

Production workflows for native linked data descriptions (WP5)

Summary and background

The partners will develop generalizable workflows for production scenarios that are not addressed by workflows based on vendor data pipelines or legacy metadata conversion. We will use local collections representative of the breadth and depth of research libraries' materials to refine these workflows with real items and use cases, and to build upon previous experimentation with RDF-based description and cataloging workflows in music, film, maps, annotated books, and other high-demand, unique collections, in some cases pairing metadata workflows with local digitization efforts.

Harvard Library collections

Harvard Library has selected sets of materials for this project: a subset of its film collection that formed the basis for past linked data experimentation and ontology development; a subset of its maps collection; and a collection from the Eda Kuhn Loeb Music Library designated for an already funded digitization pipeline.

Working with the film collection will allow us to model best practice for end-to-end production workflows with RDF-based description, yielding a dataset with considerable potential for links between the Library's RDF data and Wikimedia, connecting our entities to the broader semantic web. The resulting data will supplement the converted moving image dataset as part of our efforts to integrate linked data within vended discovery environments.

The work on the maps collection will serve as a scalable prototype for libraries to share larger collections and different types of digital content with Wikimedia. The chosen collection includes historic paper map digital images that support the development of a model workflow for integrating linked data descriptions that feature historically significant geospatial elements with Wikimedia. Harvard will work with Wikimedia to expose digital image content on the web, with relevant metadata and connections to Wikidata entities, and explore enhancement of converted legacy descriptions in the cloud-based editing environment (WP1) as well as native Wikidata tools.

With digitization being a key means for libraries to open their unique primary source materials to the wider world through accessibility on the open web, work on the collection from the Eda Kuhn Loeb Music Library will allow the Harvard LD4P team to identify special considerations when pairing linked data description workflow with digitization.

Cornell University collection

Cornell has selected a subset of its music recordings collection to pilot native RDF cataloging workflows. The collection of 45-rpm vinyl records by Frank Sinatra and other popular music recording artists of the 1950s-1980s was selected because of the considerable potential for links between our RDF data and Wikimedia. We will pilot native RDF cataloging of these recordings within the cloud-based editing environment (WP1), providing iterative feedback to the development team. As part of this work, will investigate Wikimedia and additional data sources for connecting our entities to the broader semantic web. The resulting data will supplement the converted music dataset as part of our implementation of a discovery environment, affording the opportunity to assess natively created RDF data alongside converted data, a core question we've begun to address alongside partners during the LD4P Phase 1 project.

Lacking traditional MARC data, these materials present an opportunity to bring finer-grained descriptions and Wikidata integration into production workflows, and to assess their effectiveness in meeting discovery needs not addressed by current descriptive practices. The work will provide iterative feedback to the cloud-based editing environment development team.

The results will supplement the dataset available for our implementation of a discovery environment with highly structured descriptions of rare materials with high interest in and beyond academia. Exposure of these materials on the web will be key to adoption of linked data practices for describing the primary sources and distinctive collections that are increasingly core for research libraries and central to production workflows.

The work will also provide focus for metadata practitioners and leverage local knowledge of the collections to produce rich descriptions that demonstrate the utility of natively-created linked data and inform development of generalizable workflows for varied domains and formats.