UCLA

UCLA Previously Published Works

Title

Reflections on the PCC Wikidata Pilot at UCLA Library: Undertaking the PCC Learning Objectives

Permalink

https://escholarship.org/uc/item/91w437s4

Journal

Cataloging & Classification Quarterly, 61(7-8)

Authors

Zhang, Erica

Biswas, Paromita

Dagher, Iman

Publication Date

2023-11-17

Peer reviewed

Reflections on the PCC Wikidata Pilot at UCLA Library: Undertaking the PCC Learning Objectives

Erica Zhang (0000-0003-0835-9643), Paromita Biswas (0000-0002-1008-5365), and Iman Dagher (0000-0002-1793-8578)

UCLA Library, UCLA, Los Angeles, California, USA

In 2020, the Program for Cooperative Cataloging (PCC) Task Group on Identity Management in NACO sponsored a 14-month PCC Wikidata Pilot, complete with learning objectives, for participants to experiment with Wikidata, an open linked data platform. UCLA Library joined the Pilot to create and edit Wikidata items related to UCLA Library's collections and UCLA Library entities. With the Pilot's conclusion, the UCLA Library Pilot team reflected on lessons learned. By assessing UCLA Library's experience against the Pilot's learning objectives, the authors hope to contribute on-the-ground insights that may be relevant to PCC's progress toward identity management, and the role Wikidata may play in this transition.

Keywords: Wikidata, identity management, linked data, Program for Cooperative Cataloging, metadata

Introduction

The UCLA Library is a member of the Program for Cooperative Cataloging (PCC), "a diverse coalition of practitioners, service providers, and tool developers working together to harness the full potential of metadata to promote and sustain knowledge."[1] In its mission to create metadata that is trusted and used by libraries and cultural heritage organizations, the PCC initiates pilots that explore cutting edge developments in metadata. One recent pilot that was open to both PCC and non-PCC members is the PCC Wikidata Pilot, which UCLA Library participated in from

1

October 2020 to December 2021.[2] The Wikidata Pilot was sponsored by the PCC Task Group on Identity Management (now the Identity Management Advisory Committee), with the goal of advancing PCC's movement toward identity management. The concept of identity management prioritizes the creation of identifiers for resources, such as that represented by unique URIs, as opposed to unique text strings, and is part of a broader PCC effort to transition library data to linked data that will increase this data's visibility on the web. In this article, the authors share their reflections on UCLA Library's engagement with the Wikidata Pilot; in particular, this article assesses the Library's Pilot experience against the PCC Wikidata Pilot's learning objectives. The authors chose to organize the article around the learning objectives, as these objectives can provide a window for assessing the viability of Wikidata as a library tool for the identification and description of resources and entities. The authors also identify UCLA Librarylearning objectives not fulfilled within the Pilot timeframe. The authors hope that this on-the-ground assessment of the UCLA Library Pilot team's[3] experiences will be relevant for the PCC's evaluation of Wikidata as a useful tool for identity management that can be used by the library community in its transition to linked data. The PCC's endorsement of Wikidata work is crucial for maintaining momentum in this area and in justifying catalogers' allocation of time and labor to it, and the authors hope this article will help the PCC make informed decisions around the future of Wikidata and library metadata work.

Background

Optimizing library data for the semantic web has been an integral part of the PCC's strategic directions in recent years, as new technologies present opportunities to expose bibliographic metadata on the web and facilitate greater access to library resources. PCC's 2015-2017 Strategic

Plan identified as a strategic direction "[advancing] the community's understanding of linked data;"[4] subsequent strategic directions have expanded to include applying this understanding of linked data in the cataloging context, such as through the use of linked data vocabularies, and experimentation with linked data sandbox environments.[5] In particular, the PCC Name Authority Cooperative Program (NACO)[6] has been exploring a transition from traditional ideas of authority control – which involves catalogers creating unique text strings as access points – to identity management, which relies more on creating and using unique identifiers to manage entities.

To move in this direction, the PCC Task Group on Identity Management in NACO, a group charged in 2016 to lead the community in the transition from authority control to identity management, has sponsored three pilots so far to explore the production of bibliographic metadata as linked data:

- the PCC ISNI Pilot (2017-2018)[7] where PCC participants engaged in identity management activities, such as creating and modifying ISNIs for persons and organizations, in the ISNI database managed by the ISNI International Agency;
- the PCC URIs in MARC Pilot (2019-2021)[8], where PCC participants tested the inclusion of Uniform Resource Identifiers or URIs in bibliographic and authority records in the MARC format, the data format most commonly used for library catalog data[9]; and
- most recently, the PCC Wikidata Pilot.

