From DC to Linked Data: Some Milestones on OCLC’s Journey through the Post-MARC Landscape

Carol Jean Godby
Senior Research Scientist
From Geeks, Freaks, and People in Sensible Shoes to the DC-2016 PCC URI RWO TG

encoding application profiles in a computational model of the crosswalk

Carol Jean Godby
OCLC, USA
godby@oclc.org

Devon Smith
Eric Childress

Abstract
OCLC’s Crosswalk Web Service, as defined in the Dublin Core Metadata Vocabulary (DCMI) terms, provides a crosswalk that encodes an author and annotates the namespace 'relation between: Terms spatial.' This informs the Equivalence Expression Language (EEL) and generate production-grade triples added, and redefined, the author translation model of the EEL mapping one pair of elements:

Keywords: application profiling, MARC, metadata crosswalks

Library Linked Data in the Cloud
OCLC’s Experiments with New Models of Resource Description

Carol Jean Godby
Shenghui Wang
Jeffrey K. Mixter

Special Session Title: PCC Task Group on URIs in MARC Report
Day/Time: TBD (October 13-16, 2016)
Conveners:
Jackie Shiah, Coordinator, Resource Description Group, George Washington University Libraries, http://orcid.org/0000-0003-3214-8546
Terry Reese, Head, Digital Initiatives, Ohio State University Libraries, http://reese@net; http://orcid.org/0000-0002-0180-5388
Session Language: English

Target audience: Librarians, technologists, administrators, data provider, publishers

Abstract 30-50 words: The PCC Task Group Report will report on the work from the inception of the TG to date. This will include a technical discussion around entitlement of MARC library data and comments on the status of existing infrastructure within the Libraries to support this work.
OCLC’s published linked datasets

- WorldCat Works
- WorldCat Catalog
- FAST
- ISNI
- VIAF
The library community was there from the start.

Building frameworks and Web vocabularies

Publishing datasets

A dozen years were required to build the necessary Semantic Web standards.

Conventions and best practices followed.
“The Dublin Core is not intended to supplant other resource descriptions, but rather to complement them.”

The Dublin Core metadata elements describe essential features of electronic documents (or “document-like objects”) that support resource discovery…

Dublin Core descriptors are intrinsic, extensible, syntax-independent, optional, repeatable, and modifiable
An explosion of Web standards Development

...And first glimpses of life after MARC
A Metadata Registry for the Semantic Web

Rachel Heery
Research & Development
UKOLN
<r.heery@ukoln.ac.uk>

Harry Wagner
Office of Research
OCLC / DCMI
<wagnerh@oclc.org>

Abstract

The Semantic Web activity is a W3C project whose goal is to enable a web of content that has clear-cut, unambiguous meaning. This vision is based on a declaration of schemas in metadata registries. However, many of the issues regarding registries are challenging. Additionally, registry issues are often difficult to describe and comprehend without a working example.

What Are The Benefits?

- Authoritative source of information
  - Current and trusted information
- In-depth information about metadata
  - Relationships between terms
  - Vocabulary classification schemes
- Extensibility
  - Terms, application profiles, documents, etc.
  - Add new application services
- Metadata reuse & inter-registry cooperation
  - Discovery is key to reuse
- Internationalization
  - Registry designed as an i18n application
  - i18n tools are integral part of application
- Best practice
  - Interaction promotes best practice
- Affiliate program
A maturing OCLC research project

Encoding Application Profiles in a Computational Model of the Crosswalk

Abstract

OCLC’s Crosswalk, a crosswalk that extends Dublin Core to MARC, can be used to annotate terms for MARC, metadata, and online cataloging. The Crosswalk uses equivalence terms to match with MARC records. Equivalence terms also support the translation of MARC records into Dublin Core and vice versa. Keywords: Dublin Core, MARC, metadata
Pointers to a new direction

LCSH, SKOS and Linked Data

Ed Summers
Library of Congress, USA
edsu@loc.gov

Clay Redding
Library of Congress, USA
cred@loc.gov

Abstract
A technique for converting Library of Congress Knowledge Organization System (SKOS) RDF are highlighted, as well as possible points for extending vocabularies such as Dublin Core. An application of data on the Web is also described.

