Princeton Project Proposal

Overview

In March 2015, Princeton University acquired the personal library of Algerian-born French philosopher Jacques Derrida (1930-2004). Of the roughly 16,000 published books and other items in the library, a significant number have been heavily annotated, and many bear personal dedications to Derrida from other philosophers and theorists, including Roland Barthes, Giorgio Agamben, and Emmanuel Levinas. To date, approximately three thousand items with dedications have been identified. Taking this unique collection as a proving ground, the overarching goal of Princeton's LD4P project is to explore, develop, and implement linked data standards for the description of special collections materials. The project will focus on three general areas:

Foundations

The project will involve collaboration with colleagues from Columbia University Library, Cornell University Library, and the Bibliographic Standards Committee of the American Library Association's Rare Books & Manuscripts Section (RBMS). The goal of the project will be to develop models, vocabularies, and best practices for the item-level description of special collections materials using RDF. An emphasis on the item as a physical entity is central to this subproject because standards for copy-specific metadata have not yet been adequately defined by other linked data efforts in the library domain.

Princeton's work will be informed by current initiatives like the Linked Open Data for Special Collections project of the University of Illinois at Urbana-Champaign (UIUC). UIUC's work will be particularly valuable as a foundation for developing crosswalks to convert existing metadata formats, like Encoded Archival Description (EAD), to BIBFRAME. However, Princeton's focus is distinct from UIUC's in that its efforts will center, primarily, on original resource description for special collections materials, whereas UIUC's work is geared toward digital collections and "best practices for transforming legacy metadata," with an emphasis on the schema.org vocabulary employed by major Web search engines. Princeton's work, by contrast, will help address the need for higher-level data modeling in the special collections domain. In this regard, it forms part of the larger LD4P effort to develop domain-specific extensibility mechanisms for the core BIBFRAME vocabulary.

First, project participants will evaluate the current BIBFRAME model and vocabulary in relation to rare books cataloging and identify areas for extension or modification. Their focus will then turn to ontology and standards development in the rare books and manuscripts domain, addressing the gaps that have been identified. One area of resource description that merits particular consideration—and that is particularly relevant to Princeton's work with the Derrida collection—is that of annotations. Although emerging standards like the W3C's Open Annotation Data Model and Web Annotation Data Model have made strides toward representing and sharing user-generated annotations on the Web, these high-level protocols do not address the semantics of manuscript annotations specifically, or the relation between original and transcribed annotations (as reflected, for example, in digital facsimile editions).

Specifications

The project will build on existing work to define data conversion routines and help specify requirements for Web-based linked data creation tools. It is anticipated that the project's data modeling work will lead to extensions and refinements of the conversion tools that have been developed by the Library of Congress, and to the creation of new metadata crosswalks and conversion routines to share with the broader library community. For example, Princeton's Rare Books and Special Collections Department is currently working to create a finding aid for the Derrida collection using the EAD format. This data will be leveraged as part of Princeton's LD4P participation, and a basic set of EAD-to- BIBFRAME conversion scripts will be created. Princeton already has a set of stylesheets to transform EAD components into stand-alone bibliographic records in an XML format, based on analysis for a project to incorporate data from finding aids into the Primo discovery system. Thus, new work on EAD-to- BIBFRAME would primarily fall into the area of developing new properties to account for hierarchical description. In addition, as part of Princeton's involvement with the BIBFRAME testbed initiative organized by the Library of Congress, its Cataloging and Metadata Services division has already begun to develop an experimental linked data editor using the W3C XForms standard. As an independent activity in parallel to Princeton's LD4P involvement, initial development of the editing tool, called the Cataloger's Workbench Editor, will be completed, with the goal of making it available for others to test.

Production

Princeton will create original resource descriptions for a representative selection of items in the Derrida collection that include personal dedications addressed to Derrida. The items with dedications were significant to Derrida: he filled several rooms with them, stored in alphabetical order by author. The relationships encoded in these dedications will allow project participants to produce an RDF data set that can be used by scholars who are interested in studying Derrida's social and intellectual networks.

Objectives

- Evaluate existing ontologies and models, specifically BIBFRAME and the Web Annotation Data Model.
- Work with the RBMS Bibliographic Standards Committee and Cornell University Library to define an extension ontology (RBMO) for the description of rare materials.
- Transform and convert EAD data and MARC records into BIBFRAME for a representative selection of items in the Derrida collection.
- Create original or enhanced RDF descriptions for items in the previously identified subset of the Derrida collection that contain personal dedications.
- Prototype a lightweight, standards-based online editing tool for linked data creation and enhancement.

Community

- Princeton LD4P team (librarians from Princeton University's Cataloging & Metadata Services and Rare Books & Special Collections); Project coordinator: Princeton Cataloging & Metadata Services Director
- Cornell University Library LD4P team
- Representatives from the RBMS Bibliographic Standards Committee
- Princeton Center for the Digital Humanities and related faculty

Deliverables

- BIBFRAME extension ontology for original resource description of special collections materials
- Data set comprising original or enhanced descriptions, published to a project triple store or made available through a published RDF data dump
- EAD-to-BIBFRAME conversion scripts, published to a code-sharing repository like GitHub
- Code and documentation for the Cataloger's Workbench Editor, published to a code-sharing repository like GitHub