OCLC RESEARCH REPORT

Creating Library Linked Data with Wikibase

Lessons Learned from Project Passage

Jean Godby, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, Holly Tomren

OCLC

Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage

Jean Godby
OCLC Research

Karen Smith-Yoshimura
OCLC Research

Bruce Washburn
OCLC Research

Kalan Knudson Davis
University of Minnesota

Karen Detling
National Library of Medicine

Christine Fernsebner Eslao
Harvard University

Steven Folsom
Cornell University

Xiaoli Li
University of California, Davis

Marc McGee
Harvard University

Karen Miller
Northwestern University

Honor Moody
Harvard University

Craig Thomas
Harvard University

Holly Tomren
Temple University

OCLC®

August 2019

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-135-4
DOI: 10.25333/faq3-ax08
OCLC Control Number: 1110105996

ORCID iDs

Jean Godby https://orcid.org/0000-0003-0085-2134
Karen Smith-Yoshimura https://orcid.org/0000-0002-8757-2962
Bruce Washburn http://orcid.org/0000-0003-4396-7345
Kalan Knudson Davis https://orcid.org/0000-0002-1032-6042
Christine Fernsebner Eslao https://orcid.org/0000-0002-7621-916X
Steven Folsom https://orcid.org/0000-0003-3427-5769
Xiaoli Li https://orcid.org/0000-0001-5362-2151
Marc McGee https://orcid.org/0000-0001-5757-1494
Karen Miller https://orcid.org/0000-0002-9597-2376
Craig Thomas https://orcid.org/0000-0002-4027-7907
Holly Tomren https://orcid.org/0000-0002-6062-1138

Please direct correspondence to:
OCLC Research
oclcresearch@oclc.org

Suggested citation:
Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, and Holly Tomren. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. Dublin, OH: OCLC Research. https://doi.org/10.25333/faq3-ax08.

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

The OCLC Research linked data Wikibase prototype ("Project Passage") provided a sandbox in which librarians from 16 US institutions could experiment with creating linked data to describe resources—without requiring knowledge of the technical machinery of linked data. This report provides an overview of the context in which the prototype was developed, how the Wikibase platform was adapted for use by librarians, and eight use cases where pilot participants (co-authors of this report) describe their experience of creating metadata for resources in various formats and languages using the Wikibase editing interface. During the ten months of the pilot, the participants gained insight in both the potential of linked data in library cataloging workflows and the gaps that must be addressed before machine-readable semantic data can be fully adopted. Among the lessons learned:

- The building blocks of Wikibase can be used to create structured data with a precision that exceeds current library standards.
- The Wikibase platform enables user-driven ontology design but raises concerns about how to manage and maintain ontologies.
- The Wikibase platform, supplemented with OCLC's enhancements and stand-alone utilities, enables librarians to see the results of their effort in a discovery interface without leaving the metadata-creation workflow.
- Robust tools are required for local data management.
- To populate knowledge graphs with library metadata, tools that facilitate the import and enhancement of data created elsewhere are recommended.
- The pilot underscored the need for interoperability between data sources, both for ingest and export.
- The traditional distinction between authority and bibliographic data disappears in a Wikibase description.

After the pilot, the co-authors of this report discussed the long-term issues raised by their Project Passage experiences and the potential impact of linked data in library resource description workflows. The following key issues were identified:

- The transition from human-readable records to knowledge graphs represents a paradigm shift. Although some current tasks and practices will still be necessary, others will become obsolete, and some new tasks will be needed.
- Crowd-sourcing has the potential for enriching the knowledge graphs created by libraries, as Wikibase offers options for soliciting, accepting, evaluating, and managing user contributions.
- Interoperability between Wikibase, Wikidata, and other linked data sources remains an open question deserving follow-up investigation.

The report concludes with findings, reflections, and areas for future research.

INTRODUCTION

For decades, libraries have created and shared bibliographic descriptions using the Machine-Readable Cataloging (MARC) format. Hundreds of millions of MARC records are now available in online catalogs that show what libraries hold and make accessible to their patrons. MARC records also power key library activities such as resource sharing and collection development. But, because the MARC standards were developed about 50 years ago, MARC records can be understood only by library systems and are generally incomprehensible to data consumers elsewhere.

Since about 2008, standards experts in the library community have looked to linked data as potentially transformative. The linked data conventions proposed by Tim Berners-Lee in 2006[1] are envisioned as the innovation that will replace record-based descriptions encoded in MARC and other library-community standards that are consumed primarily by human readers. Instead, descriptions will make use of information in knowledge graphs featuring descriptions of the persons, places, or things of interest—commonly known as "entities"—that are connected to other entities via machine-understandable relationships. When fully available and activated by robust applications, knowledge graphs can replace the data silos that only libraries can use.

Among the proponents of this vision is the Library of Congress, whose Bibliographic Framework Initiative (BIBFRAME)[2] is designed to replace MARC with linked data principles that make bibliographic data more useful both within and outside the library community. In addition, the Program for Cooperative Cataloging (PCC) has defined six strategic directions for 2018-2021,[3] three of which are focused on linked data:

- Apply the library community's understanding of linked data by providing opportunities for practical, hands-on experimentation.
- Accelerate the movement toward ubiquitous identifier creation and identity management at the network level by increasing the numbers and proportions of entities receiving persistent identifiers.
- Expand the use of linked data value vocabularies to augment, and, where feasible, replace existing metadata practices.

During the transition to linked data, libraries are stretching to manage an increasing number of resources with reduced staffing devoted to metadata creation. In this context, linked data is viewed as both a solution and a problem. Does linking require less human effort than record creation? Or does the opening of library practices to a wider variety of resources create value-added linking opportunities that require more human effort? Has the effort already spent on linked data produced enough high-quality datasets? After all, critical mass is required for users to discover library holdings using search engines and experience what it means for libraries to be more fully integrated into the internet.

Although the number of linked data sources published by the library community has increased over the years—which includes the catalogs of national libraries, and identifier hubs such as the Library of Congress' id.loc.gov and OCLC's Virtual International Authority File (VIAF)—linked data implementations require major investments in resource-description standards, systems environments, and workflows that are beyond what most organizations can afford. This problem plagues most library linked data efforts. To muster the required investments, advocates must demonstrate a benefit. But benefits cannot be demonstrated without implementing successful linked data services.

BARRIERS TO ADOPTION OF LINKED DATA IN THE LIBRARY RESOURCE-DESCRIPTION WORKFLOW

Surveys of linked data implementors conducted at OCLC since 2015 have revealed that most linked data experimentation proceeds on a separate track from current descriptive practice, without much cross-fertilization.[4] Only the most recent results have identified the first offerings from linked data service providers.[5] Moreover, the benefits of linked data are not yet visible on library community data because supporting applications have not achieved enough technical maturity. Thus, metadata librarians report that linked data editing tools are difficult to use and require too much knowledge of the technical underpinnings of linked data. Many of these themes are mentioned in the response by Oslo Public Library to the 2018 survey:

> As far as I can see, Oslo public library is still the first and only library with its production catalogue and original cataloguing workflows done directly with linked data. As well as rolling out linked data, we rolled out an entirely new system concept. Some of what we were doing had not been done before, so we had to invent ways of doing things, some of the technologies were new enough to be undocumented and libraries immature, and the learning curve for developers, UX, and everyone has been steep.[6]

Barriers to the adoption of linked data are conceptual as well as technical. For example, the Online Dictionary for Library and Information Science defines the traditional role of the cataloger as preparing bibliographic records representing items acquired by the library, including bibliographic description, subject analysis, and classification.[7] Since the adoption of cultural heritage resource description theory, cataloging has become a process of making observations on an expanded set of resources and making assertions between entities contained within the bibliographic universe, which has recently been expanded to include digital and born-digital materials and research data sets.

Resource description in a linked data environment requires substantial training and education of catalogers, but is not dissimilar to the current practice of analyzing an item in order to record structured access points for persons, corporate bodies, subjects, events, and other bibliographic entities. Murray and Tillett frame this analysis as an attempt to figure out what is "out there" in the bibliographic universe.[8] But what has been missing until now are the processes and the ecosystem required for creating and managing native linked data.

The pilot study described in this report focuses on the conceptual issues raised by the adoption of linked data in the library resource description workflow. The pilot is grounded in the metadata librarian's existential questions: How do I create a linked data representation for the resource I am looking at right now? How does this process compare with the method I may have already used to describe the resource? And what can I do with the results? Echoing the PCC's strategic initiative, these questions are explored in a software ecosystem that permits practical, hands-on

experimentation. But just as important, this environment also promises to expose more of the benefits of linked data than is possible with the current generation of prototypes developed in the library community.

PROJECT PASSAGE: THE OCLC RESEARCH LINKED DATA WIKIBASE PROTOTYPE

Rather than creating a suite of linked data applications from scratch, the OCLC Passage staff were inspired by the Wikimedia Movement,[9] which has produced the so-called Wiki constellation of open-source software packages and the familiar applications that are built from them. These Wiki resources include[10]:

- **Wikipedia**: a free, multilingual, web-based encyclopedia
- **MediaWiki**: a free and open-source wiki software package
- **Wikidata:** a collaboratively edited structured dataset used by Wikimedia sister projects and others
- **Wikibase:** a MediaWiki extension for storing and managing structured data

Like Wikidata, Passage is built on MediaWiki and Wikibase. At the beginning of the pilot, the OCLC project team considered editing directly in Wikidata, but opted for using a new and separate instance of Wikibase for managing the project data. This decision was motivated by the following:

- interests in evaluating the software, its customization and configuration options, and its scalability
- expectations that the pilot would be experimenting with new property entities associated with bibliographic entities, which might not be relevant to, or accepted for inclusion in, Wikidata within the timeframe of the project
- concerns about managing privacy for the participants' data, which was experimental in nature and not necessarily intended for publication
- an awareness that notability[11] requirements for Wikidata may differ from those for library-community resources

> The Wikibase environment features an openness that contrasts markedly with the user interfaces developed for the library community by the commercial sector.

The Wikibase implementation is attractive because it reflects a sophisticated understanding of what is required to support crowd-sourcing across a global community of users, which is being evaluated by OCLC and many libraries as a model for modernizing the practice of cooperative cataloging and authority management. The Wikibase platform supports input from both humans and machines. For example, change-history logs record both human and machine edits, bringing a transparency that supports user engagement. Discussion pages enable human editors to share opinions about how best to create or modify a description. In addition, the software and data model are both malleable, a feature that was especially valuable for the Passage pilot. Although the Wikidata ontology was

ingested into Passage, some pre-defined properties were later modified to meet the more precise descriptive needs of metadata librarians. And new properties were added after they were proposed by the participants and reviewed. In all of these respects, the Wikibase environment features an openness that contrasts markedly with the user interfaces developed for the library community by the commercial sector.

Wikibase also offers many technical advantages as a platform for experimentation with library resource-description workflows, including:

- The platform is a mature end-to-end system, extending from data ingest and editing to data serialization in RDF, or the Resource Description Framework.
- The platform is natively multilingual. It embeds Unicode, allowing users to create and search in more non-Latin scripts than are available in most library systems.
- A linked data database, or "triplestore," is synchronized with the structured data in Wikibase. This configuration supports SPARQL, a semantic web query language operating on RDF, as well as APIs, or application programming interfaces, that enable the incorporation of third-party additions.
- The dataset available from Wikidata has definitions for hundreds of thousands of entities, or linked data real-world objects, in varying levels of completeness, to which library-community and other identifiers have already been applied.
- Wikidata features a community-developed and managed ontology that can be reused and adapted in any Wikibase instance.

Most important, this technical detail is mostly hidden from human users of the Wikibase applications. Thus, metadata librarians familiar with current workflows can easily interact with the editing interface to create resource descriptions in a new idiom. In other words, the Wikibase platform offers the promise that the transformation from human-readable records to machine-understandable knowledge graphs can occur in the library metadata creation workflow without presupposing any knowledge of RDF, Turtle, triples, or other details in a linked data implementation.

Many of these arguments were also discussed in a white paper published in April 2019 by the Association of Research Libraries, which mentions the OCLC pilot in its environmental scan. According to the report:

> Wikibase provides something that a number of other platforms in the linked data environment have not: a dynamic, machine- and human-contributable and readable environment that supports diversity of language and data structure from the beginning. The growing number of Wikibase implementations...suggests opportunities for scholarly and GLAM [Galleries, Libraries, Archives and Museums] reuse of the software as a generic data store. Investment in community infrastructure means less lock-in to proprietary platforms and systems.[12]

THE TERMS OF THE PILOT

The Passage pilot provided a sandbox for a qualitative, hands-on, and interactive study of resource description by metadata librarians. In late 2017, OCLC identified an opportunity to extend the conversation it had begun with libraries around entity descriptions. The 2015 Person Entity Lookup Pilot,[13] which prototyped the use of search and display tools, left the participants and OCLC eager to explore the possibilities of enhancing entity descriptions, and creating new ones, in an environment that was linked data native.

The initial call for participation made it clear that the Passage pilot would require a high degree of engagement and experimentation. To support this goal, OCLC set up a project structure in which participants and a cross-divisional OCLC team comprising OCLC Research staff, product managers, and production engineers could address open questions and obstacles, as well as discuss activities that participants completed on their own between live sessions. Rich interaction that led directly to the creation of this report occurred in three channels:

1. monthly progress updates, all of which concluded with a list of enhancements recommended by the participants that would be implemented in the following month
2. weekly virtual "office hours" that explored focused topics, or use cases, in discussions that were frequently led by the participants
3. a space in the OCLC Community Center,[14] where participants could ask and answer questions, discuss problems and homework assignments, share documents, and interact with OCLC staff

The primary goal was to confront the challenges of linked data in current workflows, which would contribute to strategic discussion of post-MARC metadata standards and inform OCLC's Global Product Management roadmap for future metadata applications and services. The pilot ended in September 2018, and live access to OCLC's Wikibase instance and related applications is no longer available. The data is preserved through screenshots and links to the Internet Archive, which are cited throughout this report.[15]

To create an environment optimized for experimentation with library-community priorities, OCLC's local instance of the Wikibase/MediaWiki software package operated on a dataset imported from Wikidata containing entities associated with the library-community identifiers from sources such as VIAF, FAST, id.loc.gov, and WorldCat. At the start of the pilot, the OCLC team acted primarily as tutors. They oriented pilot participants to the features and functionality of the Wikibase platform and demonstrated its use in sample resource descriptions. They assigned homework and offered recommendations for best practices. As the pilot progressed, the participants took the lead describing their own resources, offering feedback on what did or didn't work, and suggesting improvements. OCLC staff used this input to make modifications and enhancements to their copy of the Wikibase software, for which another round of feedback was solicited.

THE PILOT STUDY PARTICIPANTS

Representatives from 16 academic, research, public, and national libraries participated at the height of the Passage pilot:

- American University
- Brigham Young University
- Cleveland Public Library
- Cornell University Library
- Harvard University
- Michigan State University
- National Library of Medicine
- North Carolina State University
- Northwestern University

- Princeton University
- Smithsonian Library
- Temple University
- University of California, Davis Library
- University of Minnesota
- University of New Hampshire
- Yale University

The participants had a variety of motivations to participate in this pilot, which included:

- gain hands-on cataloging experience in a linked data–friendly environment, understand how this work mirrors or diverges from current authority and cataloging workflows, and share experiences with other participants
- understand how linked data relationships between entities might affect search, retrieval, and discovery
- learn about working with Wikidata properties in a Wikibase environment
- envision how linked data could be integrated with other components of existing metadata environments
- apply lessons learned from participation in this prototype to local linked data projects
- understand and potentially influence OCLC directions for linked data use and creation

In addition, Passage focused on new description, not the conversion of existing bibliographic or authority records to linked data, which was the focus of many library-community projects. Finally, Passage represented an open-ended experiment, not a project with production goals and anticipated deliverables. This freedom to explore implications and ask questions identified many gaps to address in follow-up research and development. But this freedom also allowed the participants to discover for themselves that machine-understandable semantic data is a useful outcome of the library metadata creation workflow and can be easily produced.

Many of the Passage participants had participated in other library linked data projects, including the BIBFRAME initiative, the Andrew W. Mellon Foundation–funded Linked Data for Production[16] and Linked Data for Production: Pathway to Implementation projects,[17] various PCC task groups, as well as projects conducted at their own institutions. At the beginning and end of the Passage pilot, the participants were asked to reflect on how the OCLC project compared with their previous experiences. They were drawn to participate because Passage was built on Wikibase, the software foundation for Wikidata. As the Passage pilot progressed, Wikidata was growing in importance to the library community as measured by the number of relevant presentations at professional conferences in 2017 and 2018. A surge of linked data projects that consume Wikidata was also detected in the 2018 International Survey of Linked Data for Implementers.[18]

Nevertheless, these developments represent evidence of interest more than hands-on editing experience by members of the library community. Live statistics accessible from Wikiscan[19] show that, as of April 2019, 66% of Wikidata edits had been contributed by automated processes, or bots,[20] not human editors. In addition, bots had created 89% of the new Wikidata pages, or "item entities," as defined in the section "A Wikibase article, page, or 'entity'" on page 15. Thus, the Passage pilot is also a usability test of the Wikibase editing interface.

HOW THIS REPORT IS ORGANIZED

The remaining sections tell a simple story. First, they describe the system that was deployed in the Passage Pilot. They show how the system was used as the environment to conduct resource-description experiments that produce linked data instead of human-readable bibliographic and authority records. And they discuss the most important lessons that were learned from the experience.

The following section is an overview of the Wikibase/MediaWiki functionality that powers Wikidata, starting with the features revealed in the user's experience of the Wikidata editing interface. The discussion broadens to a review of the high-level system architecture components, which provides a technical context that pinpoints the places where OCLC made the most enhancements in the Passage pilot. The section ends with an outline of the three phases of the Passage pilot, which describe OCLC's modes of engagement with the library-community participants and provides a user-requirements context for the enhancements.

The section "Studies of Resource Description in Passage," starting on page 27, is the core of this report. It consists of contributions from pilot participants that describe eight use cases discussed during the Passage office hours sessions. The cases include both bibliographic and authority descriptions—or, in the linked data idiom, real-world objects representing creative works as well as the people, organizations, places, things, and events that are linked through "author," "publisher," "subject," or other named relationships to the works. Analysis and reflection are free-ranging. Discussion includes how to select an important object to describe, how to identify and correct gaps in the Wikidata ontology, and how the new tasks compare with current practice, both positive and negative.

The report concludes with important lessons learned from the use cases, reflections on the Passage pilot, and areas of future activity.

An Overview of the MediaWiki/Wikibase Platform and OCLC Project Prototypes

This section gives a brief overview of the technical details of the Passage implementation. It starts with a non-technical discussion of the user's view of the outcome of the editing process, which generates a set of mostly structured data about an object or thing of interest. Later sub-sections describe the flow of user-supplied data through a system that includes components for intake or ingest, storage, and display or export. The final sub-section describes the user-centered design of the Passage pilot and introduces OCLC's new applications.

A WIKIBASE ARTICLE, PAGE, OR "ENTITY"

The technical discussion starts with an overview of the editing process in Wikidata, which is a technical peer of the Passage prototype because both applications are built on the MediaWiki/Wikibase platform.

Figure 1 shows a typical output, a Wikidata page containing facts about the science fiction author Douglas Adams. This graphic from Wikimedia Commons identifies the most important elements of the page, most of which are encoded as structured data. Concepts such as "entity," "fingerprint," "item," "property," "rank," and "qualifier" are frequent topics of discussion in the Wikibase editing experience and are used throughout this report. These terms are defined in the Wikidata Glossary and Wikidata Help portal.[21]

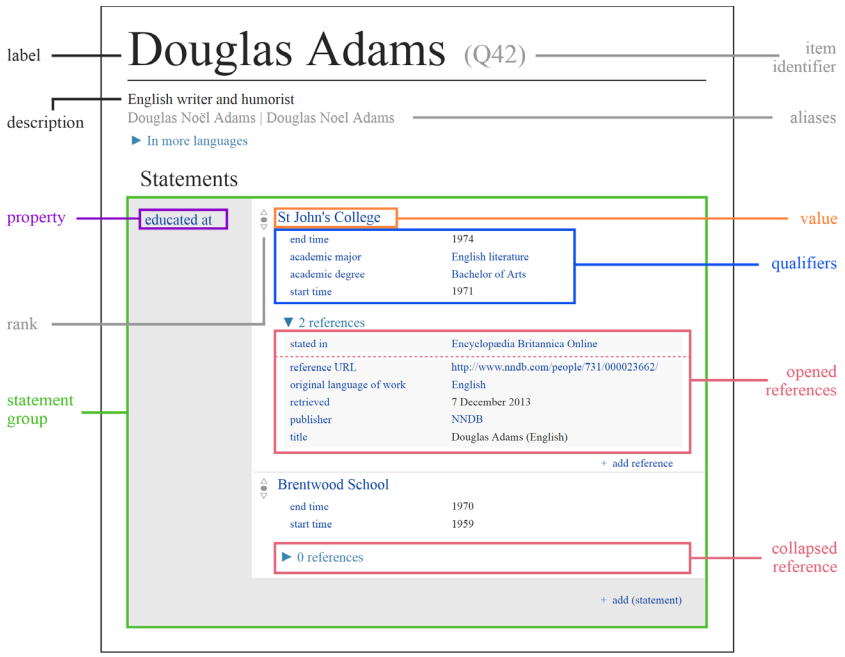


FIGURE 1. Structural components in the Wikibase editing page, Charlie Kritschmar (WMDE) [CC0], from Wikimedia Commons.[22]

Figure 1 represents an entity,[23] or the content of a Wikibase page that refers to a disambiguated Wikibase item or property, defined below. A fingerprint,[24] or "fingerprint data," appears at the top of the entity and consists of human-readable text that uniquely identifies the item or property. In figure 1, the fingerprint label[25] is "Douglas Adams." The fingerprint description[26] is "English writer and humorist," and the fingerprint aliases[27] are "Douglas Noël Adams" and "Douglas Noel Adams." Fingerprint data can be entered in any language and script supported by Wikibase, and is designed to support user-initiated search and retrieval, item selection, disambiguation, and duplicate detection.

Other important elements in the Wikibase entity template that are called out in the figure:

- **Items**[28] represent "all the things in human knowledge, including topics, concepts, and objects." Figure 1 is an item entity for the person named Douglas Adams who wrote *The Hitchhiker's Guide to the Galaxy* and other works of science fiction.
- A **property**[29] represents a relationship between the primary entity and another entity or string literal. In figure 1, Douglas Adams has the "educated at" relationship with St. John's College.
- A **qualifier**[30] can be used to replicate the nuanced meaning of a syntactically complex sentence in natural language. Figure 1 illustrates one example. A simple structured statement that can be paraphrased as "Douglas Adams was educated at St. John's College" can be elevated to "Douglas Adams was educated at St. John's College from 1971 to 1974, during which he earned the Bachelor of Arts degree" with the property qualifiers, "start time," "end time," and "academic degree." Later in the report several use cases are presented to illustrate similar applications of property qualifiers.
- A **rank**[31] is a value of "preferred," "normal," or "deprecated" assigned to a statement by a human editor or Wikibase utility. The statements shown in figure 1 have the default "normal" rank, as indicated by the black dot in the middle of the icon positioned at the left of each property object. The rank may be raised to "preferred," "as long as it is regarded

as sufficiently reliable," according to the Wikibase Data Model Primer.[32] Or the rank may be lowered to "deprecated" if it is known to be erroneous. Although ranks can be interpreted as measures of quality or trustworthiness, they are also used to acknowledge disagreement or diverse viewpoints. Statement ranks are explored in some of the use cases, and the statement rank is identified as a potentially useful measure for evaluating crowd-sourced contributions to library resource descriptions.

At a high level, the process of building an item entity is conceptually simple. Users initiate an editing session from the Wikidata home page or a local installation of the Wikibase software platform, and take the following steps:

1. Declare a new item entity.
2. Add fingerprint data.
3. Add one or more structured statements to associate properties for the item with other item entities or literal strings, modified with property qualifiers, if necessary.
4. Cite a reference for a statement, which boosts its rank.

An additional task could be performed in the Passage Wikibase instance: a human editor could modify the ontology inherited from Wikidata if a description needed a property that had not been previously defined. But in Wikidata, editorial access to the ontology is restricted.[33] If a proposed modification to the Wikidata property set gains more supporters than opponents, the change is implemented by a property creator or administrator.

The terms defined above identify the most important features of the Wikibase editing experience. But users who are familiar with linked data also recognize concepts that transcend the boundaries of the Wikibase platform. For example, a Wikibase item is commonly interpreted as equivalent to a real-world "thing," "entity," or "object," as specified by linked data conventions.[34]

The Wikibase architecture empowers users to interact with a simple and intuitive editing interface to create and consume linked data. Users can accomplish this ambitious goal without having to write software or manipulate raw RDF.

And when the user declares an item or property entity, the Wikibase editing application associates it with a persistent and globally unique URL such as https://www.wikidata.org/wiki/Q42, the item URL for Douglas Adams. This URL is commonly interpreted as a URI (or Uniform Resource Identifier),[35] as defined by linked data advocates, which is globally unique and persistently associated with the British author.

In addition, the structured data that is the typical outcome of an editing session can be exported as linked data to a triplestore,[36] a specialized database that can be queried using the powerful SPARQL

query language.[37] The Wikibase architecture empowers users to interact with a simple and intuitive editing interface to create and consume linked data. Users can accomplish this ambitious goal without having to write software or manipulate raw RDF.

For more guidance on editing in the Wikibase environment, please consult the Wikidata Help pages[38] and the tours of the editing workflow provided there.[39] In the ARL report cited earlier, the section "A Brief History and Introduction to Wikidata" has an overview of the editing process for audiences in the library community.[40] The OCLC Works in Progress webinar "Introduction to Wikidata for Librarians," conducted in June 2018, covers this material in more technical detail.[41]

THE WIKIBASE ENTITY ECOSYSTEM

An open-source package available from the Wikimedia Foundation creates the environment or platform in which pages such as the Douglas Adams entity shown above can be created, edited, displayed, and stored. Wikidata is a software application built from Wikibase components. The Passage prototype is OCLC's experimental sibling of Wikidata, which is also built from the Wikibase software, applications, and data model. In addition, the Passage installation incorporates data resources produced by a global community of Wikidata editors, such as the help portal, glossary, and ontology, as well as a subset of item entities. During the Passage pilot, local copies of data inherited from Wikidata were modified, and the Wikibase software suite was extended with the Explorer and Retriever, two new applications developed at OCLC that fill in gaps in the library resource-description workflow.

MediaWiki and Wikibase

As noted earlier, MediaWiki provides a wiki software foundation with a range of useful features, including a robust API for reading and writing data and support for managing user permissions, page editing history, article discussions, and an index for unstructured text documents. The Wikibase extension to MediaWiki also provides a way to store structured data as linked data entities and their relationships. As data is created, modified, and deleted, the Text Search Index is automatically updated to support search and retrieval, duplicate detection, and disambiguation. Finally, a Wikibase editing interface is available for creating, updating, merging, and deleting item and property entities. The connections between these systems are depicted in figure 2.

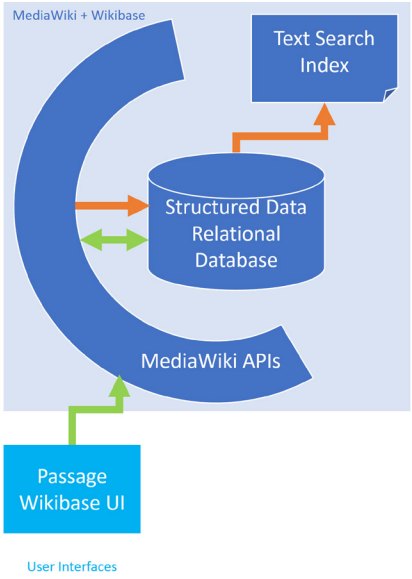


FIGURE 2. MediaWiki and Wikibase environment.