UCLA Library, as a member of PCC, has participated in all three of PCC's linked data pilots, contributing entities in the ISNI registry, adding actionable URIs in MARC authority records, and working within Wikidata in the latest Pilot, the subject of this article.[10]

The PCC Wikidata Pilot & Learning Objectives

Wikidata has increasingly become a knowledge graph of interest to the library community over the years, and is explicitly called out in the 2018-2022 PCC Strategic Directions.[11] Wikidata is an open, multilingual, collaboratively-maintained knowledge graph hosted by the Wikimedia Foundation, and contains millions of entities, called items, represented as structured data with unique identifiers, that can be read and edited by humans and machines alike.[12] As an open linked data platform Wikidata has emerged as "a high-demand identifier hub, with considerable community buy-in and recognition as an identity reconciliation service," making it "ideal in the implementation of identity management principles" that facilitates "a NACO-Lite approach that would lower the barrier to authority creation and maintenance by PCC members."[13] Consequently, to explore cataloging and metadata work in Wikidata, the PCC Task Group on Identity Management in NACO launched the PCC Wikidata Pilot in September 2020.

During the Pilot kick-off meeting, the Task Group outlined four broad categories of learning objectives around experiences of creating and editing entities in Wikidata, automation possibilities in Wikidata, training resources available, and participation in the broader Wikidata community and its norms and practices. As the co-chair of the Task Group shared, these objectives, which were also informed by community input, reflected their interest in learning "how to contribute to Wikidata and think about the tools that they [library participants] use in library work that would need to work in combination with the tools associated with Wikidata to create efficient workflows."[14] (Table 1)

Table 1. PCC Wikidata Pilot Learning Objectives

- Comparison to other registries
    - Production rate: Time to create an identifier & describe entity
    - Requirement for documenting sources for data elements
    - Proportion of needed entities found in the database
    - UI ease of use for data entry & workflow integration with other tools
    - What it's like relating one entity to another
    - How duplicate entries are surfaced and dealt with
    - Ease of finding & using documented best practices
- Assessing Productivity and QA Tools
    - Batch searching
    - Batch loading
    - Notifications of changes made to our prior contributions
    - Quality assurance tools that can bring to light problematic data
    - Detection of data conflicts when we attempt to use data
- Advisable Training Resources
- Participating in the Community
    - The developers
    - Funding of initiatives
    - The community decision-making process
    - What it is like to propose a new property and when it seems warranted
    - Can we enrich mappings between Wikidata properties and corresponding lists on loc.gov? What to do when properties do not match
    - What it is like to propose new/changed policies
    - With new/changed policies, what is done with editing of legacy data?
    - Persistence of the data, in light of the Notability policy & shared editing policy
    - Editing data when underlying policies do not match
    - How much library goals align with community and can be advanced

The Pilot was initially scheduled to run for one year, but due to continued interest, was extended to the end of 2021. At the final meeting of the PCC Wikidata Pilot, the Task Group facilitated a discussion with Pilot participants in December 2021 regarding the Pilot's learning objectives, and potential next steps for the PCC and Wikidata.[15] The Task Group also solicited Pilot participants' experiences with Wikidata via a PCC Wikidata Pilot Assessment survey, which opened from March 31, 2022 to April 11, 2022. The survey was divided into several sections, including Outputs, Training, Tools, Collaboration, and Wikidata and the PCC. The

Task Group then shared results of the survey at the PCC Joint Operations Committee Meeting held virtually in May 2022.[16]

The PCC Wikidata Pilot at UCLA Library

UCLA Library joined the Wikidata Pilot in October 2020 with 16 participating individuals. The timing of the Pilot was particularly fortuitous, as many UCLA Library employees were working remotely full-time due to the COVID-19 pandemic, and looking to pivot to work that did not require access to physical materials; this made the Wikidata Pilot an ideal workload assignment. The majority of participants were members of UCLA Library's Resource Acquisitions and Metadata Services (RAMS) department, but the Pilot also involved colleagues from Library Special Collections (LSC), UCLA Library's Digital Library Program, and the UCLA Library Center for Oral History Research. Participants brought different backgrounds and expertise in metadata, from traditional MARC cataloging - including copy cataloging, original cataloging, NACO work, and Subject Authority Cooperative Program (SACO)[17] work - to archival description and digital collections metadata work using standards such as Dublin Core.