Keywords: metadata; semantic web; controlled vocabularies; identifiers

1. Introduction
Since 1902 the mission of the Cataloging in Publication (CIP) program was to provide a list of library materials typically in the American libraries around the United States, and the world. The cataloging of library materials typically includes a description of items and their characteristics. Subject cataloging on the other hand is a technique for organizing items like the Library of Congress Subject Headings (LC Classification) that are used in describing the subject matter of an item. This technique is intended to make the subject matter of the item easily understandable.

The purpose of this paper is to describe a technique for converting Library of Congress Subject Headings (LCSH) into SKOS that enables the data to be linked to other SKOS vocabularies. This technique is based on the concept of Linked Data, which is an extension of the Semantic Web.

Cool URLs for the DDC: Towards Web-Scale Accessibility of a

FAST Linked Data
FAST (Faceted Application of Subject Terminology) is an enumerative, faceted subject heading schema derived from the Library of Congress Subject Headings (LCSH). The purpose of adopting the LCSH with a simplified syntax to create FAST is to retain the very rich vocabulary of LCSH while making the schema easier to understand, control, apply, and use.

The schema maintains upward compatibility with LCSH, and any valid set of LC subject headings can be converted to FAST headings.

FAST Linked Data
Linked Data is one of the underpinnings of the Semantic Web, the effort to make the meaning of information on the Web more understandable to computers. These Linked Data authorities are formatted using schema.org and SKOS (Simple Knowledge Organization System).

The FAST Authority file contains links to LCSH Authorities as well as other authoritative sources such as VIAF, GeoNames, and Wikipedia. We will continue to add other links where possible.

The use of the FAST authorities is open and fast is made available under an Open Data Commons Attribution (ODC-BY) License. It is also available for download under this license at http://www.oclc.org/research/activities/fast/download.htm. OCLC will update FAST periodically, at least twice yearly.

FAST Linked Data API
The SRW/SRU API to access the FAST Linked Data is described at the OCLC Developers Network: https://www.oclc.org/developers/developer/web-services/fast-api/linked-data.en.html

OCLC Experimental
Congress Subject Headings (LC-SH) available for computer processing as MARC, and more recently as MARCXML. The conventions...
Storm clouds gather...

And so forth... and so on

Problem: Inadequate representation of bibliographic metadata.

Effect on the cataloging process: Difficulties in accurately and effectively identifying and retrieving information.

Effect on the end user: Time-consuming and potentially frustrating searches for information.

Problem: Metadata is stored in a standard format and is not easily transferable to other systems or applications.

Effect on the Crosswalk: It is necessary to maintain a large number of mappings that are not used.
An early status report

The seminal statement of why librarianship is compatible with linked data principles…

A call to action…

…by an international blue-ribbon committee of linked data experts.

Nearly all are longstanding members of the Dublin Core community.
Drafting Linked Data statements with DC vocabulary

Source: “LCSH, SKOS and Linked Data” (Summers, et al 2008)

Source: “Library Linked Data in the Cloud” (Godby, et al. 2015)
Analysis of International Linked Data Survey for Implementers

Karen Smith-Yoshimura
OCLC Research
smithyok@oclc.org

DOI: 10.1045/july2016-smith-yoshimura

Printer-friendly Version
“DCMI @ 20: Past, Present, and Future”

Keynote presentation by Eric Miller
Applying a Linked Data-Compliant Model: the Usage of the Europeana Data Model by the Deutsche Digitalbibliothek.
Applying Concepts of Linked Data to Local Digital Collections to Enhance Access and Searchability.
Linked Data Mapping Cultures: An Evaluation of Metadata Usage and Distribution in a Linked Data Environment.
The 1:1 Principle in the Age of Linked Data.
Training the Trainers for Linked Data.
BIBFRAME – Expressing and Connecting Bibliographic Data.
Ecco! A Linked Open Data Service for Collaborative Named Entity Resolution.
Some enduring ideas from Dublin Core…

- Can be modified and extended.
- Interoperable. Designed for reuse. Coexists with other standards.
- Dublin Core is a vehicle for contributing expertise from the library community to the web (and vice versa).
• A vocabulary that uses state-of-the-art metadata management practices
• An international standard
• A social network
• A community of practice
Comments?

Carol Jean Godby
Senior Research Scientist
godby@oclc.org

Because what is known must be shared.