The Triplestore, SPARQL Query API, and the SPARQL Query user interface

An RDF triplestore and SPARQL Query API complement the MediaWiki/Wikibase platform. Data in the triplestore is automatically synchronized at regular intervals with data from Wikibase. The user interface for the Wikidata Query Service, relabeled as the Passage Query Service, provides an easy way to construct and send SPARQL Queries to the API and its underlying triplestore. The SPARQL Query interface has built-in sample queries, data visualizations for timelines, images, and more. The connections between the interface and its underlying systems are shown in figure 3.

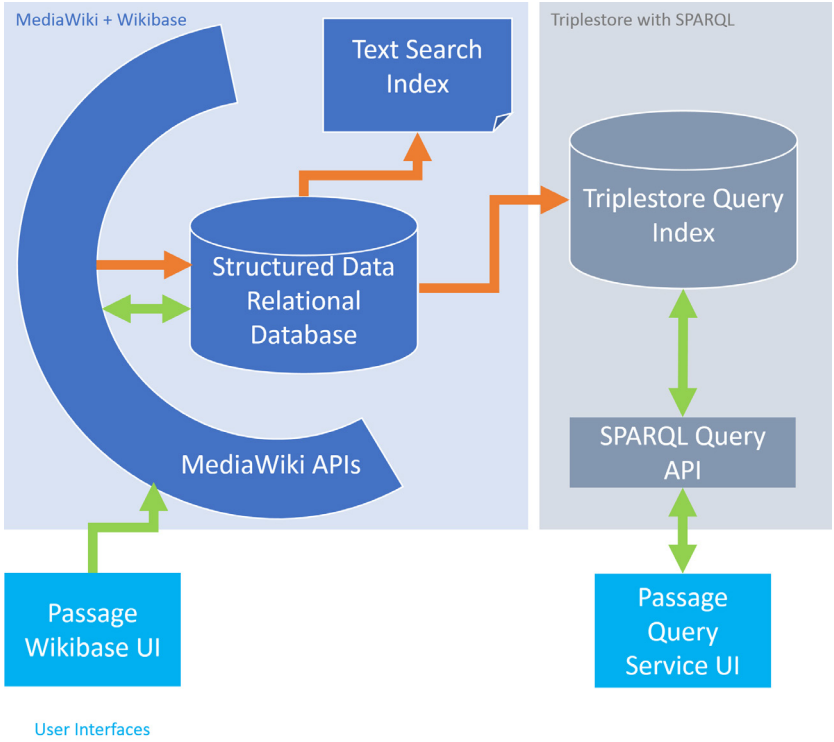


FIGURE 3. MediaWiki and Wikibase environment with a parallel Triplestore, SPARQL Query API, and SPARQL Query user interface.

Applications that use the MediaWiki/Wikibase environment and the SPARQL Query service: OpenRefine and the Passage Explorer

OCLC prototyped an alternative interface for exploratory searching, the Passage Explorer, described in more detail below. It uses both the MediaWiki Search APIs and the SPARQL Query API, along with external resources like DBPedia, to show more context for individual entities.

OpenRefine[42] is a popular tool in the library community for cleaning and normalizing data and reconciling local data with external resources.[43] The OpenRefine API utilizes both the MediaWiki APIs and SPARQL Query API and can be used by the OpenRefine application and other compatible client applications.

These additional user interfaces and their connections are depicted in figure 4.

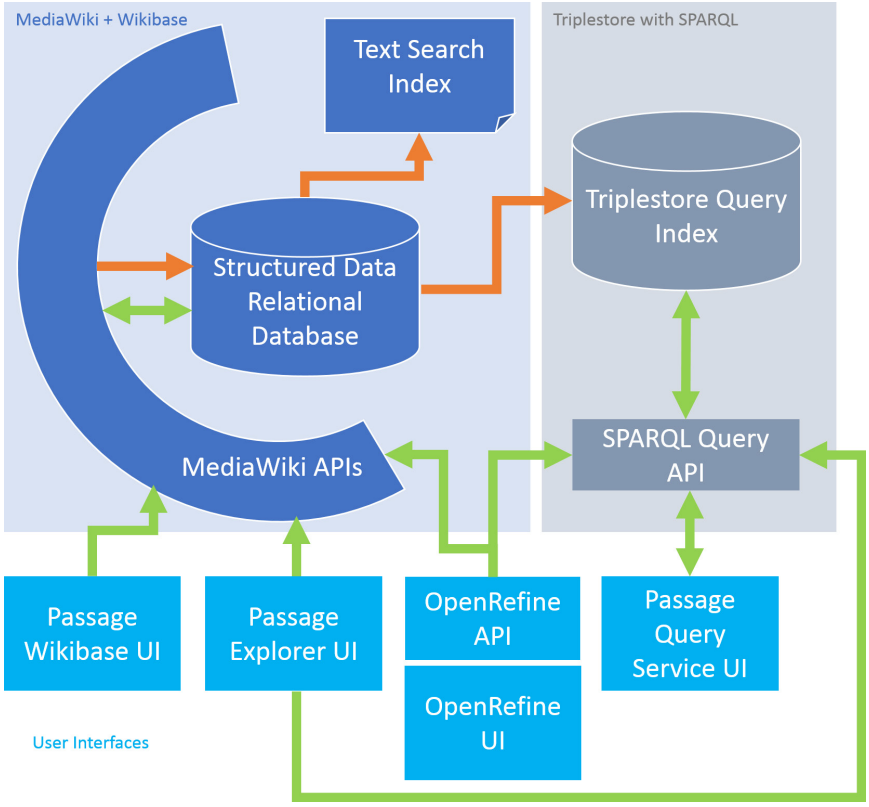


FIGURE 4. Applications that use the MediaWiki/Wikibase environment and the SPARQL Query service: the Project Passage Explorer, and OpenRefine.

Data import with PyWikibot and the Passage Retriever

Batch data import is managed by preparing entity data in upstream processes for loading. OCLC used a software library for the Python programming language named Pywikibot,[44] which calls the MediaWiki authentication and writes APIs as the mechanism for batch data import.

To help Passage users augment the data, OCLC prototyped a new web application called the Retriever to search for entity data in Wikidata, VIAF, and FAST, and import that data into Passage (via a Pywikibot script), as shown in figure 5.[45]

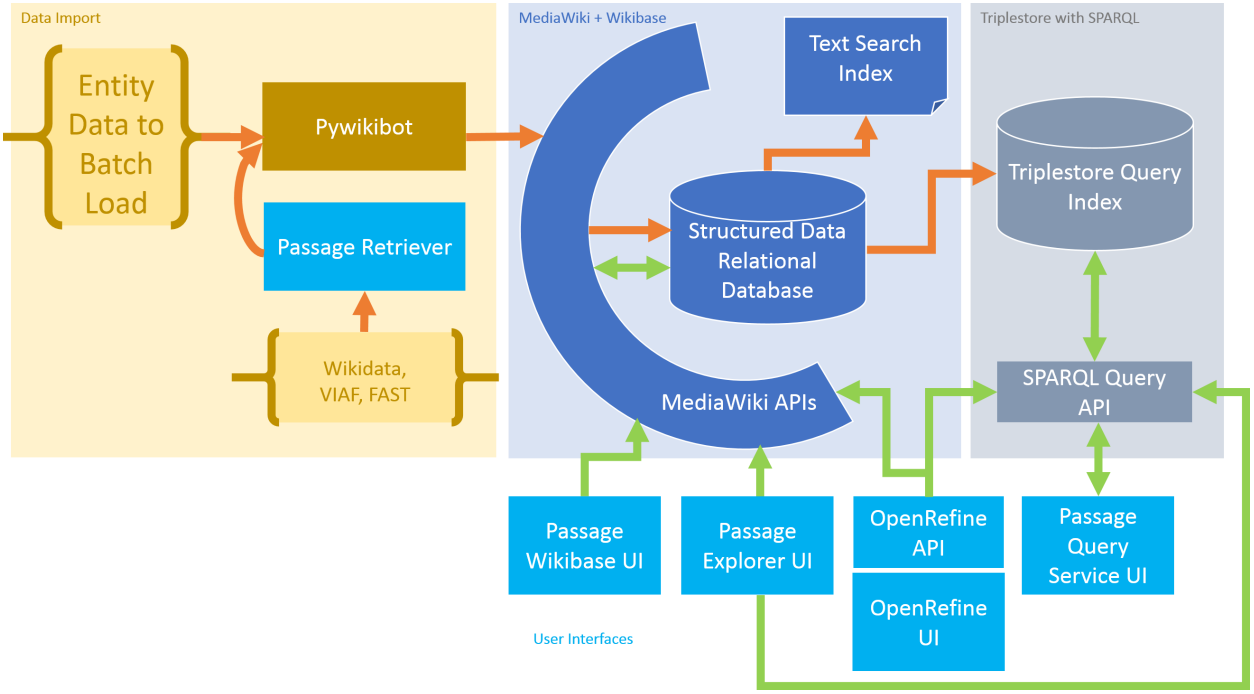


FIGURE 5. Data import with Pywikibot and the Passage Retriever.

In short, figure 5 shows a high-level view of the entity ecosystem that was implemented in the Passage pilot. The MediaWiki/Wikibase software foundation, supplemented with other openly available tools like the SPARQL Query system and Pywikibot for data loading, provides a powerful and flexible platform from which to quickly launch a linked data management system, and to which extensions and new fit-for-purpose applications can be easily added.

THE BASELINE DATASET

OCLC performed two kinds of analysis to prepare the data for initial ingest into Passage.

First, procedures were developed for filtering, cleaning, normalizing, and reconciling source data. These are typical tasks in the library resource description workflow, which decompose record-based text to machine-understandable collections of statements about real-world objects—or promoting strings to things, in the linked data idiom.

In addition, the overlap between Wikidata and library-community resource descriptions was analyzed to identify candidates for seeding the initial Passage database. In operational terms, the process selected 1.2 million Wikidata items that contain identifiers from library-community linked data resources. The seeding process initially extracted Wikidata items containing VIAF identifiers, but later iterations also included identifiers from FAST, ID.loc.gov, the Getty vocabularies, and WorldCat.

This exercise also raised modeling issues, especially with creative works. For example, WorldCat record identifiers closely correspond to editions or FRBR manifestations, but many Wikidata assertions about creative works seem to apply to more abstract work-like objects instead. This was the subject of lively debate during the Passage pilot and will be discussed in more detail in the next section of this report.

For less contentious resources, the association of identifiers from multiple resources had two beneficial effects. First, many properties defined in multiple vocabularies could be aligned through RDF "same as" relationships, which were collected into a union list, or codex. But properties that did not match could sometimes be interpreted as enrichments to library community models, another benefit. For example, the model of "person" in the MARC 21 Format for Authority Data[46] conforms to a thesaurus-like structure that highlights established headings, as well as alternate headings and cross-references. But when a personal name authority record is matched with a Wikidata entity, the result is an association of authorized headings with biographical details such as a person's gender, photograph, country of citizenship, and names of family members. This enhanced description is revealed to end users in the Passage Explorer discovery interface.

THREE PHASES OF USER-CENTERED DEVELOPMENT AND TESTING

A concurrent objective in the development of the Passage pilot software was the hands-on involvement of metadata librarians from the 16 institutions listed in the Introduction. The high-level objectives were to learn about the Wikibase editing environment, interpret the task of item creation in the context of library resource-description practice, and test the hypothesis that the objectives and outcomes of the new workflow address unmet needs of library catalogers. Along the way, OCLC's engineers developed tools, made improvements to the Passage copy of the Wikidata ontology and software libraries, and identified gaps to be addressed in future work. They also acted as teachers, conducting tutorials and facilitating hands-on practice sessions. This work happened in three phases.

Phase 1: Introducing Wikibase, evaluating the initial dataset, promoting SPARQL queries

In Phase 1, OCLC staff introduced the baseline Passage prototype to a small group of pilot participants from the University of California at Davis, Cornell University, and Montana State University, all of whom have expertise in library cataloging, authority control, or linked data. The initial database contained 1.2 million Wikibase items imported from Wikidata that were especially rich in library-community identifiers, which were reconciled and indexed by an offline process at OCLC. These preparatory steps ensured that all references to the English novelist Jane Austen, for example, resolved to the same Wikbase item, and that all name variants associated with library-community URIs would be searchable and visible in the browse index.

OCLC's engineers argued that a locally installed instance of the software that powers Wikidata creates an appropriate "sandbox" for hands-on experimentation with linked data and would support the library metadata-creation workflow. After this initial discussion, the main goal of Phase 1 was pedagogical. The group worked together to master the Wikidata framework and become power users. OCLC staff created engagement activities such as treasure hunts and exercises which featured resources that had already been described in installed production workflows. These exercises generated much discussion of similarities and differences in the current and new forms of description.

Early in the discussion, it became clear that the Wikidata editing interface was confusing because it does not reveal the full extent of the knowledge that is created when entities are linked through named relationships. The OCLC team demonstrated this potential by issuing queries through the Project Passage SPARQL UI, which (like the corresponding Wikidata Query Service) accesses the Wikibase SPARQL query API.[47] This tool is more powerful than keyword searching, and the examples the OCLC team created gave participants unfamiliar with SPARQL a starting place for creating and storing their own searches.

The results of four SPARQL queries issued on the Passage dataset are shown in table 1. As indicated in the table, queries that satisfy a knowledge-discovery need reveal that both Jorge Luis Borges and Marcel Proust were librarians, as well as other facts that might not be widely known. Query results may also be submitted to Wikibase visualization tools, such as the translation history timeline that was assembled from Passage items related to *Sein und Zeit*.

TABLE 1. SPARQL queries on the Passage dataset and some possible applications

Purpose	Query	Sample Results
Knowledge discovery	Books authored by librarians[48]	• *Book of Imaginary* Beings by Jorge Luis Borges • *In Search of Lost Time* by Marcel Proust
Knowledge discovery	Authors with the most translations[49]	• Leo Tolstoy • Roald Dahl • Marguerite Duras • Ed Greenwood • Stanislaw Lem
Visualization	Translation history timeline of *Sein und Zeit* by Martin Heidegger[50]	Discussed in the section on use case on works, translations, and Chinese-language descriptions (see p. 29)
Error detection	People born between 1800 and 1880 without a specified death date[51]	• Louis Jean Désiré Delaistre • Félix Achille Saint-Aulaire • Michel-Placide Justin • Thomas Smallwood • Jean-François-Barthélémy Cokken

Phase 2: Implementing requested features, testing constraints, and promoting the Explorer

In this phase, technical development included the following enhancements:

- More responsive search: For example, "Douglas Adams" is an expected query to Wikidata in the native Wikibase editing interface. The Passage version of the interface also accommodates searches on identifiers, such as the Wikibase-generated ID for Douglas Adams, Q42.
- Refinements to statements and claims: The Wikibase editing interface promotes statements to claims with the addition of a source of evidence. Given the library community's core values of authority and trustworthiness, this option was promoted to a requirement.
- Discovery of constraint violations: To encourage other quality standards, the Wikibase item- and property-based constraints were also extended. Once in place, Wikibase gadgets and SPARQL queries can be used to look for constraint violations as a quality-assurance measure. For example, the "author" property could have a value type constraint that flags as a violation any item that is not "person," "pseudonym," or "organization."
- Improved batch-ingest facility: To facilitate uploading data from Passage participants' local files, an API was created to provide a linked data reconciliation endpoint compatible with OpenRefine,[52] the widely used tool for cleaning, normalizing, and assigning identifiers to string data.

Engagement activities featured the description of library resources that are difficult to encode for algorithmic processing in current library metadata standards. For example, OCLC staff explored the representation of relationships among published books and their translations that facilitate discovery and retrieval. The first outcome was a set of Wikibase items for Martin Heidegger's book *Sein und Zeit,* a work of philosophy published in 1927 and translated into dozens of languages, including English. This case motivated the development of the Passage Explorer, which supports discovery, both for end users and for practitioners in the metadata creation workflow, by revealing relationships that are not explicitly exposed through the editing interface.

This feature provides catalogers with instant feedback about how their work affects the discovery experience, both for those engaged in the metadata creation workflow and for library patrons. As a result, the authors of every use case claim that the Explorer is an indispensable tool in the metadata creation workflow.

The Explorer view of Douglas Adams is shown in figure 6. The display is populated with results from live queries to three sources:

1. The RDF triplestore synchronized with the relational database, which is accessed via SPARQL. The list of publications shown in figure 6 is assembled from a SPARQL query across the entire Passage dataset.
2. Third-party sources such as Wikimedia Commons, where the photograph is from, and DBpedia where the biographical sketch is extracted from.
3. The Wikibase relational database, which can display changes that have just been entered through the editing interface. This dynamic feature of the Explorer cannot be illustrated in a static image, but is one of the most consequential technical innovations in the Passage pilot. This feature provides catalogers with instant feedback about how their work affects the discovery experience, both for those engaged in the metadata creation workflow and for library patrons. As a result, the authors of every use case described in the next section of this report claim that the Explorer is an indispensable tool in the metadata creation workflow.

Project Passage Explorer Item Q526648

Search for items in project passage

Douglas Adams

About

CC BY-SA 2.0

Wikimedia commons image

Douglas Noel Adams (11 March 1952 – 11 May 2001) was an English author, scriptwriter, essayist, humorist, satirist and dramatist. Adams is best known as the author of The Hitchhiker's Guide to the Galaxy, which originated in 1978 as a BBC radio comedy before developing into a "trilogy" of five books that sold more than 15 million copies in his lifetime and generated a television series, several stage plays, comics, a computer game, and in 2005 a feature film. Adams's contribution to UK radio is commemorated in The Radio Academy's Hall of Fame. Adams also wrote Dirk Gently's Holistic Detective Agency (1987) and The Long Dark Tea-Time of the Soul (1988), and co-wrote The Meaning of Liff (1983), The Deeper Meaning of Liff (1990), Last Chance to See (1990), and three stories for the television series Doctor Who; he also served as script editor for the show's seventeenth season in 1979. A posthumous collection of his works, including an unfinished novel, was published as The Salmon of Doubt in 2002. Adams was known as an advocate for environmentalism and conservation, as a lover of fast cars, cameras, technological innovation and the Apple Macintosh, and as a "devout atheist".

From dbpedia

English writer and humorist

Also known as: DNA | Douglas N. Adams | Douglas Noel Adams | Douglas Noël Adams | Adams, Douglas

Author of

Last Chance to See
BBC radio documentary series

Mostly Harmless
1992 book by Douglas Adams

The Hitchhiker's Guide to the Galaxy (English)
1979-1992 series of five books by Douglas Adams

The Restaurant at the End of the Universe
1980 book by Douglas Adams

Life, the Universe and Everything (English)
1982 book by Douglas Adams

The Hitch Hiker's Guide to the Galaxy (1979 edition)
edition of book

Dirk Gently's Holistic Detective Agency
1987 book by Douglas Adams

The Salmon of Doubt
2002 book by Douglas Adams

The Long Dark Tea-Time of the Soul
1988 Book by Douglas Adams

Douglas Adams's Guide to The Hitchhiker's Guide to the Galaxy (English)
guide by Douglas Adams to The Hitchhiker's Guide to the Galaxy

FIGURE 6. The Wikibase entity for Douglas Adams, as represented in the Passage Explorer interface. View a larger image online.

Phase 3: Enlarging the set of participants, developing documentation, holding office hours, and testing the Retriever

In the final phase of the pilot, the number of participating libraries expanded from three to 16, prompting the OCLC team to formalize the process of orienting new members and educating them about what had been accomplished so far in the pilot. One result was an increased reliance on the List of Properties,[53] as well as new entries in the installed Wikidata help portal, which was continuously enhanced with tutorials, primers, and other documentation relevant to the library resource-description tasks in Passage. These entries addressed topics such as Guidance for Creating and Editing; the Data Model; Using Identifiers in Creative Work Items; Adding Labels, Descriptions, and Aliases in Multiple Languages; Adding Dates in Passage; Creative Works and Translations; Events; The Retriever; and the OpenRefine Reconciliation API.[54]

The primary engagement activity in Phase 3 were the virtual office hours hosted by OCLC, which were held weekly from June through September 2018. Each session featured a walk through of a resource-description task, often led by a pilot participant. Technical development addressed issues that increased the ease, accuracy, and efficiency of executing these tasks. Suggested technical improvements included enrichments to the template for creating the Wikibase fingerprint labels and descriptions, the addition of WorldCat as a data source for the Explorer, and additions to the ontology imported from Wikidata.

During Phase 3, OCLC also developed the Passage Retriever, a web application designed to speed up the process of creating Passage item entities using data from external sources such as Wikidata, VIAF, FAST, and local files. With the Retriever, it is possible to search for things that are described in those sources and select relevant matches, a task that is onerous in current workflows. The Retriever automatically converts the external data into the Wikibase template and readies it for loading into a new Passage item by performing a few preprocessing tasks. For example, the Retriever flags potential duplicate entities by matching external identifiers to items already cataloged in Passage. The Retriever helps users edit fingerprint labels, descriptions, and aliases, and to identify and delete erroneous statements or adjust the ranks assigned to individual statements prior to creating a new entity in Passage. These Retriever workflow steps are illustrated in figure 7.

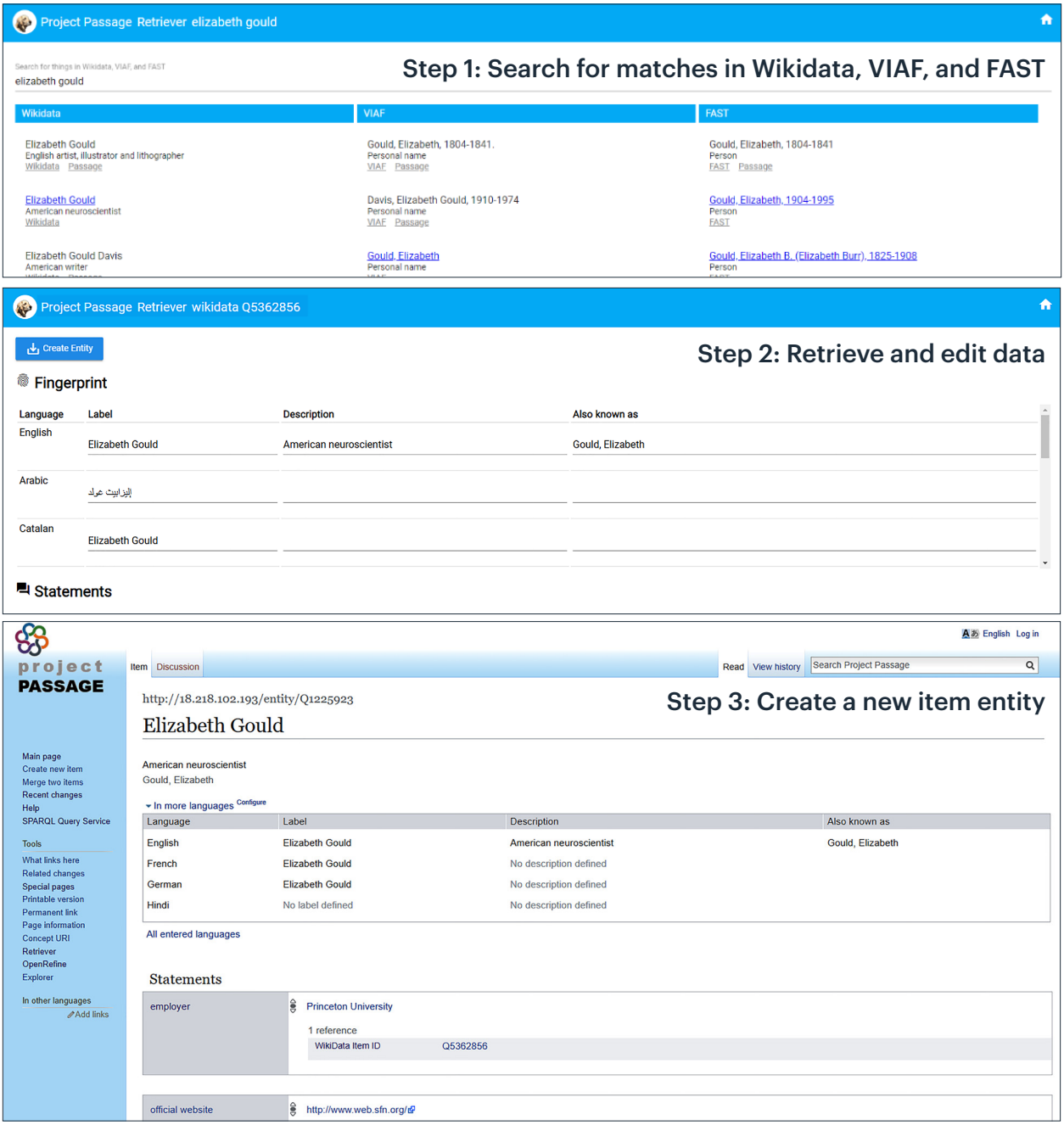


FIGURE 7. Project Passage Retriever workflow: Searching for matches, retrieving and editing data, and creating a new entity. Follow the links for larger views of the interface.

A ROBUST SOFTWARE FOUNDATION

In sum, the Passage implementation demonstrated that the Wikibase and MediaWiki software libraries are a robust foundation for tools that facilitate the ingest, storage, indexing, and retrieval of semantic data, which is natively organized not as records, but as the real-world objects defined by linked data conventions. The Wikibase applications and APIs also support mature intake environments for the human editing workflow, as well as the automated batch of data encoded in legacy formats. The applications are also designed with an abstraction layer that permits subject matter experts to focus on data curation and not the technical machinery of linked data. These features introduce significant innovation to the task of library resource description and make the generation of linked data accessible to many more practitioners.

As a result, the Passage engineering effort could be focused on the outer edges of the Wikibase process flow, where the unique requirements of library-community use cases are most likely to be addressed. And, where new applications were needed to help participants adapt to this new cataloging workflow, the software foundation proved to be a flexible and adaptable basis on which to prototype new applications, as OCLC did with the Passage Explorer and the Passage Retriever.

Studies of Resource Description in Passage

USE CASES IN LIBRARY METADATA CREATION USING THE WIKIBASE EDITING INTERFACE

The Passage editing workflow was enhanced with two utilities developed at OCLC during the course of the pilot: the Explorer, a discovery interface operating on the Passage dataset; and the Retriever, which identifies relevant descriptions created elsewhere and imports them into Wikibase for further refinement. Both were developed in response to requests by the pilot participants for tools that eased the task of description and showed the effect of work in progress.

The use cases described here were provided by the authors of the report. They are excerpted from a corpus of 20 hours of live interaction with the full set of participating institutions in virtual meetings conducted during the pilot. These recordings are supplemented with written commentary in the OCLC Community Center and informal discussion at each participating institution that was reported in public presentations referenced throughout this report.

Some of the participants were familiar with linked data and the Wikimedia ecosystem and had editing experience with Wikidata. But the Passage pilot compelled a deeper encounter for most participants. They were challenged to map their knowledge of current standards and workflows to the immediate task of describing the resource in front of them—in a novel format, and with models and applications inspired by different assumptions and priorities. This experience yielded many insights about the current workflow and a potential new one, all of which are discussed at a high level in the section Lessons Learned and Reflections on page 57.

The participants' familiarity with linked data also introduced a terminological problem at the beginning of the pilot because the Wikidata glossary[55] proposes new definitions for some terms that are already difficult and overloaded. For example, the Wikibase *item* corresponds to a linked data entity, or real-world object; and the Wikibase *entity* is commonly interpreted as an RDF-encoded web page that is "about" a real-world object, which is resolvable from a persistent and globally unique URI. This report uses the Wikidata definitions, unless otherwise noted.