Among the various motivations for participating in the Pilot, Pilot team members found particularly exciting the opportunity to have UCLA Library's oral history collections, special collections, as well as UCLA faculty, departments, and building more visible on the semantic web via Wikidata - the "open data knowledge base". Pilot team members were also interested in exploring the integration of Wikidata into cataloging workflows, including the use of Wikidata as an alternative vocabulary in bibliographic records for creator entities not currently in the Library of Congress name authority file (LC NAF), as well as using Wikidata to enrich NACO

authority records through the OCLC Connexion Authority Creation Toolkit developed and maintained by Gary Strawn at Northwestern University.[18] Additionally, Pilot team members were interested in adding scripts such as Armenian in Wikidata, scripts which were not available for use in Library of Congress authority files during the Pilot period. Under the leadership of the UCLA Library Pilot team coordinator, Pilot team members proposed various topics of interest, and organized project teams to work on these topic-specific projects. Membership in these project teams was fluid, and Pilot team members could join multiple projects as well as shift from one project to another throughout the Pilot. Over the course of the Pilot, members engaged in Wikidata projects around the following entities:

- UCLA buildings and places

- UCLA faculty

- Oral histories from the UCLA Library Center for Oral History Research

- UCLA Library special collections

- Open access serials housed in University of California's institutional repository, eScholarship

- Entities represented in non-Roman languages and scripts

To facilitate collaborative learning around Wikidata, the Pilot team ran weekly meetings to develop a foundational understanding of Wikidata, then transitioned to bi-weekly training meetings to discuss more complex Wikidata concepts, application profiles, data modeling, SPARQL querying, and batch tools. The Pilot team also held separate monthly meetings to report on the progress of the various projects and discuss next steps.

By the end of the Pilot, the Pilot team had collectively created over 1,000 new Wikidata items, and edited over 2,000 existing items (Figure 1). The team also successfully proposed four new Wikidata properties related to describing special collections, buildings, and serials (Table 2).
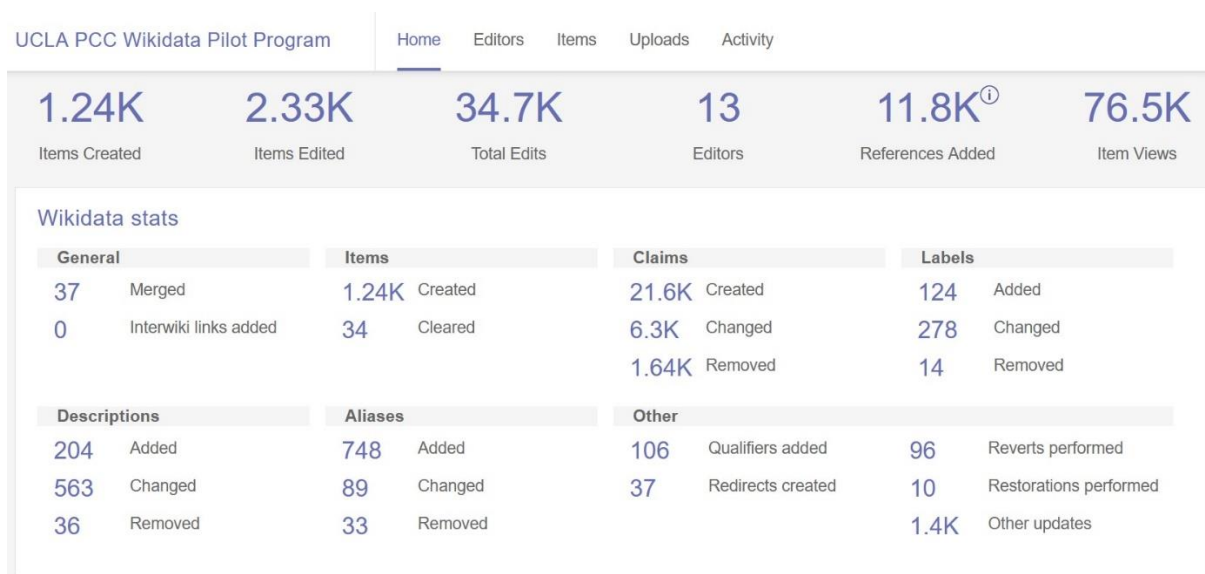

Figure 1: UCLA PCC Wikidata Pilot Program Dashboard. University of California, Los Angeles, "UCLA PCC Wikidata Pilot Program," Wiki Edu, accessed May 10, 2022, https://outreachdashboard.wmflabs.org/courses/University_of_California,_Los_Angeles/UCLA_PCC _Wikidata_Pilot_Program/home.

