

Cornell Project Proposal

Overview

Cornell University Library (CUL) proposes to undertake work in the following two areas.

- Community development of library ontology extension for the rare materials community
- Original metadata creation for noncommercial LPs from the Afrika Bambaataa collection

The focus of the original LD4L project was on transforming legacy metadata to explore how it can be leveraged in an RDF environment. However, legacy metadata will invariably have built-in constraints reflecting the context in which it was originally created. We believe it will be possible to create richer metadata if we aim to create RDF descriptions from the outset, and create our data models with original production in mind. The ability to take advantage of data published by nonlibrary sources can not only yield production economies, but also allow library metadata to connect more effectively with other information on the web.

The potential of linked data approaches is, in many ways, best examined through their application to specialized domains, which offer the opportunity to extend existing vocabularies and reach a well defined user community. CUL proposes two related projects pursuing this approach, each in collaboration with domain partners.

Rare Materials Ontology

Working with the RBMS Bibliographic Standards Committee (BSC), Princeton University and Columbia University, Cornell plans to define an extension ontology for the description of rare materials. The data particularly relevant to rare materials often pertain to the instance and item level; however, item-level data (e.g. provenance, binding) are not well defined by current library linked data models. Indeed, the original BIBFRAME model lacked a true Item class altogether.

The rare materials community has a tradition of extending mainstream library vocabularies; however, this work was performed within the constraints of legacy environments and legacy standards, which have not always well supported these materials. Linked data, with its emphasis on extensibility as well as reusability, offers a highly promising framework for addressing this long-standing need.

The project ontology team will build on work underway by BSC to identify concepts important for the description of rare materials as well as work by the Special Collection Ontology Language Project. The BSC effort builds on the Descriptive Cataloging of Rare Materials series produced by BSC and used extensively in the cataloging of rare materials. Cornell proposes to work with partner organizations to take the conceptual work produced by BSC and create an extension ontology. By working with BSC, the project partners will follow the LD4P philosophy of incorporating community engagement into the process with the intended goal of community adoption. Following the development of the ontology, we will engage catalogers at a variety of institutions to test the model and provide feedback.

Objectives

- To investigate how BIBFRAME can be augmented by other existing vocabularies for describing rare materials and sound recordings, and to publish a working ontology for this purpose
- To investigate further data modeling for handling materiality, provenance, annotations and other concepts relevant to describing rare materials
- To engage with the rare material community on the development of an ontology extension
- To publish a vocabulary in which the rare materials communities can publish their data as linked data

Native Linked Data Creation for Bambaataa LPs

Cornell proposes to perform original metadata creation for approximately 1,000 non-commercial LPs in the Afrika Bambaataa collection natively in RDF (Resource Description Framework). This collection forms part of Cornell's extensive archive devoted to this founding figure of Hip Hop culture, which is in turn part of Cornell's larger Hip Hop Collection. This collection supports a growing body of international scholars and educators who study Hip Hop not only for its global influence on popular music, art, and style, but also for its role in articulating social and political issues. As such, it offers a promising testbed for exploring the use of linked data techniques to extend the reach of descriptive metadata. Legacy MARC and EAD metadata for other parts of the collection will also lend itself to linked data conversion, reconciliation, and enhancement work, building on experience gained in this area in LD4L.

This metadata creation effort will make use of a customized Vitro instance, which will have the LD4L BIBFRAME ontology loaded with appropriate extensions as outlined below. An important aspect of the proposed work is not only to test the practical application of the revised BIBFRAME ontology developed as a result of LD4L, but also to consider how it can be extended by other ontologies. A further objective to be pursued through the associated LD4L Labs proposal (see Section 3 of the LD4L Labs narrative) is to prototype public-facing services that will leverage these richer descriptions to offer integration with non-library resources.

This work will complement the Stanford LD4P proposal, in association with the Music Library Association and other partners, to develop a Performed Music Ontology based on BIBFRAME. The Performed Music Ontology aims to extend BIBFRAME to accommodate performed music. The Cornell Bambaataa project will use this model as a basis for further local development catering to the specific needs of the Bambaataa noncommercial LP collection, with a particular emphasis on areas not fully expressed in current content standards for music cataloging, such as events and annotations. In addition, the Cornell project will have a substantial in-house tool development and original metadata creation component. The Cornell project will thus provide an opportunity to test the cataloging of performed music materials in the field using native RDF creation tools as well as examine how the additional needs of a local collection can be met.

The Bambaataa project will also be undertaken concurrently with the Rare Materials Ontology project and will similarly take into account its findings.

Objectives

- To produce a prototype RDF cataloging tool based on Vitro and progressively improve it through an iterative process of development and in-the-field use
- To produce natively created RDF descriptions for approximately 1,000 noncommercial LPs in the Afrika Bambaataa collection
- To evaluate, in collaboration with Stanford's Performed Music Ontology group, current BIBFRAME provisions for describing sound recordings and special collections material, and to provide further input on BIBFRAME development
- To investigate augmentation and integration of LC value vocabularies with those from other sources (e.g. MusicBrainz)
- To test integration of natively created RDF descriptions with descriptions generated through conversion from MARC