The use cases were not scripted, or even planned. Instead, pilot participants volunteered to lead discussions of self-selected problems in the office hour sessions. OCLC staff facilitated the sessions with tutorials and help pages that recommended properties for each class of resource that participants elected to describe.[56] OCLC's Passage engineers also offered guidance on where to start—with what this report calls the Wikibase "item of interest"—i.e., the person, place, organization, thing, event, or creative work that motivates the act of description. Once identified, the item of interest is described using the editing procedure outlined earlier. Through the editing interface, the user declares an item, and the system responds by generating an HTML form for an item entity and associates it with a unique and persistent entity URL. The user completes the description with the literal strings that make up the fingerprint data, and adds one or more statements and claims. In a late stage of the pilot, a second workflow was tested, which started with descriptions created elsewhere that were imported into the Wikibase editor using OCLC's Retriever utility.

The presentation of the use cases has been normalized slightly to aid comprehension. The first step is to introduce the item, establishing the ground truth of a real thing and commenting about why it is significant. The next step is to create the text glosses representing the fingerprint data required by the Wikibase editor. The third and most challenging step is the creation of relationships to other items through the use of Wikibase statements. Since the result is a fragment of a knowledge graph, which can be difficult to visualize, most of the use cases are illustrated with a graphic like the one shown in figure 8. The boxes identify item entities, and yellow boxes identify items of interest; all are populated with properties that link to other item entities through named relationships.

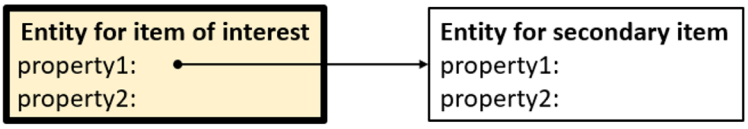


FIGURE 8. A schematic knowledge graph fragment.

In the framework of current workflows, the use cases represent experiments in both bibliographic and authority description and comprise:

- two studies of non-English description conducted by separate institutions
- studies of image resources conducted by four institutions
- two studies of archives and special collections
- a 15th-century musical work associated with an ecclesiastical event

Grouped thematically, these studies build a body of evidence that illustrate the lessons learned in the Passage pilot and longer-term perspectives (starting on page 59). The first two lessons—about editing workflow and ontology development—were relevant to every use case.

As originally envisioned, the Passage editing workflow would emphasize the creation of structured statements using the ontology imported from Wikidata as a baseline. But, as noted in the brief overview of the Wikibase editing procedure, the human editor must also supply the labels and brief descriptions that comprise the fingerprint data. For a well-known author such as Douglas Adams, the fingerprint data was straightforward; the editor needed only to supply variant forms of his name ("Douglas Noel Adams," "Douglas Noël Adams," "Douglas N. Adams") and a brief note about his occupation ("British author and humorist"). But the creation of fingerprint data in the Passage experiments turned out to be complex and frequently problematic.

Although the description task was designed to be open-ended, OCLC made two scoping decisions regarding the model of creative works. This is such a large and important topic for the library metadata creation workflow that it could not be given a comprehensive treatment during a short pilot. Instead, OCLC encouraged participants to focus on what is highlighted by linked data principles: the work as the first realization of a creative impulse, which has a physical presence that can be experienced through at least one sensory medium.

For a translated book, the case described in the section "Works, translations, and Chinese-language descriptions" below, the scoping decision generated an interest in the characteristics of the original edition, to which the translations could be associated using the Wikidata "translated from" property. And for image objects, the subject of the section "Visual resources" on page 39, the task was to describe the original physical map, photograph or poster, which could be associated with a web-accessible image using the Wikidata "digital representation" property. These issues are on the critical path to a model of creative works and are consistent with the outcomes of OCLC's previous research.[57] But community-wide effort is required for a fuller resolution, as noted in Lesson 2 on page 61.

WORKS, TRANSLATIONS, AND CHINESE-LANGUAGE DESCRIPTIONS

Two use cases explored the Wikibase/MediaWiki solution for multilingual resource description in the library metadata-creation workflow, which was available to Passage in the initial software installation. For better technical context, it's helpful to take a step back.

In an application such as Wikidata, the goal is to create a single entity—for Douglas Adams, for example—which can be understood and possibly enhanced by a worldwide, multilingual audience of information seekers. Using the editing interface, an English-speaking editor of the Douglas Adams entity can write a set of structured statements that can be consumed by readers of German, Chinese, Wolof, Tagalog, or nearly 300 other languages with the press of a button. The editor doesn't need to know any of these languages, and the reader can access the editor's work simply by changing the browser setting to display the data in a different language and script.

The impact is most evident when the dataset created by Wikidata—and, by analogy, Passage—contains no duplicate entities. This outcome is an aspiration, of course, because an English-speaking editor may be working on the Douglas Adams entity while a Catalan-speaking editor is creating a second entity. Fortunately, automated processes (or bots)[58] available from the Wikibase/MediaWiki software package may scan a Wikibase dataset to identify and resolve entities that resolve to the same person or another real-world object. Duplicates can also be merged by human editors.

When the Wikibase solution is used for multilingual resource description, at least one of the following tasks is performed:

- A subject matter specialist or metadata librarian creates a model and populates it with structured data for the study of translation and related areas of inquiry that are fundamentally about multilingualism.
- The editor sets the browser display to a language of choice and adds fingerprint data. For example, the Wikidata entity for Douglas Adams[59] has been modified by German editors, who added the fingerprint description "britischer Schriftsteller." But because the entity has not yet been touched by an Aragonese editor, as of April 2019, the display defaults to the English string "No description defined."
- With the browser display set to the language of choice, the human editor creates one or more

structured statements representing verifiable facts about an item of interest. Readers can consume the result in their own preferred languages by changing the browser setting to a different language.

The first task is the objective of the study by Karen Smith-Yoshimura and her OCLC colleagues. The studies conducted by Xiaoli Li and her colleagues at the University of California Davis Library performed the other two tasks. They show how easy it is for a Chinese-speaking cataloger to create and enhance Wikibase entities in Passage for important people and places in Chinese history. These studies show that a workflow featuring data organized as collections of machine-understandable statements about real world objects can yield precise results and make some current practices obsolete. This experience is also discussed in Li (2018).[60]

A translated work of philosophy

The cream of the world's cultural and knowledge heritage is shared in translation, a time-tested channel for learning about other cultures. According to a 2013 analysis of WorldCat by OCLC Research, 7% of the 45 million unique authors identified at that time had written works that had been translated into at least one language.[61] A 2017 OCLC study showed that Wikidata has entities for 180,000 books.[62] Most have multiple translations, one of the best clues that these are the among the works that have had the most impact on readers worldwide.

Because many of these entities also have OCLC Control Numbers, or OCNs, they can be compared with corresponding MARC descriptions accessible from WorldCat. Results show that WorldCat has good representations of the writing systems for some languages such as Chinese and Japanese, but primarily romanization-only text strings for others written in non-Latin scripts. Metadata for most Russian and Hindi works, for example, is represented only in transliteration, which can hinder the ability of readers of those languages to discover and retrieve the works they are interested in. In contrast, Wikidata entities can be expressed in the writing systems of any language supported by Wikibase, and are thus more likely than MARC records to contain text that matches the queries of native speakers of languages written in non-Latin scripts.

In contrast, Wikidata entities can be expressed in the writing systems of any language supported by Wikibase, and are thus more likely than MARC records to contain text that matches the queries of native speakers of languages written in non-Latin scripts.

The description of a book and its translations was one of the first use cases conducted in the Passage pilot, and one of the first homework assignments for the participants. To orient new users, OCLC added a page to the Passage instance of the Wikidata help portal, excerpted in table 2, which recommends a template for populating item entities for creative works and their translations.

TABLE 2. Recommended Wikibase properties for works and their translations (Source)[63]

Work	Translation
Title (original title)	Title (the translated title)
Language (of original)	Language (of translation)
Instance of (book)	Instance of (translation)
Author(s) (of original)	Translator(s)
Earliest known publication date	Earliest known publication date (of translation)
	Translated from (selected from auto-suggest)

The Passage participants also discussed the addition of a "transliteration" property, since not everyone can read non-Latin scripts. As a proof of concept, a "transliteration" property was added to the Passage Wikidata ontology, with sub-properties for schemes such as ALA/LC romanization, Wade-Giles, and kana representation. Other transliteration schemes could be added as needed.

The book selected for description was *Sein und Zeit*, one of the most important works written by the German philosopher Martin Heidegger, which was first published in 1927. According to the English-language Wikipedia entry,[64] *Sein und Zeit* was a groundbreaking work of early 20th-century philosophy, which influenced the philosophies of existentialism, hermeneutics, and deconstruction, as well as the then-incipient study of human perception. Accordingly, this book has a complex publication history containing multiple editions and translations. It has also been the subject of countless reviews, analyses, histories, and other scholarly studies.

OCLC's interest in the representation of *Sein und Zeit* in the bibliographic universe predates the Passage pilot. For example, *Sein und Zeit* is among the works of philosophy that make up a corpus of metadata extracted from WorldCat that is being studied in the CatVis project,[65] supported by a grant from the Netherlands Organisation for Scientific Research,[66] whose goal is to create a state-of-the-art visual-analytics toolkit that addresses the needs of digital humanities researchers as they engage with library collections.

An initial analysis showed that the relationships shown in table 2, although technically expressible in MARC, cannot be transformed automatically to linked data because of missing or incorrect data values and an abundance of text strings in multiple languages of cataloging, sometimes in the same record.[67] Because of these obstacles, visualizations that identify meaningful navigation paths for exploratory search and knowledge discovery have not yet been successfully created from MARC records.

As noted in the introduction to this section, OCLC staff made a scoping decision regarding creative works: that the concrete and observable features of library resources should be emphasized instead of the high-level abstractions captured in FRBR and related library-community models. In the case of *Sein und Zeit*, the effect of this decision was to favor the characteristics of the first edition and the set of known translations. Figure 9 shows a photograph of the first edition published in 1927.

FIGURE 9. A photograph of the first edition of Martin Heidegger's *Sein und Zeit*, autographed by the author (inset). View image source online. View the item in the Passage Explorer or Passage Wikidata UI.

After declaring an entity for the item of interest, the human editor is prompted for fingerprint data, which enables users of the Wikibase editing application to discover entities described in their chosen language. This task creates a challenge for an editor working in English about a book written in German because the resulting description will be a mixture of the two languages. OCLC addressed this problem in the tutorial Adding Labels, Descriptions, and Aliases in Multiple Languages,[68] published on the Passage help portal. The fingerprint data shown in table 3 conforms to these recommendations. The English-language fingerprint data for *Sein und Zeit* is shown in the second row, followed by the English fingerprint data for the English translation, *Being and Time*.

TABLE 3. English-language fingerprint labels for *Sein und Zeit* and an English translation.

Label Language of user	**Description** Language of user	**Also known as** Language of user*
Being and Time	Philosophy book by Martin Heidegger	*Sein und Zeit*
Being and Time (English)	English translation of Heidegger's *Sein und Zeit* (translators: McQuarrie and Robinson)	*Sein und Zeit*

* May be supplemented with aliases in other languages associated with the item.

The fingerprint data is designed to be comprehensible to a monolingual English reader, but when the label is a book title, the language of display can create confusion. Here, the labels in the fingerprint data give the mistaken impression that *Being and Time* is a translation of *Being and Time*, not *Sein und Zeit*. The fingerprint descriptions hint at the correct relationship, but it is expressed more rigorously in the machine-understandable statements for the Wikibase entity accessible from the canonical URL for the original work: the language is German, the title is *Sein und Zeit*, and the publisher is Max Niemeyer Verlag. A link to the photo of the book shown in figure 9 provides additional ground truth.[69]

Figure 10 shows a simplified view of entities representing the original German edition of *Sein und Zeit*, as well as the English, Hungarian, and Italian translations. The quoted text at the top of each entity is the fingerprint label that is displayed in a browser when the language is set to English, and is the same for each entity: *Being and Time*. The entity for the item of interest, shown in yellow, is the original German-language edition, whose title is *Sein und Zeit*. It is appended with "(German)" to indicate that the language of the title string is different from the language of display. The objects of the "title" property for the Hungarian and Italian translations follow the same conventions. The arrows identify all three translations as derivatives of the same German original. The "translation" entities also illustrate the use of three properties from the "translated works" template shown in table 2: "translated from," "language (of translation)," and "translator."

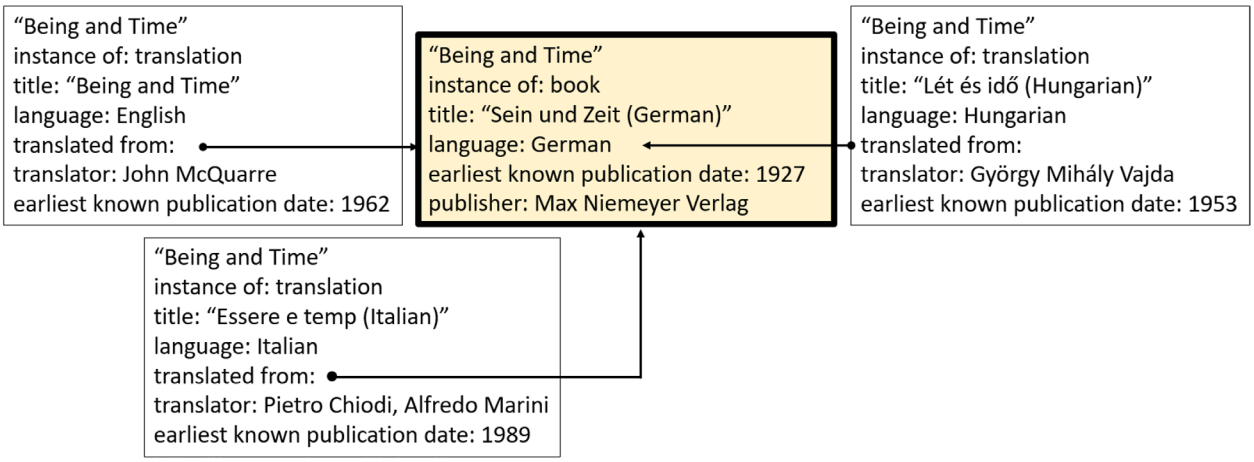


FIGURE 10. Wikibase entities for *"Sein und Zeit"* and three translations, formatted for display in English.

The "translated from" property establishes a clear relationship between the English text and its German-language source. But the corresponding "translated to" property linking the original to the translations has not been populated. Nevertheless, these connections can be viewed from the Passage Explorer view of *Sein und Zeit* displayed in English, shown in figure 11. The list compiled under the "Translated" heading on the right is generated in real time from a SPARQL query ranging over the entire Passage dataset. Thus, it is only when the discovery interface reveals the relationships between the original German work with multiple translations that the structured data created through the editing interface starts to make sense and reveal its value.

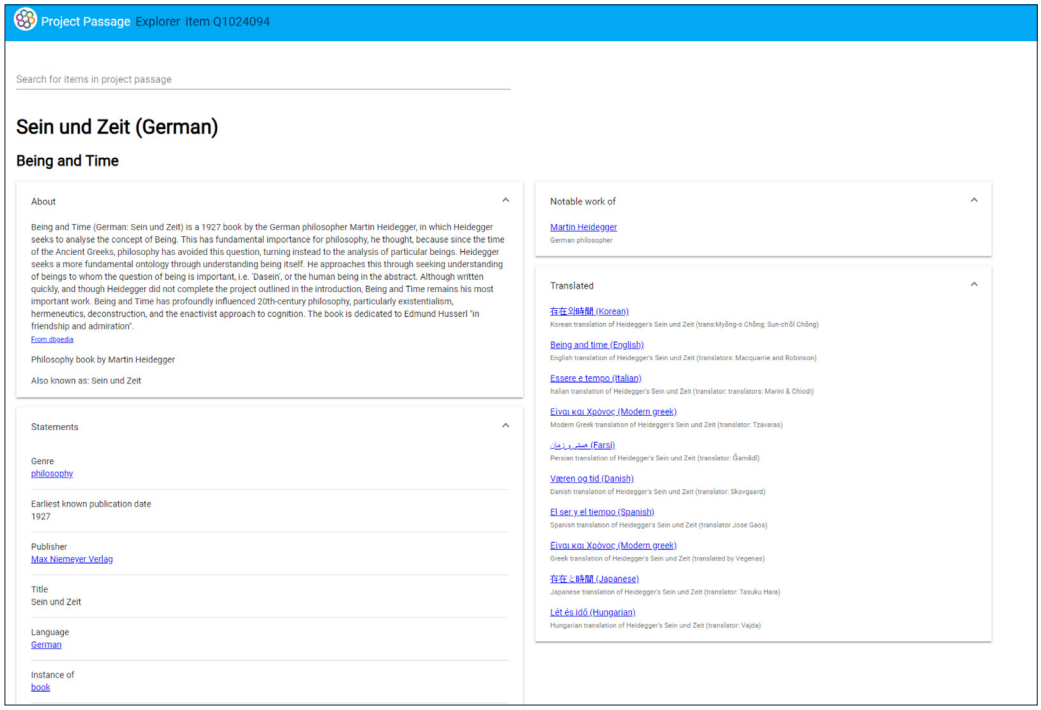


FIGURE 11. The Passage Explorer view of translations of "*Sein und Zeit*." View a larger image online.

The pilot participants discussed the merits of a best practice recommendation to include reciprocal links for every translation. But this solution was rejected as too difficult to maintain because the translation source would have to be updated with a "translated to" property whenever a new translation item entity for *Sein und Zeit* is added to the dataset. The Explorer display is a proof of concept that the missing links can be computed instead.[70]

As a related exercise in knowledge discovery, OCLC issued a query on the Passage RDF dataset using the SPARQL query interface installed with the Wikibase platform. This query requested Passage entities that represent the translations of *Sein und Zeit*. Each result included, if present, the title and translator expressed in the target language and script, the displayable English name of the target language, and the earliest known date associated with the work. The results were formatted with a Wikibase visualization tool that generates a timeline. An excerpt is shown in figure 12.

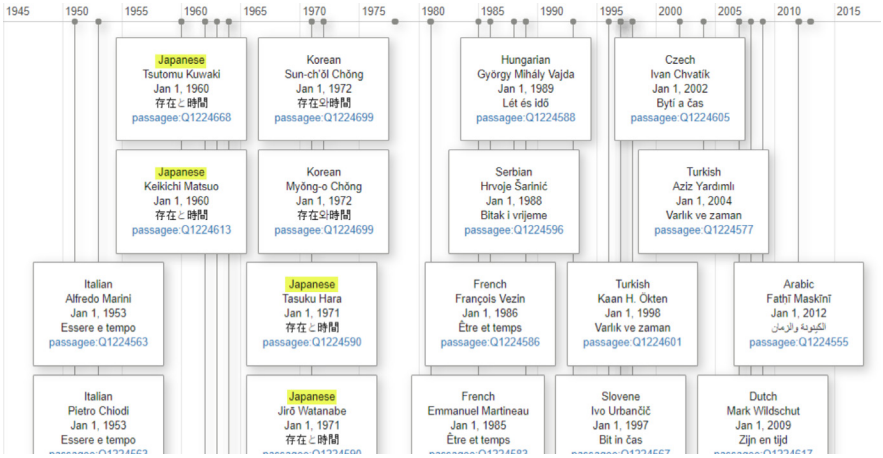


FIGURE 12. A timeline of Passage descriptions of translations from "*Sein und Zeit*." View image source online.

Many details of this visualization are important for scholarly inquiry. First, the visual history provides a snapshot of the path traveled by *Sein und Zeit* across linguistic and cultural boundaries. The image also identifies multiple translations into the same language. For example, the four entities highlighted in yellow represent Japanese translations, each by a different translator. When languages are as different as German and Japanese, there can be marked differences in translations, which is why it is crucial to identify the respective translators. Finally, and perhaps most fundamentally, the item entities for translations into languages expressed in non-Latin scripts, such as Είναι και Χρόνος (Greek)[71] and هستی و زمان (Persian),[72] show that the Wikibase platform promotes accessibility for readers of these texts in different languages. In contrast, the corresponding WorldCat records for these translations contain only romanizations, which are less meaningful to native speakers.

In sum, the visualizations generated by SPARQL queries on descriptions of works and their translations as represented in the Passage pilot are already compelling and exceed what can be accomplished with raw MARC records.

Wikibase "person" and "place" entities in Chinese

Catalogers at the University of California Davis examined the implications of the Wikibase solution for non-English resource description in the authority-control workflow.

A person: Sun Yat-sen

In the first test, a UC Davis cataloger created a Passage entity for Sun Yat-sen (孫中山), a physician, author, philosopher, and the founding father of the Republic of China. Working first in the English display and later in the Chinese display, the cataloger created many different forms of Sun Yat-sen's name and entered them as fingerprint data. The entity was populated with statements containing the properties glossed in the English display as "occupation," "sex or gender," "birth date," "death date," and "place of death" but were simultaneously available in the Chinese display with Chinese-language glosses. An additional statement links to the image shown in figure 13 in a statement using the "Wikimedia Commons" property. After populating the entity with English and Chinese text, the UC Davis team tested the effects of their changes on search and retrieval in the editing interface.

Significantly, the native support of non-Latin writing systems offered by the Wikibase software platform challenges the usefulness of the current library practice of designating a single authorized text string for a Chinese person. For the item associated with identifier Q459121, the Davis team found that either "Sun Yat-sen" or "孫中山" could be used as a preferred label, depending on the language setting of the interface for display or searching.

FIGURE 13. A photograph of Sun Yat-sen. View Image source on Wikimedia Commons. View the item in the Passage Explorer or Passage Wikidata UI.

Figure 14 shows the browser displays available for English and Chinese. "Sun Yat-sen" is the preferred label when the display language is set to English, while "孫中山" is the preferred label when the display language is set to Chinese "中文." Similarly, when the interface language is set to English, "Sun Yat-sen" appears as the first entry on the search results; but when the interface language is Chinese (中文), "孫中山" appears as the first entry in the search results. The simplified forms of the name (such as 孙中山) are also listed as aliases.

http://18.218.102.193/entity/Q459121

Sun Yat-sen

Chinese physician, politician and revolutionary

edit

Sun I-hsien | Sun Yixian | Sun Yi Xian | Sun Wen | Sun Zhongshan | Sun Yat Sen | Sun, Yat-sen

In more languages Configure

Language	Label	Description	Also known as
English	Sun Yat-sen	Chinese physician, politician and revolutionary	Sun I-hsien Sun Yixian Sun Yi Xian Sun Wen Sun Zhongshan Sun Yat Sen Sun, Yat-sen
Chinese	孙中山	中华民国的主要创立者，中国国民党创党总理	孙逸仙 孙文 孙中山 孫逸仙 孫文

Item 讨论 阅读 查看历史 更多 搜索Project Passage

http://18.218.102.193/entity/Q459121

孙中山

中华民国的主要创立者，中国国民党创党总理

编辑

孙逸仙 | 孙文 | 孙中山 | 孫逸仙 | 孫文

更多语言 配置

语言	标签	描述	别名
中文	孙中山	中华民国的主要创立者，中国国民党创党总理	孙逸仙 孙文 孙中山 孫逸仙 孫文
英语	Sun Yat-sen	Chinese physician, politician and revolutionary	Sun I-hsien Sun Yixian Sun Yi Xian

FIGURE 14. English and Chinese displays of the entity representing Sun Yat-sen in the Wikibase editor.

This case convinced the UC Davis team that it may not be the best use of a cataloger's time to dictate a single authorized name string for a person or other real-world object in every language. Perhaps a more important task is to create fingerprint data in as many languages as possible so users can choose what works best for them. This case also demonstrates that Chinese-language catalogers need not take time to provide romanizations of Chinese resources, as speakers of Hindi, Korean, Russian, and other languages would create their own transliterations for the names of globally important people such as Sun Yat-sen:

सन यात-सेन (Hindi)

쑨원 (Korean)

Сунь Ятсен (Russian)

Figure 15 shows additional gaps that could be filled. Data imported from Wikidata shows that the Spanish and German fingerprints have already been improved with language-appropriate descriptions, but do not contain "Also known as" strings. The Polish data shows only an altered label, but is associated with the description inherited from English, the default display language. And the Tok Pisin fingerprint shows only the label and description imported from English, indicating that it has not yet been touched by a native speaker.

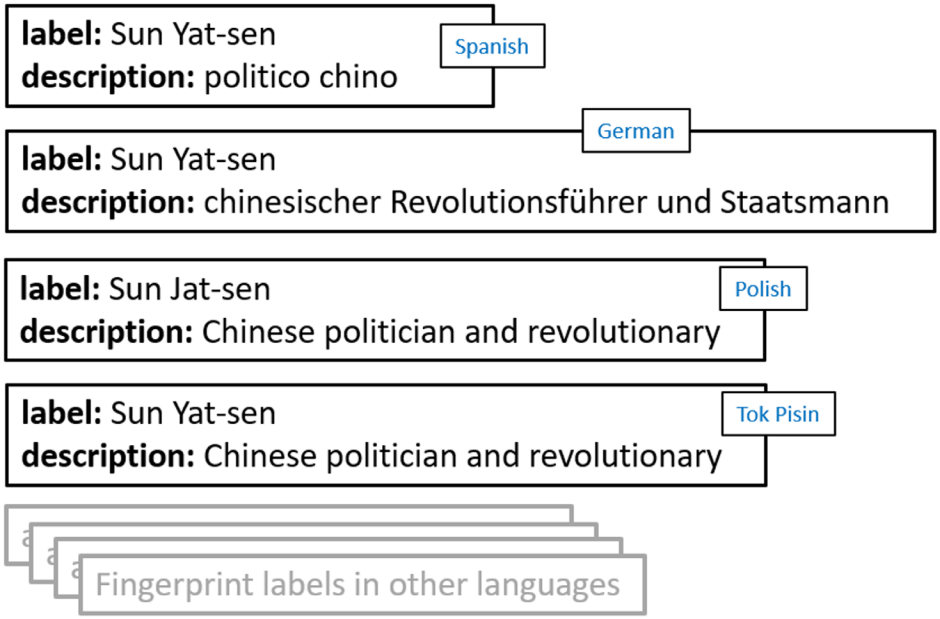


FIGURE 15. Fingerprint data for Sun Yat-sen in languages other than Chinese or English. View Wikidata source online.

The enhanced fingerprint data would improve the search and retrieval experience for speakers of languages other than Chinese, English, Spanish, and German. Authorized name strings would no longer be required for the task of disambiguation in the metadata creation workflow because the Wikibase multilingual functionality would do the work instead. By increasing the number of text strings associated with the founder of the Republic of China, the cataloger increases the odds that an information seeker will hit one of them.

In summary, the work on Sun Yat-sen was done by a cataloger who had minimal training on how to use the Wikibase editing tool, but she reported that the act of creating a "person" entity in the Passage editing interface was straightforward. The main task was to identify a person's essential attributes and apply the best predicates for encoding them, all of which were defined in the ontology ingested into Passage from Wikidata. Once created, it was an easy task for a bilingual speaker to switch back and forth between the English and Chinese displays and enhance them with additional fingerprint data.

A place: Bianliang

In the second use case, the UC Davis team compared the procedure for describing a historical Chinese place name in Passage with that prescribed by the current authority-control workflow. With its long history, China has many ancient place names that are no longer in use. But catalogers have been creating authority records for these names because they appear in works of literature, visual art, and other library resources that are important to researchers. One complication is that many historical Chinese place names lasted only as long as a named historical period. This fact cannot be represented as machine-actionable data in a MARC authority record, but it can be recorded as a structured statement in a Passage entity, as the next two figures show. In figure 16, which shows an image of the OCLC Connexion interface for NACO participants, text in the 670 "Source Data" field of the MARC authority record states only that Bianliang was later known by a different name.