Table 2. Project statistics.

Project	# of Items created/edited	Properties proposed
Buildings/Places/Architects	180	UCLA Space Inventory LocID (P9794)
Faculty	665	
Corporate Bodies/Academic Units	134	

Serials	22	issuing agent of work (P9901)
Special Collections (Persons)	42	
Special Collections (Collections)	183	Online Archive of California finding aid ID (P9592) Densho Encyclopedia ID (P9938)
Oral Histories (Persons)	240	
Oral Histories (Collections)	37	

Artifacts of the Pilot team's work, including application profiles - guidelines for describing different types of entities with statements - can be found on the team's WikiProject page.[19] Not only did Pilot team members contribute entities to a global and open linked data platform, the experience of collaboratively learning to work in Wikidata led to intense and productive discussions around the process of creating native linked data, the possibilities as well as limits of batch tools, and the challenges and opportunities of working in cooperation with a global community whose membership is comprised primarily of non-catalogers.

Literature Review

There has not been any scholarship on assessing Pilot participants' experiences against the learning objectives identified at the beginning of the PCC Pilot. Much of the literature that currently exists about the PCC Wikidata Pilot can be found on WikiProject pages that Pilot participants have created, and/or presentations that Pilot participants have given at monthly Pilot meetings or in relevant venues such as the annual LD4 Conference on Linked Data.[20] Of this body of work, few participants identified as their goals any of the learning objectives, although

some participants such as the University of Washington listed objectives related to gaining an understanding of Wikidata that reflect the larger PCC learning objectives.[21] Other participants such as Texas State, San Diego State, and Brandeis University Library mention an alignment of their Pilot projects with the broader PCC Strategic Directions, namely SD4, which highlights the need to move towards "ubiquitous identifier creation and identity management."[22] A number of participants listed projects related to exploring Wikidata and authority control which can speak to the "Comparison to other registries" learning objective, such as authority workflows for entities not in LC NAF (Library of Congress), inclusion of Wikidata items in authority files or bibliographic platforms (University of Utah; Harry Ransom Center; Casalini Libri),[23] as well as Wikidata and potential integration into NACO work. Participants were also interested in coverage of resources and entities that were otherwise not included in national authority files and databases such as LC NAF and Library of Congress Subject Headings (LCSH). For example, there was work on increasing discovery and access to library collections by creating Wikidata items for underrepresented collections and the people found in collections, such as Illinois State University's project on performers found in digitized collections of circus books[24]; IUPUI's work on subjects and communities in Indiana in support of their Scholarly Profiles as a Service project[25]; University of Delaware's project to increase visibility for persons related to Delaware or the university that would be limited by Wikipedia's notability criteria[26]; and Philadelphia Museum of Art's project to create entities for African American artists, to name a few.[27]

During the Pilot, many participants also presented on their Pilot projects and their reflections about working in Wikidata either at PCC Wikidata Pilot monthly meetings or at other external venues.[28] However, none of these presentations provided a direct assessment of participants' projects vis-à-vis the PCC's learning objectives for the Pilot. The results of the Task

Group's post-Pilot survey contained questions related to aspects of the learning objectives, such as whether participants felt there was sufficient existing training materials, and participants' experience with existing tools, but these results also don't necessarily address specifically the PCC learning objectives set out at the beginning of the Pilot.[29]

UCLA Library and the PCC Learning Objectives

The headings in this section correspond to the PCC Learning Objective bullets found on the PCC Wikidata Pilot page. In writing this article, the authors consulted with a member of the Task Group for clarification of some objectives as needed. The authors also identify learning objectives not fulfilled within the Pilot timeframe.

Comparison to other registries

Throughout the Pilot, the Pilot team members found themselves comparing Wikidata work to existing metadata workflows, with recurring comparisons between Wikidata and the LC NAF and LCSH.

