IA). The broad range of metadata standards and requirements associated with the over 240,000 items uploaded by the library, along with the flexibility of the metadata schema designed by IA, presented a unique set of challenges to overcome in the project. With the intent of creating one or more metadata application profiles for IA's schema, the library's Metadata Strategies Unit initiated work to identify the needs of library stakeholders and created an initial set of functional requirements to guide the project. This project report provides an overview of the metadata problem and project design, an update on the work, as well as insights gained by the project team.

Keywords

metadata guidelines, metadata application profiles, digital repositories, Internet Archive

1. Introduction

The University of Alberta Library has maintained an active digitization program for over twenty years [1]. In recent years, the Internet Archive's (IA) digital library platform has become a key access point for the library's digital collections. Since 2008, the library has uploaded over 240,000 items to the platform, including monographs, newspapers, maps, historical postcards and archival material. Many, but not all, of these items were previously published on the library's own digital collection platforms.

The metadata treatment of this material has varied over time, and in 2023, the Metadata Strategies Unit (MSU) began a project to consolidate and harmonize the library's metadata practices for material uploaded to IA. This project report provides an overview of the metadata problem and project design, an update on the work, as well as insights gained by the project team.

* Corresponding author.
† These authors contributed equally.
john.huck@ualberta.ca (J. Huck); ltfagnan@ualberta.ca (L. Fagnan)
0000-0003-1150-5402 (J. Huck); 0009-0000-2014-9732 (L. Fagnan)

https://doi.org/10.23106/dcmi.952447592

2. Project Description

2.1. Metadata problem

Various units within the library and stakeholders external to the library have contributed open-access material to the library's collections in IA. The metadata associated with this material and the collections in which they are held adhere to a variety of requirements and information domain standards. As a result, metadata is inconsistent across the items and between collections created and maintained by the library, making it challenging for users to navigate and difficult for the library's staff to manage.

IA's metadata schema is designed as a flexible framework with minimal constraints, allowing the platform to accommodate a wide range of material and the use of custom fields. The schema provides useful information on proper implementation of IA's fields, but additional, undocumented behaviours have been observed, making it difficult at times to predict how metadata will be displayed and organized. The flexibility of the schema and its sometimes unexpected behaviours make it challenging to implement consistent metadata practices across IA.

2.2. Project design

With broad agreement among library stakeholders that a harmonized approach to metadata would be desirable, the library's Metadata Strategies Unit (MSU) initiated a project to pursue this goal. Given IA's relatively flexible metadata schema, an early assumption was that the project would involve the creation of one or more metadata application profiles.

Metadata application profiles are a means of specifying constraints on a metadata schema for a particular application. The Singapore Framework for Dublin Core Application Profiles [2] defines basic elements of an application profile, which include an enumeration of its functional requirements and a domain model, as well as the description set profile and optional guidelines.

Malta and Baptista [3] developed a method for applying the Singapore Framework, which elaborates on the process of creating these parts of an application profile. More recently, the Dublin Core Application Profiles Working Group has developed a specification for tabular application profiles [4].

2.3. Preliminary work

Drawing loosely from this work around application profiles, the project team began the process of developing the project's vision, work plan, and an initial set of functional requirements. While MSU had previously completed work to characterize the extant metadata and consult with library stakeholders, it became clear to the project team that, in order to create a complete set of functional requirements, further preparatory work would be needed to identify all the library stakeholders, and, more importantly, their relationships to specific collections of material, as well as the degree to which they felt a sense of ownership over the existing metadata, and the role they would play in the project.

This project report provides an overview of the metadata problem and project design, an update on the work, as well as insights gained by the project team.

https://doi.org/10.23106/dcmi.952447592

References

[1] University of Alberta Library, Digital collections, 2024. URL: https://library.ualberta.ca/digital-initiatives/digital-collections.

[2] M. Nilsson, T. Baker, P. Johnston, The Singapore framework for Dublin Core™ application profiles, 2008. URL: https://www.dublincore.org/specifications/dublin-core/singapore-framework/.

[3] M. C. Malta, A. A. Baptista, Me4DCAP V0.1: A method for the development of Dublin Core application profiles, Information Services & Use 33, 2 (2013): 161–171. doi:10.3233/ISU-130706.

[4] K. Coyle et al., DC tabular application profiles (DC TAP): Primer, 2023. URL: https://www.dublincore.org/specifications/dctap/primer/.

https://doi.org/10.23106/dcmi.952447592