ARN											
							1895120				
Rec stat	c		Entered	19870714		Replaced	20041120134058.0				
Type	z		Upd status	a		Enc lvl	n		Source		
Roman	n		Ref status	a		Mod rec			Name use	a	
Govt agn	|		Auth status	a		Subj	a		Subj use	a	
Series	n		Auth/ref	a		Geo subd	n		Ser use	b	
Ser num	n		Name	n		Subdiv tp	n		Rules	c	

010			n 87875872	Functions
040			DLC $b eng $c DLC $d DLC $d OCoLC $d DLC-R $d DLC	Functions
151			Bianliang (China)	Functions
451			$w nne $a Pien-liang (China)	Functions
551			Kaifeng Fu (China)	Functions
670			Han, T.S. Pien-liang so chi, 1986: $b t.p. (Pien-liang) p. 3 of text (Pien-liang was an older name of K'ai-feng [i.e. Kaifeng Fu--Cf. Ci hai, 1980, p. 901])	Functions

FIGURE 16. OCLC Connexion screen showing a MARC authority record for "Bianliang."

In contrast, the same information in the Passage entity, excerpted in figure 17, is fully encoded for machine consumption. A statement containing the "instance of" property states that Bianliang is a city; and a statement with the "time period" property specifies that the name was extant only during the Northern Song Dynasty (960-1127), when Bianliang was its capital.

http://18.218.102.193/entity/Q1225678

汴梁

北宋東京汴梁, 现河南省开封市 编辑

汴京 | 開封市 | 東京汴梁

更多语言 配置

语言	标签	描述	别名
中文	汴梁	北宋東京汴梁, 现河南省开封市	汴京 開封市 東京汴梁
英语	Bianliang	capital of Northern Song Dynasty, today Kaifeng	Bianjing
Chinese (Hong Kong)	没有定义标签	没有定义的描述	
Chinese (Singapore)	没有定义标签	没有定义的描述	

time period 英语 北宋 编辑

0个来源

Northern Song Dynasty

添加参考文献

添加值

FIGURE 17. A Chinese-language view of the fingerprint data and time period statement in the Wikibase editing interface for the historical Chinese place name Romanized as "Bianliang." View the full data in English or Chinese.

The cataloger who created the Passage entity commented that metadata could be created much more easily in the Passage editing interface than in a MARC authority record. The Passage entity also seems more useful because it has more detail. And most important, the built-in Wikibase features for multilingual support allowed direct input in the language and writing system that is most natural to Chinese catalogers and information-seekers.

This is an observation about a single historical place name, but it addresses a problem that is important in many fields of study, including history, archeology, and cartography. In China, and in most geographic locations that have a long recorded history of human settlement, place names are likely to have changed over time, but this detail is seldom encoded as structured data in the current cataloging environment.

VISUAL RESOURCES

Four visual resources—a map, a poster, a postcard, and a photograph—were subjects of use cases. In each case, the original object is managed in a library's special collection, and the content of the resource is more image than text. In cases where the object has been digitized, the focus of description is the original object, not the digitization—a scoping decision encouraged by OCLC, as mentioned in the introduction to this section. OCLC also recommended a set of guidelines for describing creative works,[73] which includes a recommended list of properties and creator roles relevant to maps, photographs, and other resources that are primarily visual. Most of these attributes were defined in Wikidata and imported into Passage:

- instance of (map, poster, postcard, photograph, scrapbook, movie)
- published title and subtitle; or when not available, an assigned image caption
- photographer, cartographer, surveyor, engraver, designer
- depicts
- a URL that resolves to the digital representation; to Wikimedia Commons, if available
- a production date, or earliest known publication date
- owned by
- identifiers from id.loc.gov, VIAF, and other library-community or relevant external sources

The map is presented first, followed by a joint consideration of the poster, postcard, and photograph. This organization recognizes that the map and the other three images present slightly different challenges. For example, maps and other cartographic images have well-established curatorial conventions that result in MARC records for individual maps, which contain extensive structured data for properties such as "projection" or "relief" and subclasses of cartographic material, which can be mapped to Wikibase items and properties. In contrast, non-map images are generally not associated with individual MARC records or other detailed machine-understandable metadata. Posters have traditionally presented special challenges for bibliographic description because their ephemeral nature has made intellectual and physical control difficult. In addition, a poster's emphasis on imagery has never readily lent itself to string-based textual descriptions. Moreover, all three images make critical references to the ontologically complex "event" item, which is not an important feature of the map described here.

A historical map

A team at Harvard Library led by Marc McGee shared the experience of describing a digitally scanned map published in 1852, a representative resource from the Harvard Map Collection.[74] It is shown as a thumbnail image in figure 18.

The map is city scale and shows natural features, man-made structures, names of landowners, roads, and districts boundaries. It also includes an inset map of the town center of Concord, a vignette of the Monument at the Old North Bridge, and indicates building features of the homes of some famous Concord residents, including authors Ralph Waldo Emerson and Nathaniel Hawthorne. In addition, it includes a note that White Pond and Walden Pond are described from surveys made by Henry David Thoreau.

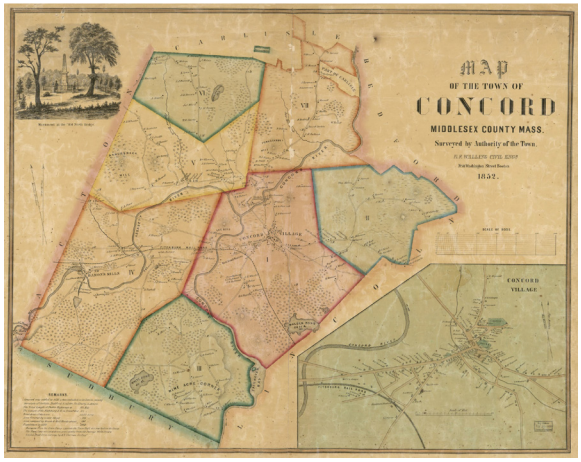


FIGURE 18. *Map of the town of Concord, Middlesex County, Mass.*: Surveyed by authority of the town by H.F. Walling, 1852. View a larger image online. View the item in the Passage Explorer or the Passage Wikibase Editing UI.

The fingerprint data, shown below, is designed to anticipate the terms that information seekers who may not be familiar with the collection might issue as queries. In this example, the resources mentioned in the fingerprint data have already been defined in Wikidata and were imported into Passage, including Concord, Middlesex County, and Henry Francis Walling.[75] These resources were linked with named relationships to the Wikibase entity that was developed for the map. The analysis produced entity-relationship pairs for map (instance of), H.F. Walling (cartographer), Concord, Mass. Middlesex County, and Massachusetts (depicts). In effect, the fingerprint data represents a human-readable but disambiguated overview of the most important connections that are expressed more rigorously in the structured data. This observation was also made in the case of *Sein und Zeit*.

Label: *Map of Concord, Massachusetts, 1852, by H.F. Walling*
Description: Map of the town of Concord, Mass. 1852—showing landowners and distinct boundaries
Also known as: *Map of the town of Concord, Middlesex County*

The published title, *Map of the town of Concord, Middlesex County, Mass,* was omitted from the fingerprint data because a string for a relatively obscure resource would likely be unfamiliar to most information seekers. But it is encoded in a structured "title" statement, where it is available for deeper knowledge discovery via SPARQL queries.

The Wikibase statements representing physical and technical metadata were patterned after MARC cataloging practice. For example, geocoordinates and scale are important structured data for maps. They are captured in typical MARC cataloging, and were expressed in Passage using the "scale" and "coordinate location" properties imported from the Wikidata ontology. The denominator of the scale "18400" was captured as a "quantity" datatype property, making it a machine-processable data point to allow for searching or sorting on scale ranges. Coordinates, expressed as latitude and longitude points, were recorded in a standard "geographic coordinate" data type property, which structures the data for re-use with geographic information systems applications. This data was used to generate the location of present-day Concord on the Google map displayed on the corresponding Explorer page.

Also typical of MARC map cataloging is the inclusion of a bounding box representing the greatest extent of the map west to east and north to south encoded as four geographic

coordinate locations (e.g., represented in a human readable format in MARC for this map as: W 71°26′30″—W 71°14′36″/N 42°31′12″—N 42°24′03″). A bounding box has the advantage of providing a more specific representation of the map, i.e., as an area instead of a point. Figure 19 shows a bounding box area of geographic coverage of the map using a Leaflet map plug-in.[76] This image is associated with a MARC record created at Harvard Library.

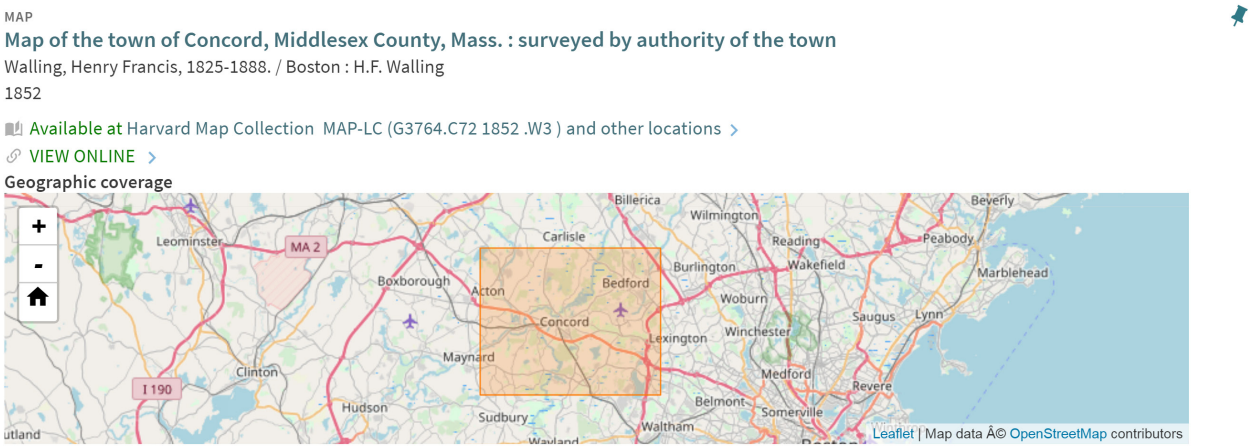


FIGURE 19. A map superimposed with a bounding box showing the location of Concord, Mass., accessible from a Harvard Library catalog record. View source online.

The encoding of the bounding box in the Wikibase entity for the map was only partially successful. As the Passage participants discovered, the "coordinate location" property defined in the Wikidata ontology requires a "geographic coordinates" data type, which does not provide for the encoding of bounding boxes. This issue was addressed when OCLC added the "geographic shape" data type property to the Passage ontology and defined it as a permissible type for the "geoshape" property. The "geoshape" property takes a Wikimedia Commons data file as an object that details the geometry of the entity in the form of a geographic file, such as the shape of the state of Massachusetts. But in the end, the "geoshape" property was not used on this map description, as a geoshape for Concord, Mass., was not available as a Wikimedia Commons file. Nevertheless, this example demonstrates one possibility for representing the geographic extent of a map.

OCLC also added Wikidata properties that would allow for physical description of the original object from which the digital object was derived, such as "height" and "width." This was a response to feedback obtained from the office hours discussions.

Content relationships were added using the Wikidata "depicts" property, which does not have a clear correspondence in MARC. In the Passage studies, this property was used in all image resources highlighted in this section. In MARC, the descriptor closest in meaning to "depicts" is "subject." But the two properties are applied in different circumstances, as the descriptions of three other digital images summarized below reveal.

In the Concord map example, "depicts" is used in two contexts. In the first, the main map depicts Concord and a segment of Middlesex County, as well as a dwelling owned by Emerson. But the sheet containing the map also contains an inset image. Since the illustration is ancillary to the main part of the map, an "image caption" qualifier was added to the "depicts" property for the entity "Obelisk Monument (Concord, Massachusetts)," referenced here using the title of the illustration. The inset map of "Concord Village" could similarly be described with a "depicts" property pointing to an entity for the village of Concord more specifically (rather than the entire town) and with an

"image caption" property transcribing the inset title, "Concord Village." These relationships permit the expression of searchable captions for each image in a compound image file, achieving a level of granularity that exceeds what can be expressed in MARC in a machine-understandable encoding.

The Concord map illustrates another nuance that can be recorded as structured data in Passage. An element commonly found on older maps and other types of materials is the inclusion of a street address for the location of the bookseller, publisher, or printer of the material (e.g., "No. 81 Washington Street Boston" on this map example). When establishing or maintaining corporate entities, recording the street address and date would facilitate historic geographic research by allowing for library material data to be sources for the building of historic geographic information systems.

During the walkthrough session, OCLC recommended the addition of an entity for the publisher (Henry Francis Walling), to which the "located at street address" property was added. In addition, the address property was modified using the predefined "point in time" qualifier, which enables a precise statement asserting that circa 1852, the address of Henry Francis Walling, the publisher of the map, was "no. 81 Washington Street, Boston." Thus, the complex connections between the map and the publisher, location, and date are rendered as accurately as possible even though the publisher and address no longer exist. These relations are captured for browsing in the Passage Explorer page for Henry Francis Walling.[77]

Many of the content relationships discussed so far are illustrated in figure 20, which is patterned after the schematic knowledge graph introduced in figure 8. Here, as in the previous image, the entity for the item of interest is the map and is highlighted with a yellow background; and the items that provide additional interpretive context for the map are shown with a white background, linked with arrows indicating property relationships. In both kinds of entities, quoted text represents fingerprint data. The description is populated with references to some already defined Wikibase items located at other nodes in the graph, including Henry David Thoreau and Ralph Waldo Emerson. Others, such as Henry Francis Walling and Obelisk Monument, were newly defined in this case.

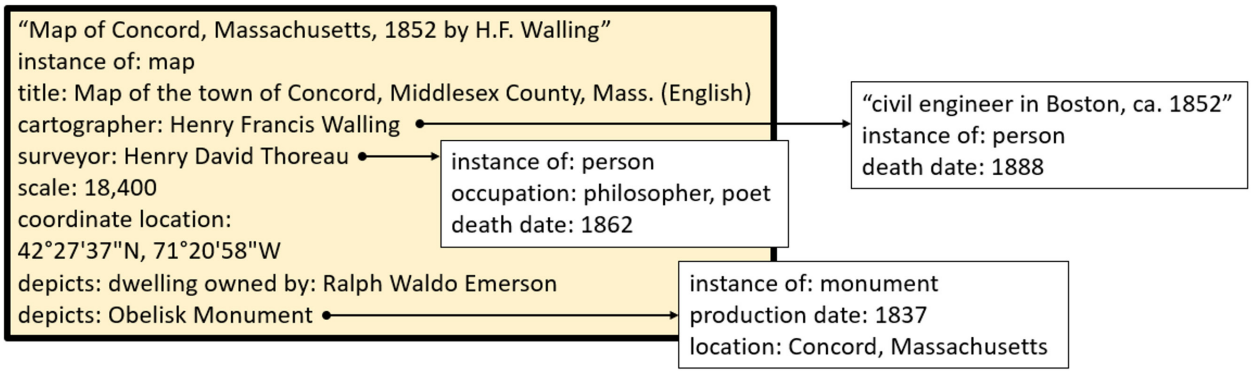


FIGURE 20. Relationships among Wikibase items for a historical map of Concord, Massachusetts. View the Passage Wikibase Editing UI.

To capture other relevant real-world relationships represented in the map, the Passage team modified the ontology imported from Wikidata. For example, Thoreau, the surveyor of two ponds on the map, was initially connected to the description by the "contributor" property, as "surveyor" was not defined as a property. Maps, especially older materials, often have multiple, specialty roles of contribution that are of research interest including: cartographer, surveyor, engraver, bookseller, etc. Having the option to be more specific than "contributor" for the roles of surveyor, engraver, etc., was desired. OCLC's engineers were able to add "surveyor" and "engraver" as properties to the Passage ontology, allowing for a more precise description of Thoreau's contribution to the map.

The map also presented the first of many opportunities to discuss date relationships during the Passage office hours. Since the publication date was printed on the map (1852), the relationship could be captured using the property "earliest known publication date" defined in the Wikidata ontology. But other "date" properties were also considered:

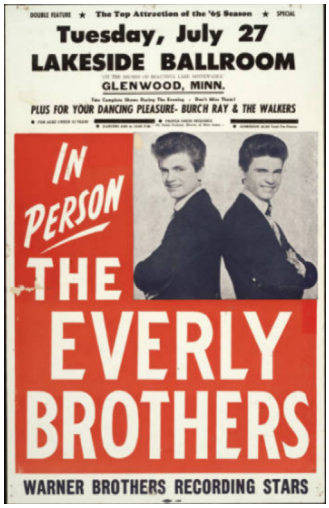
- "Inception date" describes the date that an object was brought into existence through an unspecified process, a property that would have been appropriate if the origin of the map had been unknown.
- "Production date" was added to the Passage ontology from Wikidata after the walkthrough of this use case. "Production date," defined as "year or date of production of a creative work, if different from the publication date," is more appropriate for creative works than "inception date" because the mode of creation is almost always known.
- "Date of situation" is a concept in map cataloging that represents the date of the content, or the date that the map data represents. This concept, represented by the "date depicted" property in Wikidata, was imported into Passage.
- "Point in time (date)" is defined in the Wikidata ontology as a qualifier on the "depicts" property. This encoding would establish a temporal range during which the "depicts" property is applicable.

These properties show some of the detail that can be captured by metadata librarians working with the Wikibase editing interface—in a process that is, of course, open-ended. For the map, the representation of the date was guided by MARC precedent, but the pilot participants who described the other image objects discussed below made different choices. For example, the entity for Maria Josepha illustrates the use of the "point in time" qualifiers, and the entities for archives include statements with the "inception date" property.

In summary, the Passage entity for a 19th-century map of Concord, Massachusetts, preserves nearly all the structured data in a MARC record. The Passage editing workflow also enables the open-ended addition of many other details and relationships, such as features visually represented on the map, that would be difficult or impossible to express be expressed in MARC, except as free text.

Photographs, a postcard, and a poster

The exploration of visual resources was continued in Passage office hour discussions led by librarians at the University of Minnesota, the National Library of Medicine, and Temple University. Like the map, descriptions of these resources focus on the physical object. If the object was digitized, it was associated with the online resource through the "digital representation" property defined in the Wikidata ontology.



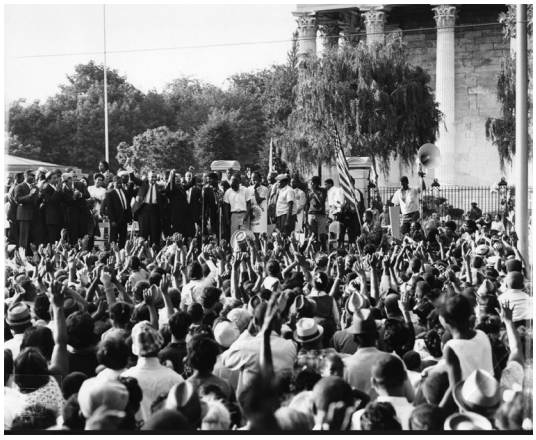
The Passage team from the University of Minnesota Library explored issues related to posters and events. Kalan Knudson Davis and her colleagues selected two posters from materials included in Minnesota Reflections,[78] a database of digitized original materials shared by cultural heritage organizations across the state. The example selected for discussion here is a poster for an Everly Brothers concert at the Lakeside Ballroom in Glenwood, Minnesota, in 1965. A thumbnail image of the poster for the Everly Brothers event is shown in figure 21. The caption has links to the full-sized image in the collection, and the relevant Passage descriptions.

FIGURE 21. A digitized image of a poster for an Everly Brothers concert in Glenwood, Minnesota, in 1965. Image source: Minnesota Reflections. View a larger image online. View the item in the Passage Explorer or Passage Wikibase Editing UI.

Karen Detling and her colleagues at the National Library of Medicine (NLM) contributed a description of a historic postcard. The postcard belongs to an archive managed by NLM's History of Medicine division, the Zwerdling Postcard Collection,[79] which contains over 2,588 postcards depicting nurses and nursing. The postcards span the decades from the 1890s to the 2010s, with many of the items dating to the golden age of the postcard, approximately 1907 to 1920. NLM is digitizing a selection of the cards for an online exhibition, including the one discussed here.

The postcard, reproduced as the digital image shown in figure 22, is a photograph of Princess Maria Josepha of Saxony. After she married Archduke Otto Franz of Austria, her title changed to Archduchess of Austria. Her brother-in-law was Franz Ferdinand, the Archduke of Austria, whose assassination in Sarajevo, Serbia, in July 1914 was one of the triggering events that started World War I. Maria Josepha worked as a nurse in World I and is shown in her uniform.

FIGURE 22. Postcard showing Princess Maria Josepha of Saxony in a nursing uniform. Image source: U.S. National Library of Medicine Digital Collections, Erzherzogin Maria Josefa. View a larger image online. View the item in Passage Explorer or the Passage Wikibase Editing UI.



The Temple University Libraries team headed by Holly Tomren created descriptions in Passage for two photographs represented in their digital collections: "Reverend Dr. Martin Luther King Jr. and Cecil B. Moore attend a protest rally at Girard College,"[80] shown in figure 23; and "Gay Student Movement Group" at Temple University shown in figure 26.

FIGURE 23. A digital image of a photograph taken at a protest rally at Girard College; Philadelphia, Pennsylvania, July 1965. Image source: Temple Digital Collections. View a larger image online. View the item in Passage Explorer or the Passage Wikibase Editing UI.

Fingerprint data for the entities representing three of the images is shown below and is populated with labels that might be entered as search queries in an application such as the Wikibase editing interface or the Passage Explorer. As in the case of the map, the pilot participants agreed that labels and descriptions should mention people, places, and events relevant to the interpretation of the image, providing readers an overview of the details encoded in the structured statements. Some are notable and were already defined in Wikidata, including the Everly Brothers, Maria Josepha, Dr. Martin Luther King Jr., Philadelphia, Girard College, and Minnesota. But others were not defined, such as the Lakeside Ballroom; Glenwood, Minnesota; and the protest rally at Girard College. Entities for these items were created during the process of describing the images.

Label: Everly Brothers performance poster, Lakeside Ballroom, Glenwood, Minn., July 27, 1965
Description: Everly Brothers performance poster, appearing with Burch Ray and the Walkers at the Lakeside Ballroom, Glenwood, Minn., July 27, 1965
Also known as: In person, the Everly Brothers

Label: Postcard of Maria Josepha
Description: Postcard of Maria Josepha in her nursing uniform during World War I
Also known as: Postkarte von Erzherzogin Maria Josefa; Postcard of Archduchess Maria Josepha; Postcard of Princess Maria Josepha; Postkarte von Prinzessin Maria Josef

Label: Reverend Dr. Martin Luther King Jr. and Cecil B. Moore attend a protest rally at Girard College
Description: Photograph of Reverend Dr. Martin Luther King Jr. and Cecil B. Moore attending a protest rally at Girard College, Philadelphia, Pennsylvania, 1965

These cases revealed that names entered as fingerprint data are potentially difficult to disambiguate or reconcile by human readers. The problem persists even if Wikibase items have been defined, because some people or places have similar names. For example, there are five historic figures named Maria Josepha of Saxony, who may be referred to by the name "Maria Josefa" or other spelling variants, according to the Wikipedia disambiguation page.[81] The item of interest in the NLM use case was the Maria Josepha who served as a nurse during World War I. This problem implies that fingerprint data may be of limited utility because it is not granular enough. When a query such as "Einstein" is submitted as a search term, the fingerprint descriptions returned as results are effective in distinguishing Albert Einstein the German-born physicist from Harry Einstein the American comedian. But when multiple Wikibase items have similar characteristics, users must do more work to distinguish among them.

After the fingerprint data was completed, attention turned to the creation of statements for each of the three objects. The creation of structured data turned out to be both easier and more challenging than current practice. The University of Minnesota team reported that this task forces metadata specialists to think about the item of interest and its named relationships to other items. It is easier in Passage because the editing interface empowers the declaration of items in a process that is much more lightweight and democratic than NACO registration or other established conventions for library authority control. But the intellectual task is more difficult because the human editor must decide how much effort is worth investing and when it makes sense to stop. In the process, Wikibase entities were established for secondary items associated with all three items of interest:

- **For the poster:** The Lakeside Ballroom; The Everly Brothers concert; Glenwood, Minnesota[82]
- **For the postcard:** nurse, Michael Zwerdling, and Postcards of Nursing[83]
- **For the photograph of the civil rights demonstration:** The Civil Rights demonstration at Girard

College, August 3, 1965; the Charles L. Blockson Afro-American Collection; John W. Mosley Photographs; John W. Mosley[84]

These new entities accomplished important goals related to creating machine-understandable data about the content of the resource.

Establishing a reference with uniquely identifying data

The entity for Princess Maria Josepha of Saxony establishes the correct reference and is connected to the postcard by the "depicts" property. In the image, Maria Josepha is wearing her nursing uniform, highlighting her work as a nurse during World War I. The initial proposal was that the "nurse" role belonged to the entity for Maria Josepha, as the object of the property "occupation." But the NLM team argued that this solution did not give enough historical context for interpreting the item of interest, i.e., the postcard. This need was addressed by adding a qualifier on the "depicts" role in the postcard entity, shown in figure 24. In the diagram, the additional qualifications are shown in the box with the dotted outline.

Together, the expanded definition of "depicts" with the object "Princess Maria Josepha of Saxony" produces a statement that can be paraphrased as "The postcard depicts Princess Maria Josepha of Saxony in the role of a nurse, circa 1914-1918." But this role was temporary, and in her longer-lasting career she performed multiple roles as a member of a royal family.

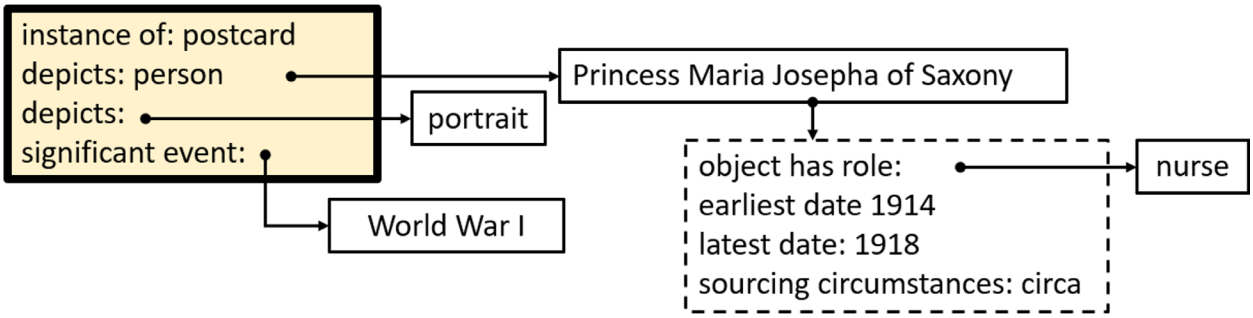


FIGURE 24. Wikibase entities for content relationships in the postcard. View the Passage Wikibase Editing UI.

Defining and naming events

The Passage help portal gives clear instruction on the treatment and description of named events.[85] The descriptions of some events, such as World War I,[86] were inherited from Wikidata, which may also be associated with established headings defined in the library community. Thus, as figure 24 shows, World War I can be entered as the predicate of a statement in the postcard entity using the Wikidata "significant event" property. In MARC, events are mentioned in the relatively new X47 series of tags, the historical use of the 650, and/or X11 series of tags for events, meetings, and meeting names.