Production rate: time to create an identifier and describe entity

Wikidata and MARC Authority Records. In current cataloging practice, creating authority records for entities related to our library collections is an important component of information organization and retrieval. Records added to national level authority files such as the LC NAF and LCSH follow complex rules of standardized heading construction outlined in cataloging manuals; catalogers spend time creating these headings that are unique, as they must be disambiguated. Even variant or related terms in authority records that allow discovery when

searching other forms of the preferred or authorized name are constructed following specific guidelines. However, this authority work is expensive, time-intensive, and not always conceivable given oftentimes limited staff with language expertise as well as extensive training requirements that guide participation in the PCC's NACO and SACO programs.[30]

By contrast, as Pilot team members experienced early in the Pilot, creating a Wikidata item can happen with just a few clicks. Unlike the rule-heavy process for creating authority records for names and subjects in MARC, Wikidata prioritizes creating a unique identifier (Q number) for the new entity or Wikidata item, with just minimal pieces of information for human readability to start: a label (which need not be unique), a short description, and any aliases if applicable. Each Wikidata item can represent anything in "human knowledge, including topics, concepts, and objects."[31] Information about an item is conveyed via statements which identify attributes of an item through the use of "properties" that have unique P numbers.[32] Adding attributes comes after the item is created, and there are no required minimum statements for a Wikidata item.[33] This contributes to the relatively quick production aspect that Pilot team members first experienced; there is no need to spend time interpreting cataloging rules to format a unique label. Composing the natural language label in Wikidata is flexible; it does not have to be unique, and disambiguation is based on the short description and statements about the item, not a carefully constructed text string label. It was not unusual for Pilot team members to come across entities in Wikidata with the same label. For example, Wikidata has several items with the same label "interviewer"; however, each of these items have their unique Q numbers as well as different descriptions to distinguish them from one another (Figure 2).

Figure 2: A search for "interviewer" brings up items with the same label; what differentiates them is their short description as well as viewing each item's statements.

Similarly, while both Wikidata and MARC authority records place value on providing justification for entries, they have different ways of doing so. In MARC authority records, at least one MARC 670[34] field citing the resource for which the authority record is being established is required. When describing entities in Wikidata, justification and literary warrant notes are not mandatory. Providing references for Wikidata statements are strongly encouraged via property constraints,[35] such as the "citation needed constraint (Q54554025)" for the property "date of birth (P569)"; however, the citation does not need to be formatted in a particular way as is required of a MARC field; instead, it can be as simple as adding a URL.

Wikidata and MARC Bibliographic Records. Another aspect of Wikidata that Pilot team

members found contributed to a faster production rate is the way Wikidata handles multilingual

descriptions of entities. Unlike in the MARC bibliographic environment where, according to

established practices, different languages of cataloging warrant separate records that aren't

always connected to each other, descriptions of the same entity in Wikidata, even in multiple

languages, are all connected to one unique identifier for the entity. What this means in terms of

production time is that Wikidata makes it easy to build on the work done in other languages.

With the help of automation, for instance, Wikidata items generated from Wikipedia articles

written in non-Latin languages automatically include statements in both the language of the

article as well as its English translation. This automatic translation is not available for the label,

description, and aliases; however, Wikidata has bots, automated tools created and run by

community contributors, that add machine translations of descriptions and aliases.[36] Pilot team

members could also manually create the English-language label that they considered appropriate

without having to apply the specific rules of the ALA-LC Romanization tables, or could even

consider using the auto-transliteration of names of a human bot, if available for that language.[37]

Additionally, more programs are being created to add multilingual content to Wikidata, such as

VanderBot, a program that works with English and Arabic data in Wikidata.[38]

Wikidata and Cataloging Rules. As easy as it is to start creating and describing Wikidata items

in a technical sense, Pilot team members who were accustomed to following rules, guidelines,

and standards that typically govern cataloging found it a challenge to work in the absence of such

guidance in Wikidata. Without application profiles and best practices, it was sometimes difficult

and time consuming to figure out an optimal level of description for an entity. For example,

while the Pilot team members benefited greatly from libraries who had already been working in Wikidata and shared practices and application profiles they had developed for similar projects, there were still decision points the team faced that were not addressed by existing documentation. One gap in best practices that Pilot team members experienced was the lack of guidance on formulating English-language labels for entities described only with non-Latin labels: to translate, transliterate according to ALA LC Romanization table, or simply write it in Latin form.

Wikidata and Data Entry. Another aspect of Wikidata that slowed down Pilot team members' work was the structure of statements used to describe a Wikidata item. Because statements are structured as triples, and many properties have items as their range values, if the value does not exist as a Wikidata item, Pilot team members would have to either make the choice to not add the statement, or invest time in first creating a separate Wikidata item for the missing value, before coming back to the original item to create a relationship to that new item. For instance, Pilot team members creating items for UCLA faculty found that when creating a statement for the faculty member's department, the department was often not in the Wikidata knowledge base. Team members would therefore have to spend additional time creating the item for the faculty member's department so they could then incorporate this information in the item for the faculty member. This is markedly different from MARC cataloging, where metadata is largely keyed as unstructured text in MARC bibliographic record fields and even in certain fields in MARC authority records.[39] MARC records can accommodate a lot of data not yet in an authority file so that the record can still be published with such "uncontrolled" terms.