In 2016, OCLC presented a discussion paper[87] to the MARC Advisory Committee that raised issues relevant to the design of FAST, which requires subjects to be separated into distinct facets, including events. This discussion paper addressed the coding of named events in the MARC 21 authority and bibliographic formats. One outcome was a new series of X47 fields in the MARC authority and bibliographic formats denoting named events. For example, the MARC field 647 identifies a named event used in a subject capacity expressed as a FAST heading. These fields are not currently approved for use in NACO authority records.

However, events whose headings have not been previously established, such as the Everly Brothers concert at the Lakeside Ballroom on 27 July 1965, or the civil rights demonstration at Girard College on 3 August 1965, are not well represented in traditional library cataloging practice or in MARC as structured data. Fortunately, the process is straightforward in the Wikibase editing interface. An "event" item is defined, and the Wikibase entity is populated, first with fingerprint labels and descriptions, and subsequently with structured statements that identify the date, time, and location of the event. Subjects can also be included, such as the headings listed in the description of the demonstration: school integration, African-American civil rights movement, and civil and political rights. Once defined, the event items simplify the descriptions of the photograph and poster. Now it is possible to say that the images "depict" an event, in addition to key participants, such as the Everly Brothers, Dr. Martin Luther King, and Cecil B. Moore. This relationship is illustrated in figure 25 below.

Building a context for interpretation

The items defined for the poster connect nationally known people and musical groups such as the Everly Brothers with relatively obscure places such as the Lakeside Ballroom and support serendipitous discovery. The same can be said about the other two images. Thus, every item relevant to the descriptions of the three images can be defined in Passage, and the only free text associated with these descriptions appears in the fingerprint data. The rest is structured, machine-understandable, and available for knowledge discovery by SPARQL queries like those that produced the timeline for the translations of *Sein und Zeit* described in the section on "Works, translations, and Chinese-language descriptions" on page 29 of this report. The same points can be also made about the description of a performance of ecclesiastical music, described below, which has more detail about the modeling of events.

These results imply that much of the text that appears in currently produced library metadata can be encoded in Passage as structured data. But the task of creating graph-organized data for image objects was sometimes disorienting or counterintuitive to the pilot participants. For example, the descriptions of photographs accessible from Temple's digital library included a variety of subject headings. In Passage, however, some of those subject headings were not applied to the description of the photograph itself, but to the people, places, or events depicted in the photograph. In some cases, the original subject could be three or four "skips" away from the entity for the photograph. For example, the image of the protest rally does not depict Philadelphia directly, but it depicts an event that took place at Girard College, which is located in Philadelphia. Thus, the Temple photograph also needed a "subject" relationship to other Wikibase entities to convey the context of the civil rights movement, desegregation of schools, and a direct connection with Philadelphia.

Some of the key relationships in this example are illustrated schematically in figure 25. Here, as in previous knowledge graph excerpts, the entity for the item of interest is in the yellow box and the related items are shown in white boxes. Fingerprint text is enclosed in quotes, and all other text represents items or properties that are linked through named relationships to Wikibase entities. The image suggests that most of the interpretive context for this image is encoded in the "depicts" statements, which have the effect of pushing critical details several nodes in the graph beyond the item of interest. Figure 25 also shows that, in contrast to the map description covered previously, the "depicts" statements do not replace but rather complement the "subject" statements inherited from the existing metadata record for this image.

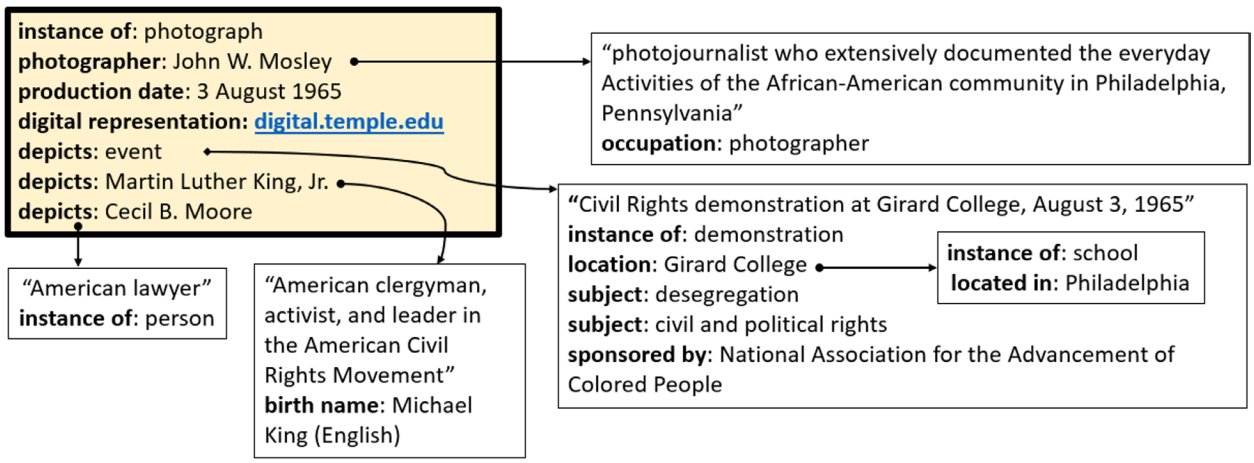


FIGURE 25. Wikibase entities for a photograph depicting an event. View the Passage Wikidata Editing UI.

It is clear that the photo of the civil rights demonstration reveals issues about context-building for items in a collection that were not fully resolved by the use of the "depicts" or "subject" properties. The experience raises questions about whether the issues can be addressed in structured data at all. For example, the second photograph discussed by Temple team—figure 26—showed a group of members of a Gay Movement student group at Temple University in 1974.[88] Because there was insufficient information to create separate entities for every named and unnamed student in the image, or the student groups to which they belonged, the Passage item entity for this photograph lists only "Temple University" and "activist group" as objects of the "depicts" property, failing to capture the LGBTQ context. The meaning was conveyed instead as free text in the fingerprint data, which contained the description "Photograph of students from Gay Movement at Temple University running a kissing booth on Valentine's Day."

FIGURE 26. A digital image of a gay student group at Temple University in 1974. Image source: Temple Digital Collections. View a larger image online.

The NLM group also reported a need to add textual annotations for images in the postcard collection, which are provided in the 520 field in a MARC record.[89] For example, one postcard in the NLM collection has the title "Ces anges des champs de bataille," which would provide no clue as to the exact nature of the image. The 520 field offers a wealth of information, not only about the image but its context:

> This is a hand-tinted French real photo studio postcard. It depicts a French nun, who is also a nurse, leaving a military hospital. She is dressed in a long gray dress, tied at the waist, with a large white collar, and long gray veil. She carries her belongings tied up in a cloth in her right hand. Her left hand holds a handkerchief and she looks sadly downward. A French soldier in blue and red uniform, is saluting her. The postcard refers to the fact that at the start of World War I military authorities believed it inappropriate for religious nursing orders to serve at military field hospitals.[90]

The description of creative works in the Passage pilot focused on the creation of structured statements that capture persistent and verifiable facts about a resource and links to related resources, as well as the labels that facilitate discovery of this data through a keyword search. But a full description may also have narrative or expository text about the resource, which establishes context, identifies content, and engages in commentary. How should such text be recorded? This issue is discussed in the "Reflections and perspectives" section on page 64, which argues that both narrative text and structured data are still necessary even after linked data has been fully adopted by the library community.

ARCHIVES

The Temple University of the civil rights demonstration is a good starting place for the description of archives in the Passage editing workflow. Now the item of interest is not the photograph, but its institutional home as an item (in the "archives" sense) in the Charles L. Blockson Afro-American Collection,[91] a leading research facility in the United States for the study of the history and culture of people of African descent. This collection has more than 500,000 resources in all formats: books, manuscripts, sheet music, pamphlets, journals, newspapers, broadsides, posters, photographs, and rare ephemera.

A subsidiary is the collection of John W. Mosley Photographs, which documents African American life in and around Philadelphia. According to the Mosley Collection digital collection description,[92] this is a "significant and extraordinary regional history collection, documenting virtually every social, cultural, and political aspect of life in the African American community during the mid-20th century." And as figure 24 above shows, John W. Mosley also took the photograph of the civil rights demonstration at Girard College.

These convergences are not serendipitous, however, because the Passage descriptions for the archive were created as a by-product of the need to describe the photograph. The process resulted in Wikibase entities for John W. Mosley, and the Blockson and Mosley collections, both associated with links to their official websites and associated with one another through the "curator of" property. Additional structured statements identify the Blockson collection as an archive housed at Temple University in Philadelphia. Since these relationships define a knowledge graph that is connected to the one shown in figure 24, the two fragments can be united.

The result is shown schematically in figure 27. Detail about the photograph available from figure 25 is suppressed, and placement of the photo inside the archival context is emphasized instead. As the figure shows, structured data can be created that positions the photograph as an archival item in the John W. Mosley Photographs collection, whose curator and institutional parent is the Charles L. Blockson archive.

FIGURE 27. Wikibase entities showing relationships in the Charles L. Blockson Afro-American Collection.

In effect, the two figures show different kinds of contexts that can be built using structured data. Figure 25 emphasizes semantic descriptors such as "subject" and "depicts," while the relationships asserted in figure 27 establish a trail of provenance, ownership, and resource management, which help the user navigate the institutions that house them. This analysis is only a sketch, but it can be theoretically scaled up with descriptions of other photographs in the collection and items in the archive.

Like the other use cases discussed in this section, the Passage collection descriptions were informed by standards and practices in the current metadata production workflow. Descriptions of archives in the United States can be created according to Describing Archives, a Content Standard (DACS),[93] and tend to be rich with narrative, which can be difficult to decompose into linked data real-world objects and their relationships to other objects. Another problem is that key relationships, such as those between people and items in the collection are not always explicitly specified, which means that it may not be clear whether a particular person is the author or subject of a resource. In addition, archival description rarely takes place at the item level, and generally requires some degree of critical reading and inference on the part of patrons to identify likely resources of interest within a collection.

Thus, the resource-description task demanded by the Passage pilot was singularly challenging even when positive outcomes were achieved, as in the Temple University use case. Other discussions were much more inconclusive. For example, the pilot participants defined different kinds of collections, such as physical collections, digital collections, partial collections, or artificial collections constructed from items that belong to multiple collections, but did not use these terms in their descriptions. The analysis of properties had similar outcomes. For example, date properties common to archival collections such as "inclusive dates" or "bulk dates" were proposed as additions to the Passage ontology, but were criticized by the participants because they erroneously implied that unpublished materials had been published.

A more fundamental set of issues was raised in the study of two archival collections by Honor Moody, Christine Fernsebner Eslao, and their Harvard colleagues. The Harvard Total Abstinence Society was a club that promoted temperance among university members from 1841 to 1843 and replaced the Harvard Anti-Wine Society, which operated from 1836 to 1840. The societies' archival records are described in a single bibliographic record in HOLLIS, the Harvard Library catalog.[94] Structured data for the organizations was created in Passage based on legacy authority records in a local MARC file.

The editing session produced Wikibase entities for the Harvard Total Abstinence Society and the Harvard Anti-Wine Society,[95] which included statements for:

- inception and dissolution dates
- reciprocal "replaces" and "replaced by" properties to connect the two societies
- "parent organization," which identifies Harvard University as the sponsor
- location in Cambridge, Massachusetts
- "instance of," which identifies the societies as associations with a social-improvement mission such as "temperance movement," another entity in Passage

These statements appear to provide some institutional and historical context for interpreting the significance of the two societies, in a style that is not qualitatively different from that created in the Temple University example. But in the Harvard instance, the structured statements sometimes implied a level of surety that was unsupported by narrative text about the two societies. The local MARC authority records included at least one 410 for the name of the club recorded as a subordinate body to Harvard University, 510s for earlier and later names, and a 678 with a free text excerpt from the organizational history note.

The use of the MARC 410 for the name of the club entered subordinately was especially problematic, as it prompted the inclusion of statements in Passage that Harvard University was the parent organization. Variant access points for subordinate entries in the authority records were largely for convenience purposes in an environment that presupposed browse indexes as a primary search tool, and the inference that Harvard was the "parent organization" is not supported by the organizational history notes for the two societies. Records for the two societies in the Library of Congress Name Authority File,[96] also stop short of direct assertions that Harvard is the parent organization. These textual descriptions imply an informal relationship with Harvard University, which is not accurately represented by the Wikidata property "parent organization." A fuzzier property such as "related organization" would be more accurate, but such generic properties become meaningless in large datasets.

Furthermore, the local authority records encode a sequential relationship between the two societies, which were duplicated in Passage using the properties "replaced by" and "replaces," as indicated earlier. But the original textual description presents a more complicated scenario. While both organizations are described by the same MARC record, each has its own history note and summary of its records. But the history notes do not tell the same story. For example, the Harvard Anti-Wine Society note states that it was replaced by the Harvard Total Abstinence Society, while the note for the latter says that it "was organized in October 1841, seemingly as a replacement for the then defunct Harvard Anti-Wine Society, with which it clearly shared some membership," a highly qualified statement. Several use cases described in this section show how Wikibase property qualifiers can be used in such circumstances, which could include hedges or other expressions of uncertainty. These issues are perhaps better addressed in free-text exposition, but progress is impeded if the text is inconsistent, inaccurate, or incomplete.

In summary, the treatment of archives in Passage is a paradox. Given the tradition of narrative and other text-based description, archives would seem to benefit most from workflows that facilitate the creation of structured data because so little is otherwise available. But the Passage experience revealed that much more community discussion is required to produce models that redraw the line between structured and textual data and creates best practices for producing both types.

A musical work associated with an event

The Consecration of Florence Cathedral, Santa Maria del Fiore, took place on the Feast of the Annunciation, 25 March 1436, upon the completion of the Cathedral dome by its architect, Filippo Brunelleschi. This event is closely related to the musical work *Nuper rosarum flores*, a ceremonial motet by the Renaissance composer Guillaume Dufay, commissioned for and performed at the consecration. The work as found in the manuscript source commonly known as "Modena B" is shown in figure 28.

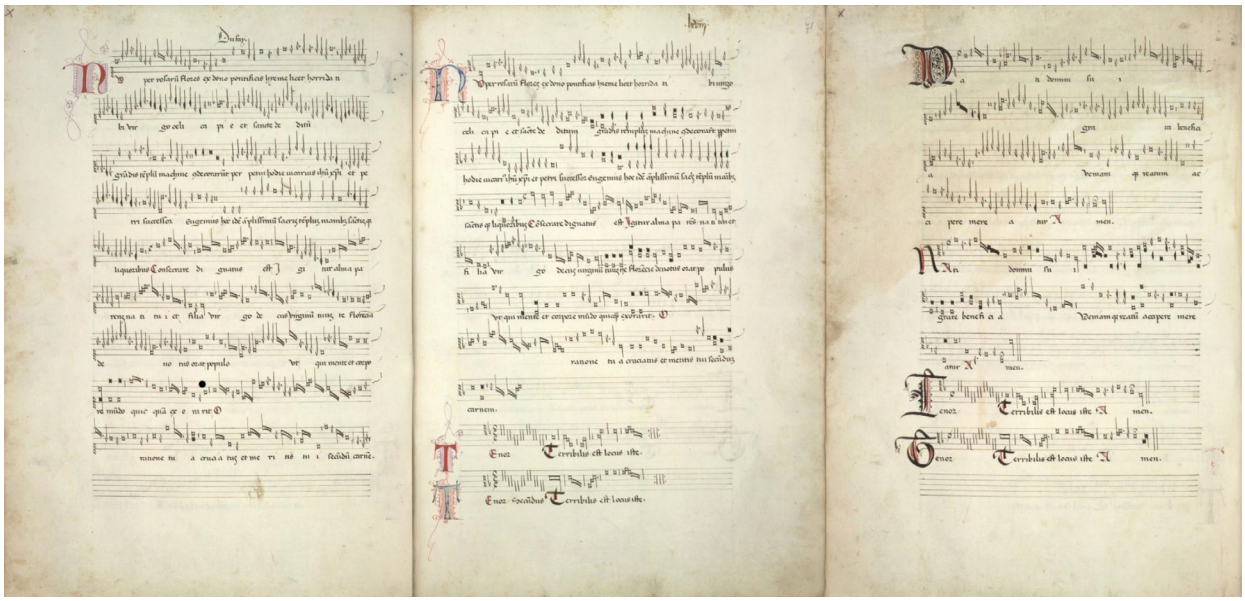


FIGURE 28. A reproduction of *Nuper rosarum flores* from the manuscript source known as "Modena B." View a larger image online. Source: Wikimedia Commons, public domain.

The relationship between Dufay's motet and the consecration of Florence Cathedral is the subject of another use case contributed by the Passage participants at Harvard University Library. The work was led by Craig Thomas.

Nuper rosarum flores is one of Dufay's more famous works, incorporating into the musical structure a portion of the Gregorian chant used for the dedication of a church, *Terribilis est locus iste*. Musicologists have long debated the proportions of the musical structure of the motet, 6:4:2:3. One theory, later discredited, attempted to relate these proportions to the proportions of the dome itself; in another, the musical proportions were related to those of Solomon's temple measured in cubits, as described in 1 Kings 6:1–20. In addition, the text of the motet refers to Pope Eugene IV and his gift of a golden rose on the high altar, the rose being an iconographic symbol for the Virgin Mary, to whom the cathedral was dedicated.

Given the historical and musicological importance of the event and the musical work as well as the web of rich interrelationships between the two, the Harvard Passage team decided that this would be an excellent use case to test the ability of linked data to express these relationships. From the standpoint of contributors to Passage, this use case illustrates some of the decision points entailed in creating and linking entities. This case discusses the treatment of events in Passage, the consecration of the cathedral (and the performance of the musical work), which were singular events. However, this use case does not address the question of periodically recurring events such as conferences and annual sporting events. It also touches on the practicalities of the treatment of inverse statements (corresponding to underlying RDF triples) between linked items in Passage.

Creating and editing items in Passage

The process began with the creation of a new entity denoting the Consecration. The fingerprint data is shown here.

Language English
Label: Consecration of Florence Cathedral
Description: Florence Cathedral, 25 March 1436, on the occasion of the completion of the dome built under the instructions of Filippo Brunelleschi
Also known as: Dedication of dome (Florence Cathedral); Dedication of Florence Cathedral; Consecration of dome (Florence Cathedral)

In early brainstorming, the Harvard team had originally given the consecration event the label, "Dedication of dome (Florence Cathedral)." However, this label was subsequently changed to "Consecration of Florence Cathedral." Although the item's consecration and dedication are both apt (and are therefore both included as values of the property, "instance of" within this item), the two terms are commonly confused. Both consecration and dedication are item entities[97] that were inherited from the Wikidata ontology ingested into Passage at the start of the project. The fingerprint descriptions of these terms include overlapping language. Consecration is defined as "a solemn dedication to a special purpose or service;" and dedication is defined as "the act of consecrating an altar, temple, church, or other sacred building."

At this event, Santa Maria del Fiore was consecrated as a place of worship and dedicated to the Virgin Mary. Of the two, consecration is arguably the preferred term since it is a sub-sub-class of event—"consecration" is a sub-class of "ceremony," a sub-class of "event"—and therefore brings out the event aspect, whereas dedication is ultimately a sub-class of "temporal item." Because of the common confusion between consecration and dedication, and between the consecration of the Cathedral and the completion of its dome, the use of "also known as" fingerprint labels to cover the various permutations—i.e., Dedication of dome (Florence Cathedral), Dedication of Florence Cathedral, and Consecration of dome (Florence Cathedral)—functions as a courtesy to users and can also help prevent future confusion and potential duplicate Passage items.

Statements were also created to cover the temporal and spatial aspects of the event. In both cases, property qualifiers defined in the Wikidata ontology were used to make the values of these properties more explicit; they were also used in the map and postcard descriptions discussed earlier. Thus, the Consecration of Florence Cathedral had an inception of 25 March 1436, which is an instance of the Feast of the Annunciation (this date chosen due to the dedication of the Cathedral to the Virgin Mary). Likewise, the event took place at the location of Florence Cathedral. The event commemorates both Brunelleschi's dome (an entity brought in from Wikidata via the Passage Retriever utility) and Florence Cathedral (which was already defined in Passage).

Turning to the musical work, the Harvard team noted that although there was no Passage item for *Nuper rosarum flores*, a brief entity for the work could be imported into Passage from Wikidata via the Retriever. The editing process added such essential information as the statements concerning the work itself (composer, title, language, genres, and subjects) as well as event-related information concerning the first performance of this work. The OCLC Passage team brought in the properties "date of first performance" and "location of first performance" from Wikidata upon request from a Passage participant. The entity also includes statements that:

- link to the Wikimedia Commons image shown in figure 28

- link the musical work to the initial event, using the Wikidata property "commemorates" to assert the "has part" relationship between *Nuper rosarum flores* and the Gregorian chant, *Terribilis est locus iste*, which necessitated the creation of a new Passage item
- link *Nuper rosarum flores* with corresponding authoritative metadata sources defined elsewhere, including those for the original Wikidata entity, the LCCN, and the VIAF identifier[98]

Figure 29 illustrates some important content relationships between the two items of interest just discussed. The outcome is the most densely populated knowledge graph developed in the Passage use cases discussed in this report. This result is perhaps a consequence of the scholarship in history and musicology that has established a set of verifiable facts about the two items shown in yellow boxes in the figure—the motet and the consecration—as well as an especially detailed ontology of events and musical genres available from Wikidata, which could also be easily extended with the terms required for this example.

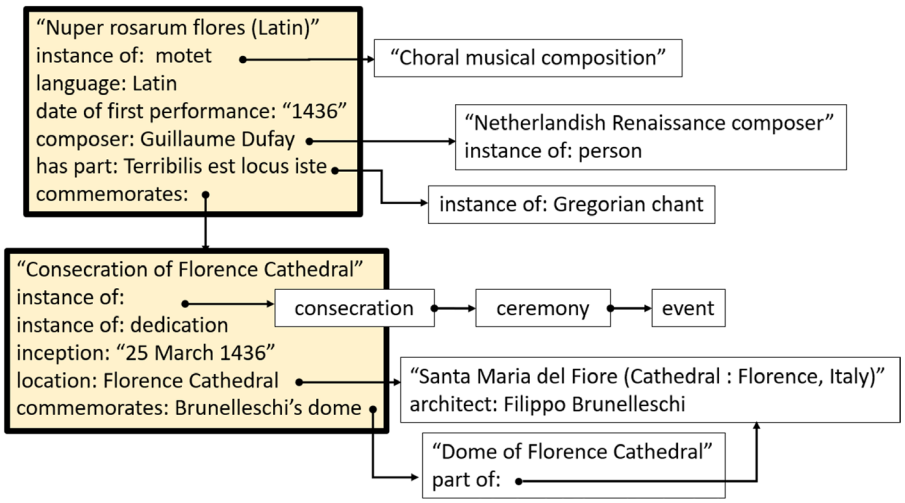


FIGURE 29. Relationships between *Nuper rosarum flores* (view the Passage Wikibase Editing UI) and the Consecration of Florence Cathedral (view the Passage Wikibase Editing UI) expressed as Wikibase item entities.

In fact, the figure under-represents the true set of connections because it shows only one-way relationships. As shown, *Nuper rosarum flores* commemorates the consecration of the Florence cathedral, but the inverse relationship is also true: the consecration is commemorated by the motet.[99] As in the study of *Sein und Zeit*, the inverse relationship can be automatically generated via a SPARQL query in the Passage dataset and revealed in the Passage Explorer, as shown in figure 30.

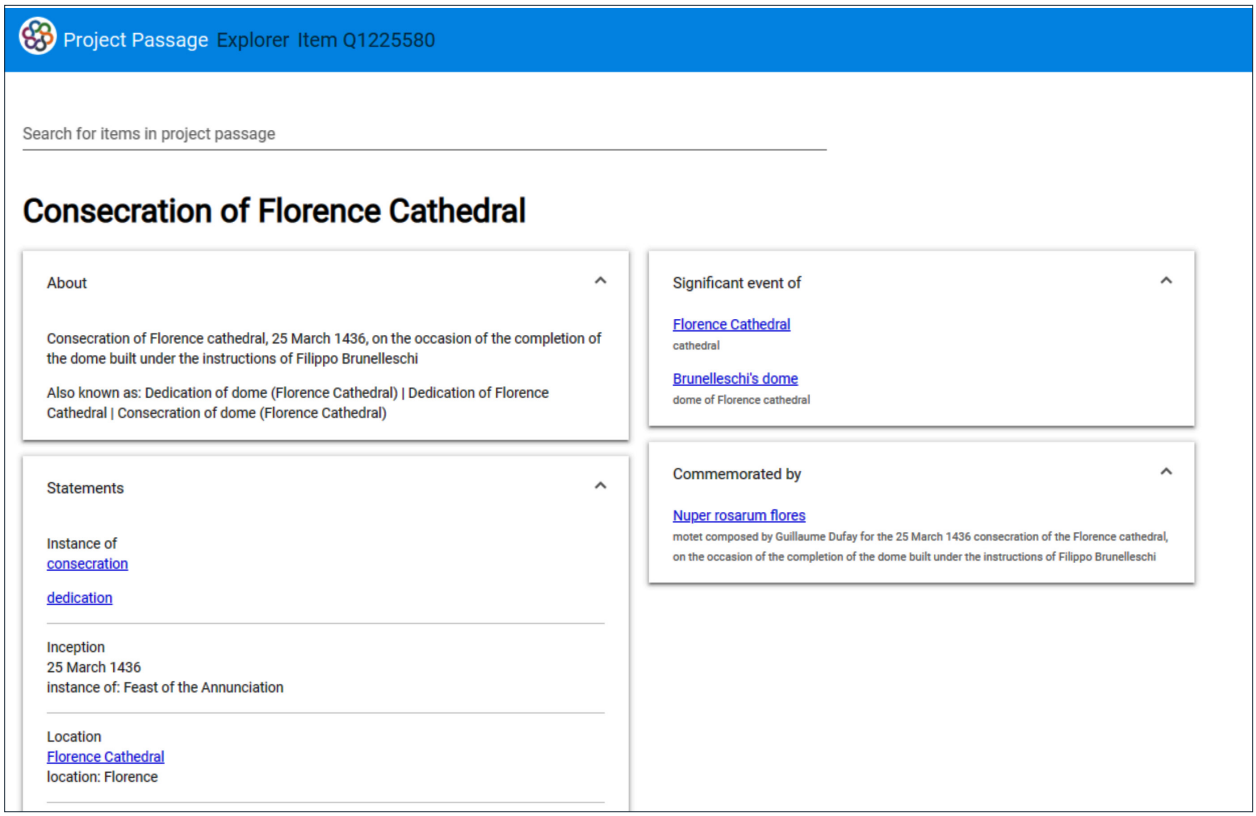


FIGURE 30. Passage Explorer view of the Consecration of Florence Cathedral, showing the computed "Commemorated by" relationship to *Nuper rosarum flores*.

The use case also reinforced the conclusion that catalogers should expect concepts in Passage and Wikidata to be organized differently than their counterparts in Library of Congress Subject Headings (LCSH) and the Library of Congress Genre/Form Terms, or LCGFT. For example, the broad musical terms, "choral music," "religious music," "vocal music," and "Renaissance music" were imported from Wikidata and are members of the "genre" class. The more specific term, "motet," however, is considered to be both an "instance" of "musical form" and a member of the ontological class "musical composition." By contrast, "motets" and "musical compositions" (the latter as a variant access point for "music") are considered to be form/genre headings in LCGFT.[100] The heading "motets" is also considered to be a subject heading denoting a broad musical form in LCSH. It is because of the placement of "motet" within the Passage/Wikidata music ontology that *Nuper rosarum flores* was asserted to be an "instance of" a motet. Asserting that it was a member of the "genre" motet would have been a constraint violation. This drove home the additional point that it is always important to ensure that one's assertions are ontologically accurate by carefully examining the Passage item for the desired term or concept.

To summarize in table 4, the description was built on the following entities imported from Wikidata and edited when necessary:

TABLE 4. Entities imported from Wikidata

Term	Entity
Named works of music	*Nuper rosarum flores*
Music ontology	Cantus firmus, Choral music, Introit, Motet, Religious music, Renaissance music, Vocal music
Named works of architecture	Brunelleschi's dome
Event ontology	Consecration, Dedication
Named events	Feast of the Annunciation

In addition, new entities were created for the event named Consecration of Florence Cathedral and the Gregorian Chant titled *Terribilis est locus iste*. Finally, the properties "date of first performance" and "location of first performance" were imported from Wikidata by OCLC staff. These items and properties create a network of relationships that can far exceed the detail currently represented in the corresponding MARC-based library authority files.

For example, the Library of Congress subject authority record for "motet" (sh 85087515) positions the term in a knowledge organization scheme that shows a relatedness to sacred vocal music (550 field, "Part songs, Sacred"). But the relationships are static, defined in the "broader" and "narrower" terms of a thesaurus. In contrast, relationships in the Wikibase entities are dynamic, encoded as named relationships that link directly to the items for motet, religious music, and related concepts defined in the Wikidata ontology and extended in Passage.

Authority records also lack machine-understandable detail associating the musical work with the 15th-century consecration event in Florence or with the topics mentioned in the accompanying lyrics, such as Brunelleschi's dome, Florence Cathedral, the Virgin Mary, and Pope Eugene IV. This sort of information might appear as text in a notes field, but such text often becomes scattered or lost. In contrast, the Wikibase entities defined in Passage for *Nuper rosarum flores* and the Consecration of Florence Cathedral act as collection points that make the relationships persistent and discoverable through keyword searching or SPARQL queries.

These relationships illustrate what can be added to a library resource description in the Wikibase editing interface by a cataloger with domain expertise. But even in the scope of current practice, the data can be improved with more explicit detail about creators, creative works, and the relationships among them. For example, the string "Dufay, Guillaume,‡d 1397-1474.‡t *Terribilis est locus iste*" is listed as an alternate name in the Library of Congress name authority record (no 98013583). But it does not contain the relationship with *Nuper rosarum flores* captured in the Passage description—namely, that Dufay is the composer of the motet into which he had incorporated a pre-existing Gregorian chant. Although the authority record identifies a "creator" relationship between Dufay and the two musical works as a by-product of the personal-name differentiation process, the association is only implied in the 100 field of the MARC 21 Format for Authority Data standard for personal names.[101] But in the Passage description, Dufay is explicitly related to Nuper rosaurm flores with the "composer" property.

SUMMARY OF USE CASES AND FIRST IMPRESSIONS

These use cases show that some were clearly successful and conducted with relative ease, an affirmation that the task of creating structured data and putting it to use makes sense in the library resource-description workflow and can be executed. But other studies were challenging and sometimes frustrating, a signal that the editing environment for creating structured data is still immature, or that the rationale for creating structured data is not always compelling. These issues can be explored in follow-up research.

One observation that is true of every use case is that all were motivated by the need to describe an especially compelling "thing," or item of interest, and establish its ground truth. But, once engaged, the participants realized that the task is open ended. As one Passage participant said, the framework implemented by the Wikibase platform is "protean," and one of the most vexing questions raised in the pilot was knowing when to stop the act of description. One consequence is that the library-community source might bear little resemblance to the Passage output, as the case of *Nuper rosarum flores* showed. Thus, the line between librarianship and scholarship is blurred, a realization that has significant consequences for library resource-description workflows.

Lessons Learned and Reflections

HOW PASSAGE DIFFERS FROM OTHER LINKED DATA PROJECTS IN THE LIBRARY COMMUNITY

Project Passage allowed editors to gain valuable insight into how to build the relationships that form the connections in structured, machine-understandable semantic data and obtain instant feedback about their work in a discovery interface. It provided a much-needed sandbox for metadata specialists to experiment freely and share their experiences with one other. This collaboration was beneficial both to the pilot participants and to OCLC staff. The pilot can be judged a success if the results inform community discussion, identify requirements, and set directions for appropriate standards and best practices, as well as products and services by OCLC and other service providers in the library sector.

Table 5 lists responses to OCLC's prompt for reflections on how the experience with Passage differed from other library linked data projects. The quotations identify the largely positive reactions to the technical characteristics of the sandbox, as well as to the modes of engagement featured in the pilot.

TABLE 5. Participants' comments about how Passage compared to other library linked data projects.

Difference	Sample quotes
Maturity	"The Passage editor is further along than many linked data editors." "Non-buggy, easy to use software" "All components were in place on the back and front end."
Openness	"Linked data was once reserved for the 'have' libraries and scoped at the institution level. Now we have the beginnings of something new. We can see and share the fruit."
Simplicity	"Very easy to connect theory and practice" "Usability, editor/interface better than any other project" "Style of editor so different, but intuitive" "The use of qualifiers hides some of the complexity for users."
Built-in support for multilingual description	"This one focused on multilingual data entry . . ."
Opportunity for feedback	"With the built-in SPARQL endpoint and query builder, it was easier to test consequences of ontology and data entry choices."
User-driven development	"Direct access to OCLC technologists" "Rapid, iterative software development"
Connection to library patrons	"The Explorer offered many of us a first taste of an actual public interface for a system based on native linked data."
Focus on new description	"Not primarily focused on automated conversion or reconciliation pipelines." "Focus on entities and non-linguistic data."
Pilot design	"Hands-on experience, homework assignments and office hours, community collaboration"

The shorter list of what did not work so well, as compiled by Pace and Tomren,[102] is largely a consequence of the design of Passage as an open-ended exploration. For example, the ontology isn't rich enough, and the Explorer and Retriever applications need more functionality. Another issue is that the pilot ended before the librarians whose work is most deeply affected by the transition from records to knowledge graphs had enough time to think about the implications of this change. But the end date also provided an opportunity for assessment.

Passage was unique because it created an environment for practitioners to experience their first deep encounter with linked data concepts in their current jobs, enabling them to make head-to-head comparisons between current and new standards of practice. The use cases abound with examples of the fine-grained analysis that was produced in every resource description that was undertaken.

But even looking beyond the property or statement in an individual Wikibase entity, the pilot participants were able to map current practices to the new paradigm, preserving some of the most important values of librarianship. For example, Fernsebner Eslao, et al., argued that the robust solution in the Wikibase platform for establishing provenance, authority, and trust at the level of the individual statement is technologically more sophisticated than corresponding library-community solutions.[103] This solution is available to librarians who are not engineers, but accomplishes a fundamental technical goal in the linked data environment: declaring the existence of an item as a linked data real-world object about which facts can be assembled and associated with a globally unique URI. Additional items and properties can also be declared in the moment, for a wide range of resource types, in a multitude of human languages and writing systems. The descriptions can link to library-community datasets, as well as to vocabularies and ontologies maintained elsewhere.

SEVEN LESSONS LEARNED

As a research study, the Passage experience provides an initial reality-check, both on the promises of linked data in the library cataloging workflow, and on the gaps that must be addressed before machine-understandable semantic data can be fully adopted as a production goal. The most important lessons learned are listed below. These are quick impressions, gleaned from the immediate experience of conducting the use cases.

Lesson 1: The building blocks of a Wikibase item entity can be translated into a straightforward procedure for creating structured data with a precision that exceeds the expressive power of current library standards. But can this data be used to build the context required to discover and interpret a resource?

Although the Wikibase editing environment was intuitive and easy to use, the pilot participants concluded that it is not ready to serve as a turnkey solution for the metadata creation workflow.

The essential task is to identify an item of interest—such as a person, organization, place, event, creative work—and to make factual statements about it. Each of the use cases started here, but a description of the thing of interest often demands the definition of related things. And even a simple description also demanded enhancements to the ontology. Thus, the linear task of assembling statements turned out to be complex because the concept space in a knowledge graph under construction is both multidimensional and full of gaps.

This is an issue of maturity, both of the ontology and the instance data that expresses it. But the impact was uneven in the Passage use cases. For example, gaps could be addressed easily in the description of *Nuper rosarum flores* and the event that inspired it because the Wikidata ontology was already richly populated with relevant concepts in musicology and types of events. But

archives and special collections presented greater challenges because it was not obvious what should be expressed as structured data and how this information could be used in a discovery interface to build an interpretation that is superior to the products of existing practice.

This outcome points to gaps in models of library special collections and the functionality of applications such as the Passage Explorer for some classes of resources. The outcome also suggests that human-readable narrative or expository text will continue to serve a useful function even after linked data has been adopted. Expert communities will need to determine where to redraw the line between structured data and text, now that a software solution such as the Wikibase platform makes it easier to create structured data than is possible with current standards and tools.

In addition to structured data, the editing workflow also requires the creation of fingerprint data, the labels and textual descriptions designed to match a user's query in a keyword search and serve as a human-readable summary at the top of a Wikibase entity. The need for fingerprint data required little justification, and its creation was often straightforward, but the use cases also raised many questions and concerns.

For example, the case of *Sein und Zeit* and its translations revealed the need to be clear about the language of display and how it is created. For structured data, the language of display is generated from a table when the user changes a browser setting. This request transforms the English property "author" to the German property "Schriftsteller," for example, and makes concomitant changes to the writing system if the requested language is expressed in a non-roman script. But the fingerprint data contains only string literals that are hand-coded by editors who are fluent in the selected language of display, as highlighted in the case of Sun Yat-sen. Fingerprint data usually matches the language of display, but it does not have a formally assigned language tag. To manage potential confusion, the Passage studies explored conventions for the selective and disciplined use of mixed-language labels.

Additional conventions were developed to manage the description of image resources that had no pre-defined labels, such as a poster or an unpublished photograph in a collection. To compensate, the pilot studies experimented with fingerprint data containing lists of Wikibase items, or linked data real-world objects, that were important for understanding the context of the object. But a fingerprint label is not guaranteed to be reconciled, disambiguated, or associated with a persistent URI. As a result, this convention generated an additional obligation to create Wikibase entities for many relatively obscure people, places, things, or events, such as the civil rights protest rally at Girard College depicted in the Temple University photograph. This obligation initiated some of the non-linear workflows described above. The use cases also showed that detailed fingerprint descriptions were sometimes created because there was no other place to put expository or narrative text in a Wikibase entity.

Taken together, the task of creating fingerprint data, like the task of creating structured data, could either be easy or difficult. How much work is required to verify that a fingerprint is uniquely associated with a resource in a production environment featuring datasets that are orders of magnitude larger than Passage or Wikidata, as well as multiple peer applications? Is fingerprint data that conforms to the conventions that evolved during the course of the Passage pilot optimized for search and retrieval? In the worst cases, the participants noticed that retrieval was effective only if the searcher succeeded in second-guessing the cataloger's input. Nevertheless, the creation of natural-language labels that are designed to match the language of a user's queries represents an important entry point into the dataset created by the Wikibase editing workflow. The problems identified by the pilot participants are identified as issues for future discussion.

Lesson 2: User-driven ontology design enabled by the Wikibase platform is a good thing. But how should it be managed for long-term sustainability?

The Passage pilot showed that a workflow centered on the creation of verifiable facts about an item of interest can produce exquisite detail. Credit for this result is due in large part to the Wikidata ontology, which is already rich enough to support innovation but is also designed to be extensible with community input. New Wikibase items and properties could be easily inserted into the ontology that Passage inherited from Wikidata to define nuances that addressed immediate needs. For example, maps, posters, and photographs led to new "date" subtypes such as "inception" and "production," as the pilot participants grappled with how to say that objects can come into being through channels that do not involve formal publication. Likewise, the musical work use case prompted the addition of event-related properties such as "date of first performance" and "location of first performance" as well as the addition of musical forms such as "motet."

The pilot participants also saw benefits in incorporating contributions from outside the library domain as properties and terminologies expand and change over time. For example, vocabularies used in social media are culturally current and apt to change more quickly than those used in bibliographic description. Incorporating new vocabulary and ensuring that it is widely understood are important goals for both the application of new terms and the specification of guidelines for how others use them. In addition, some practices in traditional authority work might be imported into the ontology, such as "use for" references, scope notes, "follows," "supersedes," and "superseded by." The application of these concepts builds a revision history that is useful for understanding the context for a term and how its meaning has changed.

Nevertheless, the pilot produced many hints that management of the ontology will be a major concern after the Passage experiment has matured. For example, some form of continuous maintenance is required to ensure that community-supplied terms do not clash. More significantly, the Wikidata ontology needs a model of creative works that addresses the library community's use cases. The Passage pilot restricted the scope of creative work to assertions that could be made about the earliest known realization of the creative impulse. But the larger issue of how to represent detail in hierarchical or tiered models of creative works such as FRBR,[104] RDA,[105] or the Library Reference Model[106] remains unaddressed. Finally, the use case on archives reported that the Wikidata and the library-community vocabularies all exhibit mixed results in their sensitivity to equity, diversity, and inclusion issues and could benefit from remediation.

In the end, the participants acknowledged the benefits of a democratically governed vocabulary while also voicing reservations about it. Democratic governance could generate conflicting proposals and multiple overlapping, but still incomplete, ontologies that would not be sustainable. Or such a model might require even more rules than the current resource-description environment has. To avoid these problems, ontologies and terminologies would perhaps need to be managed through a governance structure similar to the oversight of MARC by the MARC Advisory Committee and the PCC. If so, the library community would adopt practices similar to those that govern the creation of content in the Wikimedia ecosystem, where anyone can create a Wikipedia article or a Wikidata item, but only a small group is permitted to curate the most important features of the ontology. And those who contribute content to Wikimedia applications must learn how to bridge the cultural differences between ontology management in the library community and a more open model of ontology management.

Lesson 3: The Wikibase platform, supplemented with OCLC's enhancements and stand-alone utilities, represents a step forward in enabling librarians to obtain feedback as they work.

The Passage experience demonstrated that the discovery layer provided by the Explorer interface is crucial to the workflow for creating and editing entities and their relationships. With this, too, catalogers could gauge the impact of their work on the user experience as they worked. And the Explorer display could also offer hints that the act of description was nearing closure if it could be judged fit for purpose. For example, the Explorer view of the *Sein und Zeit* example showed the list of translations of the original German edition, which was mined from the Passage dataset and did not have to be added by catalogers at all.

Lesson 4: Robust tools are required for local data management.

The creation of MARC records and other textual descriptions that conform to library-community standards is a "what you see is what you get" experience: catalogers can see the total output of their effort, expressed in a compact and easily understood template. In contrast, linked data is always a mediated experience because it must be generated, mapped, sampled, or queried. As a result, practitioners may have difficulty seeing what an entity for a given item of interest contains and where the effort of one editor is distinguished from that of other contributors. Editors may not be able to view the nuances that they have just added, or they may discover a statement that they did not create. And if they wish to consume the data at a later step in the metadata-creation workflow, what do they request if a clearly demarcated database record is no longer an option?

This issue shows up in the missing translated-to and translated-from links in the *Sein und Zeit* use case. Fortunately, that problem was remediated with a SPARQL query issued by the Explorer across the entire Passage corpus. In fact, the SPARQL query interface and associated visualization tools may provide a useful stand-in until robust tools for local data management are built. SPARQL queries can help metadata creators discover knowledge, view subsets of data, and spot and correct errors. And SPARQL can help engineers envision what a future interface could be designed to deliver.

Lesson 5: To populate knowledge graphs with library metadata, tools that facilitate the reuse and enhancement of data created elsewhere are recommended.

To address this need, OCLC introduced the Retriever application, which issues a cataloger's string searches associated with an item of interest to Wikidata, VIAF, and local or national library authority files. Catalogers performed the essential task of reconciliation on the results list to identify items that had the same real-world references as their items of interest. Descriptions that matched

their queries could be imported into the Wikibase software environment for enhancement. In a production environment, no one wants to duplicate work that has already been done. A data architecture that minimizes duplicates is one of the most important reasons for adopting a linked data representation of library resources, and was mentioned as a key objective by the participants on the first day of the Passage pilot.

When feasible, the extraction of knowledge from legacy library data should be a high priority in the import process. To illustrate, consider a university professor who retires from the Department of Mathematics and donates her personal papers, teaching files, and publications to the university's archives. She might have an authority record that reflects her dates of birth and death in the 046 [Special Coded Dates] field. These are recorded in Extended Date Time (EDTF) format,[107] so they can be processed as data. The place of birth is recorded in 370 $a [Place of birth] and the professor's field of research—number theory—can be recorded in the 372 field [Field of Activity]. Because this is a personal name record, it is also possible to record the professor's occupation in the 374 field [Occupation], the universities where she earned her degrees, and the university at which she taught are recorded in the 373 field [Associated Group].

In the current environment, catalogers add these fields in the hope that they will be used to identify persons and corporate bodies very clearly. This information can be imperative in differentiating between authors with the same name or in determining whether authors with publications on very disparate topics are the same person. But catalogers also enter this data in good faith, putting in the extra effort to enter controlled vocabulary terms and use formatted dates, so that these fields will be searchable in future systems. That future has not yet arrived, but the Passage applications show how these fields can now be used.

Currently, the Wikibase editing interface allows for the assignment of these properties for a person or corporate body, but they are not automatically populated when an authority record is imported via the Retriever. However, because the Retriever has undergone so much improvement in the short time it was in use, it is natural to recommend that this functionality be added. If the Retriever could populate fields defined the in the Library of Congress Name Authority standard for RDA elements, then catalogers will be able to envision what their efforts can do and would be, in turn, more likely to include such useful information. This detail about relationships between creators and creative works can enhance the linked data that can be generated from this metadata.

Lesson 6: The pilot underscored the need for interoperability between resources, both for ingest and export.

One of the recurring questions that arose during the pilot was how information added to data that had been ingested from Wikidata could be exported and added back to the original source, a question also raised by data imported from library authority files. This is an instance of the synchronization problem for data publishers, articulated for the library community in Section 5 of the white paper National Strategy for Shareable Local Name Authorities, published in 2018.[108]

Lesson 7: The descriptions created in the Wikibase editing interface make the traditional distinction between authority and bibliographic records disappear.

People, places, things, organizations, and events are linked data real-world objects represented as Wikibase items, and so are creative works. All can be represented as a Wikibase entity, which is associated with a system-generated Wikibase item URL, commonly interpreted as a linked data URI, and is populated with a list of statements that refer to other entities. This common ground is why the process for describing *Nuper rosarum flores* is the same as that for Maria Josepha and Sun Yat-sen, and is the same as that for photographs of these historical figures, as well as for digitized images of these photographs and for collections of these images. In other words, one of the most important distinctions in traditional library metadata management vanishes when human-readable records are replaced with knowledge graphs. The consequences for education and training, professional specialization, the establishment of workflows, and community-wide governance and best practices remain to be analyzed.

The commonalities identified in the workflow for the description of a self-selected and disparate set of library resources can also be interpreted as a hint that Passage could gracefully shut down because the pilot succeeded in identifying and evaluating a repeatable process. But it also generated issues that demand more immediate attention.

REFLECTIONS AND PERSPECTIVES

In the final month of the Passage pilot, the team of authors for this report was assembled. Nearly all of the authors had also contributed use cases. After the office hours sessions ended, they regrouped to discuss long-term issues raised by the Passage experience, and, more generally, the adoption of linked data in the library resource description workflow.

The transition from human-readable records to knowledge graphs as a paradigm shift

The intellectual work undertaken by catalogers to describe resources in the current and new workflows has many similarities, even if the outputs look different. It is thus natural to ask: Is the transition from record creation to the management of knowledge graphs a paradigm shift? The short answer is a qualified yes. To formulate a more nuanced answer, the authors of this report discussed individual tasks in the metadata creation workflow and assigned them to three categories: new tasks; current tasks that will become obsolete; and the tasks, practices, and values that are still necessary.

New tasks

The most important new task is the materialization of a novel concept. Kalan Knudson Davis identifies the problem in her examination of the poster for an Everly Brothers concert in the Minnesota Memories collection. “As we catalog in this new environment,” she said, “we are forced to think through the answer to the question: *What entities matter to this object*?” Identifying “the entities that matter” will drive the construction of the future bibliographic universe. This imperative will perhaps supersede the traditional rationale for bibliographic description, which focuses on the “item in hand” and is informally referred to Cutter’s now-infamous “cult of the title page.”[109] In the linked data environment, the new imperative is to identify the relationships among the entities

relevant to the item of interest, which Davis characterizes as "a reconceptualization of the concept of bibliographic significance." This is the foundation of the view that the transition from records to knowledge graphs is a paradigm shift.

Tasks that become obsolete

Current tasks that will become obsolete cluster around the creation and normalization of natural language text strings in bibliographic and authority records. Some current practices now appear trivial and pointless. For example, some of the Passage co-authors admitted that the time they have spent on ISBD punctuation of text strings in MARC records was "cringe-worthy,"[110] and they welcomed a metadata creation workflow that makes this practice obsolete. But practices that enlist text strings to accomplish processing goals such as matching, linking, and disambiguation are perhaps more difficult to abolish from the current workflow. In a linked data environment, however, natural language strings are intended for human readers and do not need to be overloaded in the service of computational goals. Instead, machine-understandable identifiers do most of the work.

For example, Wikibase properties, defined with the same rigor as Wikibase items, establish machine-processable relationships between different versions of an object, such as a print publication and its digitized counterpart. One statement that results from this analysis links the digital image of the Everly Brothers poster with a paper copy. The first statement below is encoded with data available from the Passage dataset, and the second is a human-readable gloss assembled from English-language fingerprint data.

1. <https://reflections.mndigital.org/catalog/pch:57#/image/0><P235>< Q1225138>
2. The Everly Brothers at the Lakeside Ballroom, Glennwood, Minnesota—digital representation—Everly Brothers performance poster, July 27, 1965

In a MARC record, the same relationship is not guaranteed to be fully encoded because references to the subject or object are often ambiguous, and the relationship may be expressed in uncontrolled text.

But perhaps the most significant string-curation tasks that are now obsolete involve describing non-English language resources. For example, the transliteration of text from resources written in non-Latin scripts helps those who cannot understand or pronounce non-Latin scripts, but is not necessary for those who can. In a linked data environment such as Wikibase, transliteration is redundant because the data model is natively multilingual. Any need that remains for transliteration can be addressed more easily with generic tools such as Google Translate, which also supplies recorded pronunciations.

A corollary is that the MARC concept of "language of cataloging" is almost obsolete. Since a Wikibase entity consists mostly of structured data that is already associated with multilingual labels, the human editor need not specify a language of cataloging because the data is automatically available in any language supported by the Wikibase platform. Thus, when an English-speaking editor publishes the machine-understandable statement "Albert Einstein—occupation—physicist," the same statement is viewable in Spanish as "Albert Einstein—ocupación—físico," in Polish as "Albert—Einstein—occupation—fizyk," and so on. The user can see the different results simply by changing the language of display on the editing interface.

Nevertheless, a literal text string that can be interpreted as a language of cataloging must still be entered by the human editor as fingerprint data and in a few structured statements. This feature is confusing in some cases—as when an English speaker is confronted with the statement on the

English-language display that *Being and Time* is a translation of *Being and Time* (instead of the German *Sein und Zeit*). But an abundance of labels with linguistic identity is essential for the search experience, as the case of Sun Yat-sen showed.

This case also argued that identifiers make some tasks in the workflow for authority control obsolete. In a linked data environment, catalogers no longer need to establish a single preferred name string in each language of description as an unambiguous reference for the founding father of the Republic of China (such as Sun Yat-sen, 孫中山, or 孙中山). The goal of differentiation is accomplished with machine-understandable identifiers instead. As a result, identifiers serve in Wikibase as a form of "shorthand for the ensemble of metadata that is collected to describe and differentiate an identity."[111] Researchers are only beginning to explore the consequences of this observation.

What is still necessary

The category of tasks and practices that are still necessary is the most densely populated of the three. This outcome is consistent with the argument that linked data—or more broadly, a data architecture that reveals deeper semantic relationships to machine processes than is possible with current solutions—is a good fit with the domain of library resource description because core values are preserved. But the transformation from records to knowledge graphs introduces more sophisticated technological solutions to address existing requirements. Nevertheless, as Christine Fernsebner Eslao said,

> We observed that Wikibase required less training and practice than moving from one ILS (integrated library system)[112] to another. There's a conceptual shift involved, but the actual practices of description here felt like they evolved naturally out of current practice, with an emphasis on 'entification' and reuse beyond 'access points.'

Interpretive context

Interpretive *context is still needed* for most types of resources, especially for digital images and archival collections. As the Temple University team reported, the "depicts" property may differ from a traditional "subject" used in existing metadata in a digital library, but both may still be necessary. Fingerprint labels and structured statements about people, places, or things depicted in images may facilitate access and categorization, but they may not be fully successful in providing cultural and historical context. The Temple use case also showed that important structured statements in a knowledge graph may be many steps away from the entity for the primary item of interest. A mediated view or query must be designed to reveal this deeply buried relationship, which seems more complex and less trustworthy than a human-readable page.

Structured and narrative data

It is a truism that structured data facilitates machine understanding and interoperability. But the tools that make structured data easy to produce and render in a sophisticated discovery interface remain to be written. Perhaps this realization is also reflected in the configuration of the Wikimedia ecosystem. There, Wikidata is the designated place for a graph of structured data featuring factual, verified statements about items and properties. Wikipedia is a repository for human-readable exposition that provides additional context for understanding items of interest. And Wikimedia Commons contains structured data about digital media, including rights statements and technical metadata.

The division of labor can be illustrated by comparing the descriptions of Alfred Bierstadt's painting *The Rocky Mountains, Lander's Peak* in the three data stores. The Wikidata item entity[113] contains

structured statements about the painting and its current location in the Fogg Museum, as well as key biographical facts about the artist. Additional structured data about images of the painting is available from Wikimedia Commons,[114] such as file sizes and licensing information. Finally, the Wikipedia article contains a narrative about Alfred Bierstadt,[115] the Hudson River School painter who was born in Germany. In 1850, he had an artistic breakthrough on a trip to the Rocky Mountains, where he realized that he was experiencing "the best material for the artist in the world." These statements could, in principle, be rendered as structured data, but the exercise would probably require the definition of items and properties that are not otherwise useful.

The interplay between structured and narrative data was a persistent theme in the Passage use cases. The pilot participants concluded that while both are necessary for library resource description, the creation of structured data is the more urgent need. Nevertheless, Washburn and Mixter showed how the tripartite configuration of Wikipedia, Wikidata, and Wikimedia Commons can be mapped to library-community data stores such as CONTENTdm.[116] This thought experiment can be explored more deeply when the time is right to reimagine the role of narrative or expository data.

Best practices are still important

Every use case revealed that the act of creating structured data in the Wikibase required shaping. OCLC offered some guidance in the Passage help portal, but many recommendations for best practices also emerged from group discussion. Here are some examples:

- Each resource type should include statements containing key properties.
- Each structured statement should be associated with a link to a source, which provides input for calculations of authoritativeness or veracity.
- Use the terms defined in the Wikibase ontology. If an additional nuance is required for a particular task, propose an addition, but be conservative about it.
- Add fingerprint labels for real-world objects that are important for interpreting the item of interest. If these objects have not been defined, create Wikibase item entities for them.
- When describing a translated work, include an "also known as" fingerprint label in the original language and script. This convention would ensure that a result from an English-language search would reveal that *Being and Time* is a translation of *Sein und Zeit*.

The list of recommendations that could be promoted to best practices grew organically during the course of the pilot. But this list can be supplemented more deliberately by mapping the wisdom of current practice to the new workflow. For example, the concept of a uniform title serves as a normalized work title. In Passage, the entity's fingerprint data also serves to collate works and helps avoid the creation of duplicate entities. Another bibliographic concept that can be mapped to the new paradigm is the concept of "core," which can be used to define a scope for a description. Otherwise, the protean character of Passage runs the risk of creating elements that are bibliographically insignificant. Svenonius (2000) notes that the cost of making descriptions increases in proportion to the number of data elements they contain.[117] In the future, Lubetzky's question—"Is this rule necessary?"—can be rephrased as the question: "Is this [linked data] entity or relationship necessary?"[118]

Svenonius also notes that when bibliographic reality conflicts with existential reality, the real world wins out. Traditionally, it has been part of the cataloger's job to explain or correct ambiguous, unintelligible, and inaccurate statements through the use of notes or annotations in brackets. A Wikibase recommendation has a similar goal: "Wikibase will not be about the truth, but about

statements and their references."[119] This recommendation is often applied in cataloging practice through the decisions surrounding preferred names. Catalogers know that a preferred form of a name doesn't necessarily vary by language, but by the rules governing the National Bibliography that has created the bibliographic description. In Wikibase, the labels that appear when the language is set to a given value in the browser interface can be accompanied with statements of provenance. In the future, perhaps cataloging ethics will govern how and when such references and statements will be deprecated.

Once established, best practices can be translated into common-sense guidelines for subject-matter experts who may not be trained as librarians. As several use cases showed, the line between librarianship and scholarship is blurred in the production of a knowledge graph. If subject-matter experts step forward to offer even more detail, why not take advantage of the features in the Wikibase platform that empower crowd-sourcing? The outcome will be more effective, however, if the work is guided by an outline or template. In Passage, OCLC staff recommended best practices by acting primarily as mentors, writing help pages that suggested a list of key properties for each resource type. In a larger implementation, the guidelines would be established in an appropriate expert community.

Librarians will still uphold the values of authoritativeness and quality

A first encounter with the Wikimedia Foundation's content-creation culture leaves many with the impression that "anybody can say anything about anything," which appears to clash with the librarian's need to establish trust, truthfulness, authority, and quality. Of course, library "authority" files are a significant artifact of this vetting process. OCLC's engineers challenged this impression by introducing the participants to the Wikibase utilities for ensuring quality. They argued that these utilities address the same goals that librarians strive for and recommended that they be incorporated into the editing workflow. For example, OCLC recommended that each statement be annotated with a source, from which a Wikibase utility computes a reliability score that can break a tie when statements conflict. Statements can also be deprecated when there is evidence to do so.

In addition, Wikibase has an open-ended solution for incorporating identifiers into a single item entity, which invites the inference that a well-established real-world object is represented in many more datasets than a poorly documented one. Finally, the Wikibase audit trail provides human-readable evidence for gauging the trustworthiness of the data, which is implemented in the change-history log and the discussion pages. Experience with these utilities changed the initial perception of the Wiki editing culture for many pilot participants.

Reinventing crowd-sourcing

The Passage participants saw the potential of crowd-sourcing for enriching the knowledge graphs created in the Wikibase editing interface. This effort could be supported by the revision history and discussion pages that track every edit for a given Wikibase entity, each associated with a registered username and time stamp. In contrast, quality management is hampered in current resource description workflows by the fact that a MARC record can be marked only as "touched," with no written trace of what was changed by whom. And discussion takes place outside the editing environment, typically on professional listservs, where the connection to the affected content is lost. Additional effort is required to incorporate commentary from museum professionals, subject matter experts, and other interested third parties who do not usually participate in library-community listserv discussions.

Despite the robust implementation of editing history, however, the pilot participants raised concerns that crowd-sourcing in the Wikibase environment could still add unvetted information from unknown sources that would dilute the integrity of curated library data. The essential human-resources issue is that a self-selected crowd may have a range of skills and expertise, not all of which are suited to a given description task or use case. But some members of the crowd undoubtedly do have knowledge that complements or supplements that of library staff. For example, scholars who are familiar with non-English and non-Latin script materials could enrich the metadata created by librarians and archivists who lack this expertise. Even more productive relationships would be fostered if such community experts continue to interact with the staff and their patrons.

A more theoretical problem is that crowd-sourced terms may not be reflected in the controlled vocabularies managed by the library community, and may therefore lack the identifiers or URIs that maximize machine understanding of the data being created. For example, the crowd may contribute new terms that originate in social media but appear in the library community's controlled vocabularies only years later. Or the crowd-sourced terms may include acronyms that are not widely understood. But such language could make its first appearance in so-called LibGuides, or web pages used to share information and organize a library's discipline-specific resources.

To maximize the potential of crowd-sourcing while minimizing the risks, the report's co-authors identified traditional and new solutions.

An obvious starting point is a review of published research. For example, *Social Metadata for Libraries, Archives and Museums,*[120] an OCLC Research report, recommends that crowd-sourced contributions be solicited for a given purpose and with clear objectives that are governed by existing or new policies. Library staff should motivate volunteers and leverage their interest in contributing. Perhaps counterintuitively, the authors recommend that the library community should be relatively less concerned about spam or abuse because their experience suggested that most volunteers contribute in good faith. Finally, the report recommends the use of metrics that measure the usage of crowd-contributed content, a proxy of success.

Of course, another starting point is the functionality of Wikibase. How well do the utilities for scoring structured statements mentioned in the section on the Wikibase editing interface on page 15 perform when the contributor is not a librarian who is motivated enough to act on these values even without the aid of supporting technology? This question was beyond the scope of the Passage pilot and is a potentially rich topic for investigation. An attractive hypothesis is that the self-governing culture of Wikipedia and Wikidata is a model for adoption in the library resource description workflow because it demonstrates how high-quality work might be endorsed and enhanced by the crowd, while bad-faith or mediocre work can be ignored, downgraded, deprecated, or flagged for deletion.

Nevertheless, dispute resolution of crowd-sourced content in these applications is contentious because an editing culture that privileges authoritativeness and even objectivity may also propagate the biases of dominant social groups, leading to a failure to acknowledge alternative viewpoints. This issue arose in the case of the photograph of the gay student group at Temple University, which could not be fully populated with structured data because the Wikidata ontology lacks sufficient terminology for LGBTQ concepts. The problem of underrepresentation was also featured prominently in the program for WikiConference North America 2018 in sessions about the coverage of people with marginalized identities, gender equity, and right-of-center topics. [121]

If the Wikipedia/Wikidata model turns out to be inadequate for managing crowd-sourced data that addresses library-community use cases, a narrower set of issues about the shape, configuration, and oversight of third-party contributions might yield a more satisfactory solution. For example, to address the problem of uncontrolled terminology, advocates of library-community crowd-sourcing initiatives might recommend that terminology be restricted to Wikibase fingerprint data for already defined items. Such a decision would create a collection point for alternative labels and descriptions that already have identifiers.

Community contributions that are more textual than structured but exceed the scope of fingerprint data could be stored separately and treated as complementary additions that would be reviewed or curated by library metadata specialists. If so, the problem of crowd-sourced description could be construed as another argument for maintaining narrative or expository descriptions in a resource that accompanies structured data. On a larger scale, crowd-sourcing could be managed by governance structures or cooperation agreements. The PCC's Task Group on Identity Management in NACO is already investigating how the PCC could benefit from strategic partnerships and collaborations with Wikidata.[122]

But regardless of how crowd-sourcing is activated in the library community after linked data has been adopted, mature tools and applications that already populate the Wikibase software platform only expand the options for soliciting, accepting, evaluating, and managing the contributions.

Interoperability with Wikidata and other sources of linked data

At the beginning of the Passage pilot, OCLC's engineers made two decisions: to use the Wikibase platform that implements Wikidata, and to install a local instance of the software instead of contributing to Wikidata directly. The introduction of this report lists the arguments for using the Wikibase platform: the functionality reveals more of the benefits of linked data than do current library-community solutions; it is the foundation for mature applications that can be used even by those who do not have a deep technical knowledge of linked data; and it has many built-in features that support crowd-sourcing by a global and multilingual community of contributors. But the decision to create a dataset separate from Wikidata required more justification. At the outset, OCLC argued that a separate dataset was an expedient solution for the purposes of a short-lived pilot because a separate instance created a fully featured but private and configurable space to try out a wide range of ideas and minimize the consequences of failure.

Local Wikibase instances

As the Passage pilot advanced, however, more substantive arguments emerged that a separate Wikibase instance is also a reasonable long-term decision. For example, the library community has requirements and values that are separate from the Wikidata community for notability, authority, trustworthiness, and modeling. And the Wikibase editing interface did not fully meet the needs of the library resource-description workflow; the fingerprint data, in particular, presented challenges in nearly every use case. Conversely, some discussion in the Wikidata community implies that Wikidata may not have the capacity to serve as the library community's dedicated backbone because it is designed to represent all domains of human endeavor.[123]

Local Wikibase instances are also being installed elsewhere in the library community. For example, a local Wikibase instance is the technical backbone for the Mellon-funded digital humanities project Enslaved.org,[124] led by researchers from Michigan State University. Experiments with Wikibase are being planned by the French consortium ABES and the Bibliothèque nationale de France,[125] as well as the Deutsche Nationalbibliothek.[126] A related effort is the recently announced FactGrid project,[127] conducted by researchers at the Universität Erfurt, which uses a local Wikibase instance to collect factual data of interest to historians. In April 2019, the university announced a Memorandum of

Understanding[128] with the Deutsche Nationalbibliothek to base the FactGrid on the GND,[129] the integrated authority file managed cooperatively by the Deutsche Nationalbibliothek, German-speaking library networks, and associated libraries.

If this trend accelerates, library resource descriptions will continue to be represented in separate datasets even after the transition from human-readable database records to knowledge graphs has been implemented. With this in mind, it is not premature to think about how the new datasets will interoperate. In the Passage use cases, entities imported from Wikidata for *Sein und Zeit*, Sun Yat-sen, and *Nuper rosarum flores* were improved. Should they be exported back to Wikidata? Or should the corresponding Wikidata entities be enhanced only with the URIs that resolve to these library-community entities, following the examples established by FAST, VIAF, and id.loc.gov?

Direct collaboration with Wikidata

The above questions could also be interpreted as issues that are more about governance and reputation management than the logistics of data management. If so, it is relevant to note the special status of Wikidata URIs, which are already recognized in library-community linked data projects as canonical references to Albert Einstein, New York City, and many other real-world objects in the bibliographic universe. This status is codified in the proposal[130] approved by the MARC Advisory Committee in 2017 that defines the $1 subfield as a placeholder for URIs that resolve to RDF-encoded statements about real-world objects. In the examples discussed in the proposal and in follow-up documentation,[131] the $1 MARC subfield is often populated with Wikidata URIs.

Another endorsement of Wikidata is expressed in the PCC's Strategic Directions for 2018-2021,[132] which includes a directive to "accelerate the movement toward ubiquitous identifier creation and identity management at the network level." This statement supplements the PCC-sponsored Linked Data Advisory Committee's 2017 white paper, *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies*,[133] which listed Wikidata among the "hub datasets" playing an "important role" in making connections among identifiers on the semantic web. Wikidata was also named by the OCLC Research Library Partners Metadata Managers Focus Group[134] as an option for minting identifiers for names not already represented in library authority files.[135]

In 2018, the Andrew W. Mellon Foundation funded the project *Linked Data for Production: Pathway to Implementation,* or LD4P Phase 2, which, in collaboration with the PCC, plans to develop "a cloud-based sandbox editing environment in support of an expanded cohort of libraries to create and reuse linked data" and integrate "library metadata and identifiers with the Web through collaboration with Wikidata."[136]

Taken together, these trends suggest that the relationship between Wikidata and local installations such as Project Passage is one of the most significant questions deserving follow-up investigation. Some hint at a solution that recommends registration in Wikidata for the most notable people, places, organizations, and events mentioned in library data, with local management of other resources. But an alternative is a closer collaboration and a merging of concerns in the library and Wikidata and Wikimedia communities. Arguments supporting both positions are listed in the ARL white paper mentioned earlier.[137] Much more work is required to identify the most productive relationships to advance the goal of creating more and better resource descriptions that benefit both groups.

From a technical perspective, this discussion reignites the classic debate about centralized vs. federated data repositories. This topic was explored in the IMLS-funded National Strategy for Shareable Local Name Authorities National Forum, and is documented in the 2018 report and

reference model.[138] Regardless of the outcome, however, the datasets being developed will likely have an architecture that is compatible with linked data principles, which offers the promise of a deeper interoperability with data published both inside and outside the library community than is possible with current library standards.

LOOKING FORWARD

In April 2019, as this report was going to press, the surge of interest in Wikidata by the library community mentioned in the section on "The pilot study participants" on page 13 was becoming more evident. For example, the PCC was planning to host the seminar Intro to Wikidata[139] at the Library of Congress, with support from the National Library of Medicine, several Wikimedia communities in the United States, and Wikimedians in Residence at the University of Virginia and the National Institute of Occupational Safety and Health. The goal of the workshop is to guide members of the library community through the process of creating Wikidata entities for persons or corporate bodies, and to reflect on how the result compares with existing library-community modes of description.

On 23 April 2019, Hilary K. Thorsen, the Wikimedian in Residence for the Linked Data for Production project,[140] conducted the inaugural meeting of the LD4 Wikidata Affinity Group.[141] The Wikimedian will facilitate contributions of item entities to Wikidata and coordinate with the Wikidata community on issues involving notability, maturation of the ontology, and tools development. Many of the Passage pilot participants from the library community are engaged in these activities.

OCLC researchers are also playing leadership roles to foster closer ties between library and Wikimedia communities. For example, the project reported in Klein and Kyrios (2013)[142] automatically produced hundreds of thousands of links between Wikipedia and VIAF. The authors argued that this result "represents an opportunity for libraries to present their traditionally siloed data, such as catalog and authority records, in more openly accessible web platforms." More recently, the OCLC Research program Libraries Leverage Wikimedia[143] has received grant funding from the Wikimedia Foundation to produce a training curriculum for libraries. In 2016, OCLC won the John S. and James L. Knight Foundation's 2016 News Challenge for Libraries award[144] with the project Wikipedia + Libraries: Better Together. And in June 2018, OCLC researchers worked with Andrew Lih and Robert Fernandez of the Wikimedia District of Columbia to produce the webinar Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond.[145]

As a follow up to the Passage pilot, OCLC is continuing to explore the technical implications of library resource description using Wikibase. Research staff have given presentations on their experience with installing and deploying Wikibase,[146] and more knowledge-sharing events are being planned.

Resource description projects inspired by Passage are also underway at OCLC, all with collaborators in the library community. For example:

- Relationships are among structured, narrative, and image data inspired by the interplay of Wikidata, Wikipedia, and Wikimedia Commons are being explored in CONTENTdm®, which was enhanced with support for the International Image Interoperability Framework (IIIF) in 2018.[147] IIIF is an industry-standard set of specifications for application program interfaces, or APIs, designed to enable a richer and more contextualized experience of digital images. IIIF is fully compatible with the machine-understandable semantic data that is the subject of this report.
- A community-developed ontology of archives[148] proposed as an extension of Schema.org has been harmonized with the Wikidata ontology and is being implemented in ArchiveGrid. Several of the extensions have been promoted to "pending" in Schema.org (Issues 1758 and 1759).[149]

- Studies of model development that ease the transition from the current generation of library-community data resources to an entity/relationship-based architecture are being conducted. For example, Wikidata and other Wikibase URIs can be added to MARC records through the $1 subfield, according to a proposal approved by the MARC Advisory Committee in 2017.[150] This proposal was developed with input from OCLC research and product management staff.[151]

The Passage pilot represented an opportunity for all participants to gain hands-on experience creating structured data that could be exported as linked data. Among the outcomes were hundreds of new Wikibase item entities and new tools for creating and viewing results. The experience also produced knowledge and wisdom that could be shared, as well as a multitude of concrete ideas that are already giving shape to follow-up investigations. The results of this effort will help materialize the paradigm shift that is evoked by the name of the pilot. The shared goal is a "passage" from standards introduced in the 1960s to a 21st-century solution featuring structured semantic data that promises better connections between libraries and the world beyond.

ACKNOWLEDGMENTS

OCLC Research wishes to thank *all* the Project Passage participants who contributed to the pilot and shared their experiences, in addition to the co-authors of this report. We particularly appreciate those flagged by an asterisk in the list below for their special efforts or who served as funnels aggregating their respective teams' experiences.

- **American University:** Dominque Bortmas, Matthew Bright, Christine Dulaney, Alayne Mundt
- **Brigham Young University:** Greg Reeve, Elizabeth Tobiasson
- **Cleveland Public Library:** Chatham Ewing,* Arnia Wheatley*
- **Cornell University:** Jason Kovari,* Xin Li, Sarah Ross*
- **Harvard University Library:** Michelle Durocher, Tom Ma, Chew Chiat Naun,* Rebecca O'Brien, Minna Popkin, Isabel Quintana, Elise Ramsey, Robin Wendler, Anthony Terrizzi
- **Michigan State University:** Autumn Faulkner, Devin Higgin, Ranti Junus, Tim Kiser, Elisa Landaverde, Lisa Lorenzo, Lucas Wing Kau Mak, Lisa Robinson, Nicole Smeltekop
- **National Library of Medicine:** Diane Boehr, Barbara Bushman,* Phyllis Chui, Nancy Fallgren,* Tina Shrader, Alvin Stockdale,* Sharon Willis
- **North Carolina State University:** Jacob Shelby, Lynn Whittenberger
- **Northwestern University:** Michael Babinec,* Paul Burley,* Catherine Grove, Marcia Tiede, Evviva Weinraub
- **Princeton University:** Jeff Barton, Jennifer Baxmeyer, Joyce Bell,* Minjie Chen, Shuwen Cao,* Sarah Hamerman, Regine Heberlein, Brian Newhouse, Rachel Simon, Pavel Teptyuk, Don Thornbury, Luiza Wainer, Chiharu Watsky
- **Smithsonian:** Eric Bergstrom, Sue Graves, Martin Kalfatovic, Polly Khater, Bess Missell, Lesley Parilla, Suzanne Pilsk
- **Temple University:** Rachel Appel, Jasmine Clark, Matt Ducmanas, Leanne Finnigan
- **University of California, Davis:** Jared Campbell, April Chen, Dale Snapp, Carl Stahmer,* Khampheng Vongkhamchanh
- **University of Minnesota:** Christine DeZelar-Tiedman, Stephen Hearn,* Kelly Thompson, Stacie Traill
- **University of New Hampshire:** Elise Daniel, Sarah Theimer
- **Yale University:** Eva Bolkovac, Dominique Bourassa, Danijela Matkovic, Youn Noh, Rick Sarcia, Yukari Sugiyama, Tim Thompson

OCLC: John Chapman, Sara DeSmidt, Melissa Hess, Tod Matola, Jeff Mixter, Andrew K. Pace, Stephan Schindehette, Taylor Surface, Bruce Washburn, Jeff Young. Thank you to Jeanette McNicol for the excellent design of this report and to Erin M. Schadt for her editing.

NOTES

1. Berners-Lee, Tim. 2009. "Linked Data." Design Issues. W3C. Updated 18 June 2009. https://www.w3.org/DesignIssues/LinkedData.html.

2. Library of Congress. "Bibliographic Framework Initiative" BIBFRAME. Accessed 9 July, 2019. https://www.loc.gov/bibframe/.

3. Library of Congress. "*PCC (Program for Cooperative Cataloging) Strategic Directions: January 2018-December 2021.*" Cataloging and Acquisitions. Updated 24 January 2019. https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf.

4. Linked Data. n.d. "Linked Data Survey." OCLC Research (website). Accessed 10 July 2019. https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-survey.html.

5. Smith-Yoshimura, Karen. 2018. "Analysis of 2018 International Linked Data Survey for Implementers." *Code4lib Journal* no. 42 (November). https://journal.code4lib.org/articles/13867.

6. Smith-Yoshimura, Karen. 2018. "Results of Linked Data Surveys for Implementers (2014-2018)." OCLC Research. https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-survey.html.

7. Reitz, Joan M. "ODLIS: Online Dictionary for Library and Information Science." ABC-CLIO (website). Accessed 2 October 2018. https://www.abc-clio.com/ODLIS/odlis_c.aspx.

8. Murray, Ronald J., and Barbara B. Tillett. 2011. "Cataloging Theory in Search of Graph Theory and Other Ivory Towers: Object: Cultural Heritage Resource Description Networks." *Information Technology and Libraries* 30, no. 4 (December): 170+.

9. Wikipedia. "Wikimedia Movement." Updated 22 June 2019, at 07:20 (UTC). https://en.wikipedia.org/wiki/Wikimedia_movement.

10. Wikipedia: https://www.wikipedia.org/;

    Wikipedia. "MediaWiki." Updated 8 July 2019, at 01:52 (UTC). https://en.wikipedia.org/wiki/MediaWiki;

    Wikipedia. "Wikidata." Updated 10 July 2019, at 18:47 (UTC). https://en.wikipedia.org/wiki/Wikidata;

    Wikipedia. "Wikibase." Updated 12 May 2019, at 22:36 (UTC). https://en.wikipedia.org/wiki/Wikibase.

11. Wikidata. "Wikidata:Notability." 1 July 2019, at 13:38. https://www.wikidata.org/wiki/Wikidata:Notability.

12. Association of Research Libraries Task Force. 2019. *ARL White Paper on Wikidata Opportunities and Recommendations*,38. Washington, DC: Association of Research Libraries. https://upload.wikimedia.org/wikipedia/commons/7/70/ARL_White_Paper_on_Wikidata_Opportunities_and_Recommendations.pdf.

13. OCLC Research. "OCLC to Launch Linked Data Pilot with Seven Leading Libraries." News Release, no. OCL-010 2015, 11 September 2018. https://cdm15003.contentdm.oclc.org/digital/collection/p15003coll6/id/382/rec/169.

14. OCLC. "OCLC Community Center." Accessed 10 July 2019. https://www.oclc.org/community/home.en.html.

15. Slide decks for presentations involving OCLC authors are accessible from the OCLC Linked Data Wikibase Prototype project page: https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-prototype.html.

16. Duraspace. "Linked Data for Production (LD4P)." Updated 3 April 2019. https://wiki.duraspace.org/pages/viewpage.action?pageId=74515029.

17. Duraspace. "Linked Data for Production: Pathway to Implementation (LD4P2)." Updated 1 July 2019. https://wiki.duraspace.org/display/LD4P2.

18. Smith-Yoshimura, Karen. 2018. "The Rise of Wikidata as a Linked Data Source." *HangingTogether* (blog). Posted 6 August 2018. http://hangingtogether.org/?p=6775;

    Smith-Yoshimura, Karen. 2018. "Analysis of 2018 International Linked Data Survey for Implementers." *code4lib Journal* no. 42 (November). https://journal.code4lib.org/articles/13867.

19. Wikiscan. "Statistics on www.wikidata.org." http://wikidata.wikiscan.org/.

20. Wikidata. "Wikidata:Bots." Updated 4 May 2019, at 14:48. https://www.wikidata.org/wiki/Wikidata:Bots.

21. Wikidata. "Wikidata:Glossary." Updated 17 April 2019, at 12:07. https://www.wikidata.org/wiki/Wikidata:Glossary;

    Wikidata. "Help:Items." Updated 31 May 2019, at 12:20. https://www.wikidata.org/wiki/Help:Items.

22. Wikimedia Commons. "File:Datamodel in Wikidata.svg" Updated 4 July 2019, at 18:37. https://commons.wikimedia.org/wiki/File:Datamodel_in_Wikidata.svg.

23. Wikidata. "Wikidata:Glossary, Entity." Updated 17 April 2019, at 12:07. https://www.wikidata.org/wiki/Wikidata:Glossary#Entity.

24. MediaWiki. "Wikibase/DataModel/Primer." Updated 1 August 2018, at 14:02. https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer.

25. Wikidata. "Help:Label." Updated 25 January 2019, at 08:59. https://www.wikidata.org/wiki/Help:Label.

26. Wikidata. "Help:Description." Updated 18 June 2019, at 09:07. https://www.wikidata.org/wiki/Help:Description.

27. Wikidata. "Help:Aliases." Updated 19 September 2018, at 15:28. https://www.wikidata.org/wiki/Help:Aliases.

28. Wikidata. "Help:Items." Updated 31 May 2019, at 12:20. https://www.wikidata.org/wiki/Help:Items.

29. Wikidata. "Wikidata:Glossary, Property." Updated 17 April 2019, at 12:07. https://www.wikidata.org/wiki/Wikidata:Glossary#Property;

    Wikidata. "Help:Properties." Updated 10 July 2019, at 21:30. https://www.wikidata.org/wiki/Help:Properties.

30. Wikidata. "Wikidata:Glossary, Qualifier." Updated 17 April 2019, at 12:07. https://www.wikidata.org/wiki/Wikidata:Glossary#Qualifier;

    Wikidata. "Help:Qualifiers." Updated: 3 January 2018, at 22:19. https://www.wikidata.org/wiki/Help:Qualifiers.

31. Wikidata. "Help:Ranking." Updated 27 December 2018, at 16:29. https://www.wikidata.org/wiki/Help:Ranking.

32. MediaWiki. "Wikibase/DataModel/Primer, Ranks." Updated 1 August 2018, at 14:02. https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer#Ranks.

33. Wikidata. "Wikidata:Property Proposal." Updated 17 April 2019, at 07:31. https://www.wikidata.org/wiki/Wikidata:Property_proposal.

34. Berners-Lee, Tim. 2009. "Linked Data." Design Issues. W3C. Updated 18 June 2009. https://www.w3.org/DesignIssues/LinkedData.html.

35. Wikipedia. "Uniform Resource Identifier." Updated 24 June 2019, at 16:13 (UTC). https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.

36. Wikipedia. "Triplestore." Updated 17 February 2019, at 15:06 (UTC). https://en.wikipedia.org/wiki/Triplestore.

37. Wikipedia. "SPARQL." Updated 26 May 2019, at 17:44 (UTC). https://en.wikipedia.org/wiki/SPARQL.

38. Wikidata. "Help:Contents." Updated 24 June 2019, at 08:03. https://www.wikidata.org/wiki/Help:Contents.

39. Wikidata. "Wikidata:Tours." Updated 7 April 2018, at 19:37. https://www.wikidata.org/wiki/Wikidata:Tours.

40. Association of Research Libraries Task Force. 2019. *ARL White Paper on Wikidata Opportunities and Recommendations*,38. Washington, DC: Association of Research Libraries. https://upload.wikimedia.org/wikipedia/commons/7/70/ARL_White_Paper_on_Wikidata_Opportunities_and_Recommendations.pdf.

41. Lih, Andrew, and Robert Fernandez. 2018. "Works in Progress Webinar: Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond." Produced by OCLC Research, 12 June 2018. MP4 video presentation, 1:1:51. https://www.oclc.org/research/events/2018/06-12.html.

42. OpenRefine. "Welcome." Accessed 10 July 2019. http://openrefine.org/.

43. For readers in the library community who wish to learn more, Library Carpentry OpenRefine is a useful tutorial. See "Introduction to OpenRefine." Accessed 15 August 2019. https://librarycarpentry.org/lc-open-refine/01-introduction/index.html.

44. MediaWiki. "Manual:Pywikibot." Updated 13 June 2019, at 11:01. https://www.mediawiki.org/wiki/Manual:Pywikibot.

45. Project Passage. "Help:Retriever." Updated 4 September 2018. (Archived 27 Feb 2019.) https://web.archive.org/web/20190227194209/http:/www.projectpassage.org/wiki/Help:Retriever.

46. Library of Congress. "MARC 21 Format for Authority Data: Table of Contents." Network Development and MARC Standards Office. Updated 15 May 2019. https://www.loc.gov/marc/authority/.

47. Wikidata Query Service. "#Number of humans in Wikidata" https://query.wikidata.org/#%23Number%20of%20humans%20in%20Wikidata%0ASELECT%20%28COUNT%28%3Fitem%29%20AS%20%3Fcount%29%0AWHERE%20%7B%0A%09%3Fitem%20wdt%3AP31%2Fwdt%3AP279%2a%20wd%3AQ5%20.%0A%7D.

48. Project Passage SPARQL UI. "Query: Books Authored by Librarians." n.d. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-query-books-authored-by-librarians.png.

49. Project Passage SPARQL UI. "Query: Authors with the most Translations." 12 March 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-query-authors-with-most-translations.png.

50. Project Passage SPARQL UI. "Query: Sein und Zeit Translations Timeline." n.d. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-query-sein-und-zeit-translations-timeline.png.

51. Project Passage SPARQL UI. "Query: People Born 1800-1880 without Specified Death Date." (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-query-people-born-1800-1880-without-specified-death-date.png.

52. OpenRefine. "Welcome." Accessed 10 July 2019. http://openrefine.org/.

53. Project Passage. "List of Properties." n.d. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-list-of-properties.png.

54. Project Passage Wiki library resources pages (Archived 27 February 2019):

- Guidance for Creating and Editing: https://web.archive.org/web/20190227194215/http:/www.projectpassage.org/wiki/Help:Guidance_for_Creating_and_Editing_in_Project_Passage;
- The Data Model: https://web.archive.org/web/20190227194211/http:/www.projectpassage.org/wiki/Help:Data_Model;
- Using Identifiers in Creative Work Items: https://web.archive.org/web/20190227194246/http:/www.projectpassage.org/wiki/Help:Using_Identifiers_in_Creative_Work_Items;
- Adding Labels, Descriptions, and Aliases in Multiple Languages: https://web.archive.org/web/20190227194221/http:/www.projectpassage.org/wiki/Help:Adding_Labels,_Descriptions,_and_Aliases_in_Multiple_Languages;
- Adding Dates in Passage: https://web.archive.org/web/20190227194227/http:/www.projectpassage.org/wiki/Help:Dates;
- Creative Works and Translations: https://web.archive.org/web/20190227194236/http:/www.projectpassage.org/wiki/Help:Creative_Works_and_Translations;
- Events: https://web.archive.org/web/20190227194238/http:/www.projectpassage.org/wiki/Help:Events;
- The Retriever: https://web.archive.org/web/20190227194209/http:/www.projectpassage.org/wiki/Help:Retriever;
- OpenRefine Reconciliation API: https://web.archive.org/web/20190227194201/http:/www.projectpassage.org/wiki/Help:OpenRefine_Reconciliation_API.

55. Wikidata. "Wikidata:Glossary." Updated 17 April 2019, at 12:07. https://www.wikidata.org/wiki/Wikidata:Glossary https://www.wikidata.org/wiki/Wikidata:Glossary.

56. Project Passage. "Help:Guidance for Creating and Editing in Project Passage." Updated 29 May 2018. (Archived 27 February 2019). https://web.archive.org/web/20190227194215/http://www.projectpassage.org/wiki/Help:Guidance_for_Creating_and_Editing_in_Project_Passage.

57. Godby, Jean, and Diane Vizine-Goetz, 2016. "BIBFRAME and OCLC Works: Defining Models and Discovering Evidence." (PowerPoint presentation, ALA Annual Conference, Chicago, Illinois, 26 June 2017.) https://www.oclc.org/content/dam/research/presentations/godby/godby-vizinegoetz-ala2017.pptx;

Godby, Carol Jean, Shenghui Wang, and Jeffrey K. Mixter. 2015. "Modeling and Discovering Creative Works." Chap. 3 in *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*. Dublin, OH: OCLC Research. https://www.oclc.org/research/publications/2015/oclcresearch-library-linked-data-in-the-cloud.html.

58. MediaWiki. "Manual:Creating a Bot." Updated 29 May 2019, at 06:19. https://www.mediawiki.org/wiki/Manual:Creating_a_bot.

59. Wikidata. "Douglas Adams (Q42)." Updated 12 July 2019, at 02:30. https://www.wikidata.org/wiki/Q42.

60. Xiaoli Li. "Making Linked Data: Experience with OCLC's Project Passage." In: Making Linked Data – Catalogers' Views from the Front Edge. Streamed live by IGeLU and ELUNA Show&Tell, 29 August 2018. Youtube video 1:07:25. https://www.youtube.com/watch?v=6nhjuc05E4c. [Li's presentation begins 2:24.]

61. Smith-Yoshimura. 2013. "Multilingual WorldCat Represented by Translations." *HangingTogether* (blog), 12 November 2013. http://hangingtogether.org/?p=3426.

62. Smith-Yoshimura, Karen. 2017. Using the Semantic Web to Improve Knowledge of Translations (PowerPoint presentation, DCMI International Conference on Dublin Core and Metadata Applications 2017, Washington, DC, 26 October 2017. http://dcevents.dublincore.org/IntConf/dc-2017/paper/download/496/609.

63. Project Passage. "Help:Creative Works and Translations." Updated 29 May 2018. (Archived 27 February 2019). https://web.archive.org/web/20190227194236/http:/www.projectpassage.org/wiki/Help:Creative_Works_and_Translations.

64. Wikipedia. "Being and Time." Updated 12 July 2019, at 09:16 (UTC). https://en.wikipedia.org/wiki/Being_and_Time.

65. OCLC Research. "CATViS: Visual Analytics for the World's Library Data." Updated 21 June 2018. https://www.oclc.org/research/themes/data-science/catvis.html.

66. "Welcome to NWO." Accessed 12 July 2019. https://www.nwo.nl/en.

67. Smith-Yoshimura, Karen Sandra. 2017. "Using the Semantic Web to Improve Knowledge of Translations." *International Conference on Dublin Core and Metadata Applications* (2 December 2017): 27. http://dcpapers.dublincore.org/pubs/article/view/3866/2051.

68. Project Passage. "Help:Creative Works and Translations." Updated 29 May 2018. (Archived 27 February 2019. https://web.archive.org/web/20190227194221/http:/www.projectpassage.org/wiki/Help:Adding_Labels,_Descriptions,_and_Aliases_in_Multiple_Languages.

69. Bauman Rare Books. "Sein und Zeit," Martin Heidegger, First Edition, Signed. n.d. Accessed 13 July, 2019. https://www.baumanrarebooks.com/rare-books/heidegger-martin-allers-rudolph/sein-und-zeit/85260.aspx.

70. There is also a built-in Wikibase UI feature that does something similar. In the left-hand sidebar of the display, in the "Tools" section, there is a "What links here" option that returns a list of other Wikibase entities that are linked to the current one. For example, for the Being and Time work entity, that link returns a list of the translations. See Project Passage. "Pages that Link to 'Item:Q1024094." n.d. (Screenshot). http://OCLC.org/content/dam/research/images/publications/passage/passage-query-being-and-time-what-links-here.png.

71. Project Passage. "Being and Time: Modern Greek Translation of Heldegger's Sein und Zeit (translater: Tzavaras)." Updated 5 September 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-sein-und-zeit-greek.png.

72. Project Passage. "Being and Time: Persian Translation of Heldegger's Sein und Zeit (translater: Ǧamādī)." Updated 1 May 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-sein-und-zeit-persian.png.

73. Project Passage. "Help:Guidance for Creating and Editing in Project Passage." Updated 31 August 2018. (Archived 27 February 2019). https://web.archive.org/web/20190227194215/http://www.projectpassage.org/wiki/Help:Guidance_for_Creating_and_Editing_in_Project_Passage.

74 Harvard Library. "Harvard Map Collection." Accessed 13 July 2019. https://library.harvard.edu/libraries/harvard-map-collection.

75. Wikidata. "Concord (Q523240)." Updated 24 June 2019, at 20:44. https://www.wikidata.org/wiki/Q523240;

Wikidata. "Middlesex County (Q54073)." Updated 4 July 2019, at 20:13. https://www.wikidata.org/wiki/Q54073;

Wikidata. "Henry Francis Walling (Q18911728)" Updated 21 June 2019, at 13:58. https://www.wikidata.org/wiki/Q18911728.

76. Leaflet: An Open-source Javascript Library for Mobile-Friendly Interactive Maps. "Plugins." https://leafletjs.com/plugins.html.

77. Project Passage. "Henry Francis Walling." Explorer Item Q214001. n.d. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-henry-francis-walling-explorer.png.

78. "Minnesota Reflections." Accessed 13 July 2019. https://reflections.mndigital.org/.

79. NIH; US. National Library of Medicine. "Pictures of Nursing: The Zwerdling Postcard Collection." Announcements, News. Accessed 13 July 2019. https://www.nlm.nih.gov/hmd/about/exhibition/picturesofnursing.html.

80. Temple University Libraries. "Reverend Dr. Martin Luther King Jr. And Cecil B. Moore Attend a Protest Rally at Girard College." Temple Digital Collections. https://digital.library.temple.edu/digital/collection/p15037coll17/id/36%20%5bdigital%20representation%5d.

81. Wikipedia. "Maria Josepha of Saxony." Updated 26 February 2013, at 12:29 (UTC). https://en.wikipedia.org/wiki/Maria_Josepha_of_Saxony.

82. Project Passage. "Lakeside Ballroom, Glenwood, Minnesota." Updated 7 August 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-lakeside-ballroom.png;

Project Passage. "Everly Brothers Performance, Lakeside Ballroom, Glenwood, Minn., July 27, 1965." Updated 18 July 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-everly-brothers-concert.png;

Project Passage. "Glenwood, Minnesota." Updated 17 July 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-glenwood-minnesota.png.

83. Project Passage. "Nurse: Type of Health Care Provider." Updated 4 September 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-nurse.png;

Project Passage. "Michael Zwerdling: Nurse and Author of Postcards of Nursing." Updated 11 July 2018, at 16:09. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-michael-zwerdling.png;

Project Passage. "Postcards of Nursing, a Worldwide Tribute: Postcards Depicting Nursing Between 1893 and 2002 ." Updated 17 June 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-postcards-nursing.png.

84. Project Passage. "Civil Rights Demonstration at Girard College, August 3, 1965." Updated 1 May 2019. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-civil-rights-demonstration-girard-college.png;

Project Passage. "Charles L. Blockson Afro-American Collection." Updated 22 August 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-charles-l-blockson-afro-american-collection.png;

Project Passage. "John W. Mosley Photographs." Updated 22 August 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-john-w-mosley-photographs.png;

Project Passage. "John W. Mosley." Updated 21 June 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-john-w-mosley.png.

85. Project Passage. "Help:Events." Updated 31 August 2018. (Archived 27 February 2019). https://web.archive.org/web/20190227194238/http:/www.projectpassage.org/wiki/Help:Events.

86. Project Passage. "World War I." Updated 16 May 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-world-war-one.png.

87. Library of Congress. "Discussion Paper List." Especially for Librarians and Archivists. MARC Development, MARC Standards. 3 March 2019.https://www.loc.gov/marc/mac/2016/2016-dp09.html.

88. Temple University Libraries. "Gay students sponser kissing booth." Temple Digital Collections. 14 February 1974. https://digital.library.temple.edu/digital/collection/p15037coll3/id/67369.

89. WorldCat. "Ces anges des champs de bataille." Accessed 14 July 2019. https://www.worldcat.org/title/ces-anges-des-champs-de-bataille/oclc/993651639.

90. NIH; National Library of Medicine. "Ces Anges Des Champs de Bataille." Digital Collections. Accessed 14 July 2019. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101707419-img.

91. Temple University Libraries. "Charles L. Blockson Afro-American Collection." Accessed 14 July 2019. https://library.temple.edu/collections/blockson.

92. Temple University Libraries. "John W. Mosley Photographs." Temple Digital Collections. Accessed 14 July 2019. https://digital.library.temple.edu/digital/collection/p15037coll17.

93. "Describing Archives: A Content Standard, Second Edition (DACS)." Accessed 14 July 2019. https://www2.archivists.org/book/export/html/17474.

94. Harvard University Archives; Hollis. "Records of the Harvard Anti-Wine Society and of the Harvard Total Abstinence Society, 1836-1843." Accessed 14 July 2019. https://hollis.harvard.edu/primo-explore/fulldisplay?context=L&vid=HVD2&search_scope=everything&tab=everything&lang=en_US&docid=01HVD_ALMA21220161969000394l.

95. Project Passage. "Harvard Total Abstinence Society." Updated 20 August 2018. (Screenshot) https://oclc.org/content/dam/research/images/publications/passage/passage-harvard-total-abstinence-society.png;

Project Passage. "Harvard Anti-Wine Society." Updated 20 August 2018. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-harvard-anti-wine-society.png.

96. Library of Congress Authorities. "Harvard Total Abstinence Society." LCCN Permalink No2019025967." n.d. Webpage. Accessed 14 July 2019. https://lccn.loc.gov/no2019025967;

Library of Congress Authorities. "Harvard Anti-Wine Society." LCCN Permalink No2019026126." n.d. Webpage. Accessed 14 July 201. https://lccn.loc.gov/no2019026126.

97. Wikidata. "Consecration (Q125375)." Updated 12 May 2019, at 08:17. https://www.wikidata.org/wiki/Q125375;

Wikidata. "Dedication (Q1762010). Updated 12 March 2019, at 17:39. https://www.wikidata.org/wiki/Q1762010.

98. Wikidata. "Nuper rosarum flores (Q589115)." Updated 14 May 2019, at 08:17. https://www.wikidata.org/wiki/Q589115;

Library of Congress. "Dufay, Guillaume, 1397-1474. Nuper rosarum flores/Terribilis est locus iste." LC Linked Data Service: Authorities and Vocabularies. Webpage. Accessed 14 July 2019. http://id.loc.gov/authorities/names/no98013583.html;

VIAF (Virtual International Authority File. "Dufay, Guillaume, 1397-1474. Nuper rosarum flores." VIAF ID: 185706616 (Work). Accessed 14 July 2019. https://viaf.org/viaf/185706616.

99. Project Passage. "Consecration of Florence Cathedral." Explorer Item Q1225580. n.d. Accessed 14 July 2019. (Screenshot). https://oclc.org/content/dam/research/images/publications/passage/passage-consecration-florence-cathedral.png.

100. Library of Congress. "Library of Congress Genre/Form Terms PDF Files." Cataloging, Acquisitions. Updated 1 May 2019. https://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html.

101. Library of Congress. 2016. "MARC 21 Authority." Headings - General Information. Updated 17 November 2016. https://www.loc.gov/marc/authority/ad1xx3xx.html.

102. Pace, Andrew, and Holly Tomren. 2018. "From Prototype to Production: Turning Good Ideas into Useful Library Services" (PowerPoint presentation, CNI Fall 2018 Membership Meeting, Project Briefings, Washington, DC, 10 December 2018). https://www.oclc.org/content/dam/research/presentations/2018/pace-tomren-cni-prototype-to-production-20181210.pptx.

103. Eslao, Christine Fernsebner, Honor Moody, and Sara Newell. 2018. "Creating a Linked Data Library Partnership: Using Wikibase." (PowerPoint presentation). https://www.oclc.org/content/dam/research/presentations/2018/2018-eslao-linked-data-lib-partnership-wikibase.pdf.

104. IFLA. 1989. *Functional Requirements for Bibliographic Records: Final Report*. UBCIM Publications, n.s., vol. 19. Edited by Marie-France Plassard. Munich: Saur. https://www.ifla.org/files/assets/cataloguing/frbr/frbr.pdf.

105. The Library of Congress. 2019. "Resource Description and Access (RDA): Information and Resources in Preparation for RDA." Updated 5 June 2019. https://www.loc.gov/aba/rda/.

106. Riva, Pat, Patrick Le Bœuf, and Maja Žumer. 2017. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. Den Haag, Netherlands: IFLA. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf.

107. The Library of Congress. 2019. "Extended Date/Time Format (EDTF) Specification." Standards. Updated 4 February 2019. https://www.loc.gov/standards/datetime/edtf.html.

108. Casalini, Michele, Chiat Naun Chew, Chad Cluff, Michelle Durocher, Steven Folsom, Paul Frank, Janifer Gatenby, Jean Godby, Jason Kovari, Nancy Lorimer, Clifford Lynch, Peter Murray, Jeremy Myntti, Anna Neatrour, Cory Nimer, Suzanne Pilsk, Daniel Pitti, Isabel Quintana, Jing Wang, and Simeon Warner. 2018. *National Strategy for Shareable Local Name Authorities National Forum: White Paper*. Ithaka, New York: Cornell University Library *eCommons digital reposity*. https://hdl.handle.net/1813/56343.

109. Google Books has scanned the 1904 edition, in which page 24 says "Bibliographers have established a cult of the title-page." Cutter, Charles A. 1904. Rules for a Dictionary Catalog, 24. London: Library Association. https://archive.org/details/rulesforadictio06cuttgoog.

110. IFLA. 2010. "International Standard Bibliographic Description (ISBD): Consolidated Edition." Draft as of 2010-05-10. IFLA Series on Bibliographic Control, vol. 31., A.3. https://www.ifla.org/files/assets/cataloguing/isbd/isbd_wwr_20100510_clean.pdf.

111. Casalini et al., *National Strategy, 18 (see n. 108)*.

112. Wikipedia; "Integrated Library System," 9 May 2019, at 10:33 (UTC). https://en.wikipedia.org/wiki/Integrated_library_system.

113. Wikidata; "Lander's Peak (Q20588520)," 29 April 2019, at 14:58. https://www.wikidata.org/wiki/Q20588520.

114. Wikimedia Commons; "File: The Rocky Mountains, Lander's Peak (Albert Bierstadt), 1863 (oil on linen - scan).jpg," 18 June 2019, at 04:44. https://commons.wikimedia.org/wiki/File:The_Rocky_Mountains,_Lander%27s_Peak_(Albert_Bierstadt),_1863_(oil_on_linen_-_scan).jpg.

115. Wikipedia; "The Rocky Mountains, Lander's Peak," 8 May 2019, at 23:54 (UTC). https://en.wikipedia.org/wiki/The_Rocky_Mountains,_Lander%27s_Peak.

116. Washburn, Bruce, Jeff Mixter, and Jean Godby. 2018. "Using Wikibase to Manage Linked Data for Digital Collections" (PowerPoint presentation, CONTENTdm User Group Meeting, Columbus, OH, 2 August 2018. https://www.oclc.org/content/dam/research/presentations/2018/20180802-washburn-wikibase-manage-linked-data-cdmug.pptx.

117. Svenonius, Elaine. 2000. The Intellectual Foundation of Information Organization. Cambridge, MS: MIT Press.

118. Lubetzky, Seymour. 1953. *Cataloging Rules and Principles: a Critique of the A.l.a. Rules for Entry and a Proposed Design for their Revision. Prepared for the Board on Cataloging Policy and Research of the A.L.A. Division of Cataloging and Classification.* Washington: Processing Dept., Library of Congress. https://catalog.hathitrust.org/api/volumes/oclc/1373808.html.

119. MediaWiki. "Wikibase/DataModel/Primer." Updated 1 August 2018, at 14:02. https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer.

120. Smith-Yoshimura, Karen. 2012. *Social Metadata for Libraries, Archives, and Museums: Executive Summary.* Dublin, OH: OCLC Research. https://doi.org/10.25333/C3H34P.

121. WikiConference North America 2018 – Columbus. Accessed 14 July 2019. https://wikiconference.org/wiki/2018/Main_Page;

Wikiconference North America. "Submissions:2018/Strategies to Counter Systemic Bias: Writing Campaigns, New Maps, and Informed Review." Submissions: Discussion. Updated 4 September 2018. https://wikiconference.org/wiki/Submissions:2018/Strategies_to_Counter_Systemic_Bias:_Writing_Campaigns,_New_Maps,_and_Informed_Review;

Wikiconference North America. "Submissions:2018/Advancing Gender Equity: Conversations with Movement Leaders." Updated 15 October 2018. https://wikiconference.org/wiki/Submissions:2018/Advancing_Gender_Equity:_Conversations_with_Movement_Leaders;

Wikiconference North America. "Submissions:2018/Inclusion, Wikipedia, and right-of-center topics." Updated 21 October 2018. https://wikiconference.org/wiki/Submissions:2018/Inclusion,_Wikipedia,_and_right-of-center_topics.

122. The Library of Congress. 2018. "Charge for PCC Task Group on Identity Management in NACO," 5. American Bar Association, Program for Cooperative Cataloging, revised 22 May 2018. https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf.

123. Wikidata. "Wikidata:WikiCite/Roadmap." Updated 12 May 2019, at 13:07. https://www.wikidata.org/wiki/Wikidata:WikiCite/Roadmap.

124. "Enslaved: People's of the Historic Slave Trade." Accessed 9 July 2019. http://enslaved.org/.

125. "Appel d'offre pour une preuve de concept : WikiBase comme infrastructure pour le Fichier national d'entités." T*ransition bibliographique*.fr. Updated 3 avril 2019. https://www.transition-bibliographique.fr/2019-04-02-appel-offre-preuve-concept-wikibase-pour-fne/. [French]

126. Fisher, Barbara. 2019. "Authority Control meets Wikibase." Confluence. Updated 20 May 2019. https://wiki.dnb.de/display/GND/Authority+Control+meets+Wikibase.

127. FactGrid: A Database for Historians. "Main Page." n.d. Accessed 15 July 201. https://database.factgrid.de/wiki/Main_Page.

128. Simons, Olaf. 2019. "Memorandum of Understanding Between the University of Erfurt and the German National Library – to Base the Factgrid on Gnd Data in a Joint Project. *FactGrid: A Database for Historians* (blog), 17 April 2019. https://blog.factgrid.de/archives/1527. [German Version in English]

129. Ohlig, Jens, and Barbara K. Fischer. 2019. "GND meets Wikibase – Eine Kooperation: A Federal Agency Goes on an Expedition in the Wikiverse." *gnd Blog*, Deutsche Nationalbibliothek. Updated 20 May 2019 by Barbara Fischer. https://wiki.dnb.de/pages/viewpage.action?pageId=147754828.

130. Library of Congress. "MARC PROPOSAL NO. 2017-08" MARC Development Proposals List. Standards Office. Updated 16 May 2017. https://www.loc.gov/marc/mac/2017/2017-08.html.

131. Deliot, Corine, Steven Folsom, MJ Han, Nancy Lorimer, Terry Reese, and Adam Schiff. 2019. *Formulating and Obtaining URIs: A Guide to Commonly used Vocabularies and Reference Sources*. Library of Congress PCC Task Group on URIs in MARC. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf.

132. "*PCC (Program for Cooperative Cataloging) Strategic Directions: January 2018-December 2021.*" 2019. Library of Congress. http://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf.

133. Baxmeyer, Jennifer, Karen Coyle, Joanna Dyla, MJ Han, Steven Folsom, Phil Schreur, and Tim Thompson. 2017. *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies*. Library of Congress PCC Linked Data Advisory Committee. https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf.

134. OCLC Research. "OCLC Research Library Partnership Metadata Managers Focus Group." Updated 18 April 2019. http://www.oclc.org/research/themes/data-science/metadata-managers.html.

135. For a summary of these discussions, see: Smith-Yoshimura, Karen. 2018. "The Coverage of Identity Management Work." *HangingTogether* (blog). OCLC Research. Posted 8 October 2018. http://hangingtogether.org/?p=6805.

136. Futornick, Michelle. 2019. "LD4P2 Project Background and Goals." Linked Data for Production: Pathway to Implementation (LD4P2). Duraspace. Posted 14 January 2019. https://wiki.duraspace.org/display/LD4P2/LD4P2+Project+Background+and+Goals.

137. Association of Research Libraries Task Force. 2019. *ARL White Paper on Wikidata Opportunities and Recommendations*,38. Washington, DC: Association of Research Libraries. https://upload.wikimedia.org/wikipedia/commons/7/70/ARL_White_Paper_on_Wikidata_Opportunities_and_Recommendations.pdf.

138. Casalini, *National Strategy (see n. 107); Wang, Jing. 2018. Shareable Local Name Authority Reference Model (Draft)*. Cornel University *eCommons digital reposity*. https://hdl.handle.net/1813/56582.

139. Wikipedia. "Wikipedia:WikiProject University of Virginia/Program for Cooperative Cataloging May 2019." Updated 22 May 2019, at 15:31 (UTC). https://en.wikipedia.org/wiki/Wikipedia:WikiProject_University_of_Virginia/Program_for_Cooperative_Cataloging_May_2019.

140. Stanford Libraries. "Hilary K. Thorsen." Wikimedian in Residence. n.d. Accessed 15 July 2019. https://library.stanford.edu/people/thorsenh.

141. LD4 Wikidata Affinity Group. "2019-04-23 LD4 Wikidata Affinity Group Call." (Agenda and Audio recording). Google docs. https://docs.google.com/document/d/1BuszEQQxlOY14hK60Fl7n8Huvh6jEWXre0-wSvpyq84/edit#heading=h.wquti58weptm.

142. Klein, Maximilian, and Alex Kyrios. 2013. "VIAFbot and the Integration of Library Data on Wikipedia." *code4lib Journal,* no. 22. http://journal.code4lib.org/articles/8964.

143. OCLC Research. "Libraries Leverage Wikimedia." Accessed 15 July 2019. https://www.oclc.org/research/themes/community-catalysts/libraries-wikimedia.html.

144. OCLC. "OCLC Wins Knight News Challenge Award to Promote Collaboration Between Public Libraries and Wikipedia: Project Will Reinforce Libraries as Stewards of High-quality Information." News release, Dublin, Ohio, 23 June 2016. https://www.oclc.org/en/news/releases/2016/201612dublin.html;

OCLC WebJunction: The Learning Place for Libraries. "Wikipedia + Libraries: Better Together: Overview." Accessed 16 July 2019. https://www.oclc.org/content/webjunction/explore-topics/wikipedia-libraries.html.

145. Lih, Andrew, and Robert Fernandez. 2018. "Works in Progress Webinar: Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond." Produced by OCLC Research, 12 June 2018. MP4 video presentation, 1:1:51. https://www.oclc.org/research/events/2018/06-12.html.

146. OCLC Research. "Linked Data Wikibase Prototype." Linked Data from OCLC Research. Accessed 15 July 2019. https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-prototype.html.

147. IIIF (International Image Interoperability Framework). "Enabling Richer Access to the World's Images, Home." Accessed 15 July 2019. https://iiif.io/;

OCLC Research. "IIIF: Improving the Interoperability of Digital Materials." Accessed 15 July 2019. https://www.oclc.org/research/themes/data-science/iiif.html.

148. Google Groups. "Archives and Linked Data Interest Group." https://groups.google.com/forum/#!forum/archives-and-linked-data;

W3C. "Alternative 1 model proposal." Updated 21 February 2018. https://www.w3.org/community/architypes/wiki/Alternative_1_model_proposal.

149. Schema.org. "Schema.org Section: Pending." Accessed 15 July 2019. https://pending.webschemas.org/.

150. Library of Congress. "MARC PROPOSAL NO. 2017-08" MARC Development Proposals List. Standards Office. Updated 16 May 2017. https://www.loc.gov/marc/mac/2017/2017-08.html.

151. Shieh, Jackie, Terry Reese, Jean Godby, Steven Folsom, and Tiziana Possemato. 2016. "Program for Cooperative Cataloging (PCC) Task Group on URIs in MARC." Presented at the International Conference on Dublin Core and Metadata Applications, DC-2016, Copenhagen, Denmark, 14 October 2016. https://www.oclc.org/content/dam/research/presentations/2016/101416-pcc-task-group-uris-marc.pdf.

For more information on OCLC Research, visit: **oc.lc/research**

6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878
T: +1-614-764-6000
F: +1-614-764-6096
www.oclc.org/research

ISBN: 978-1-55653-135-4
DOI: 10.25333/faq3-ax08
RM-PR-216276-WWAE 1907