In sum, it can seem easy to get started creating Wikidata items, particularly when considering the machine translation of labels as compared to the training and time needed to create MARC authority records. However, the Pilot team found that the lack of established guidelines and practices, and the necessity of having to create additional items for use in statements, quickly slowed down the production rate of creating Wikidata items. Such complexities therefore need to be taken into account in any assessment of production rate for creating Wikidata items.

Requirement for documenting sources for data elements

As mentioned in the previous section, Wikidata does not require the type of citation or source note typically documented in MARC field 670 and in subfields $v and $u added to attributes found in authority records. Instead, while not mandatory, Wikidata encourages contributors to provide references at the statement level. Contributors are encouraged to use properties such as "stated in (P248)" and "reference URL (P854)" for citing publications, websites, and online databases to back claims made in statements.[40] However, when it comes to labels and aliases for Wikidata items, there is no way to attach source information that justify their usage. The Pilot team, therefore, consulted existing application profiles, such as University of Washington's application profile for faculty and staff,[41] that noted practices regarding statements requiring references - for example, "personal pronoun (P6553)" - and adapted such guidelines as applicable. Pilot team members also developed a data quality practice of adding qualifiers to statements, namely, the properties "reference URL (P854)" and "retrieved (P813)", to capture sources of information. To assist with the potentially laborious task of adding source information at such a granular level in each statement, Pilot team members made frequent use of macros or

gadgets such as the "DuplicateReferences" gadget, which helped with adding the same source information to all statements that shared that source in a particular item.

The Pilot team brought the use of references to their Wikidata work by incorporating references into application profiles. However, because references are ultimately not mandatory in Wikidata, Pilot team members sometimes encountered challenges when trying to disambiguate items or verify statements in a Wikidata item. References help with the verifiability of statements, and Pilot team members found inconsistent application of references to be a recurring data quality issue in Wikidata.

Proportion of needed entities found in the database

Wikidata is a growing platform; new items and properties are added on a daily basis. Despite this growth, however, the Pilot team members often found gaps in needed entities in the knowledge base. For instance, Wikidata entries for earlier and later titles of serials often had to be created along with items for publishers or issuing agencies for these titles.

Pilot team members also encountered gaps with properties, which take more community involvement and time to get approved for use. For instance, Pilot team members working on describing serials in Wikidata found there was no property for identifying issuing agents - entities who play a crucial role in the intellectual production of a serial and therefore need to be included in serial descriptions. Pilot team members over the course of the Pilot submitted four property proposals related to serials, buildings, and special collections.

UI [User interface] ease of use for data entry & workflow integration with other tools

Wikidata's user interface for data entry is noticeably different from interfaces of programs such as OCLC's Connexion client. Display-wise, labels for all properties and statements are meant to be human-readable. In contrast to working in Connexion, which involves primarily keying information, some of which can then be validated and hyperlinked if the text string is an exact match to headings in the LC authority files, Wikidata is more about selecting entities within the Wikidata knowledge base. That is, Pilot team members found that working in Wikidata was not so much "data entry" as it was selecting values from a predetermined list within the Wikidata knowledge base. Aside from properties for references, external identifiers, and statements related to dates, the vast majority of statements the Pilot team members worked with required Wikidata items as the range value. Thus, entering a value often meant selecting from a drop-down menu of existing Wikidata items; at the same time, Pilot team members needed to be careful selecting the appropriate item within the drop-down because of the lack of required uniqueness of any label; that sometimes meant reading through each item's page of statements for verification.

In terms of workflow integration with other tools, one area of integration that some members of the UCLA Library Pilot team explored was the integration of Wikidata with Gary Strawn's Authority Toolkit for NACO work in OCLC Connexion. This integration permits very quick enrichment of an in-progress NACO record based on using the structured data in Wikidata and other registries such as VIAF, data that would have to be laboriously keyed in otherwise by the cataloger. However, Pilot team members using this tool found that manual review of the data was still needed to ensure record validation and data quality. This tool can also generate the 024 MARC field used for recording standard identifiers from registries such as Wikidata in NACO

records, a practice first carried out by participants in the PCC URIs in MARC Pilot and now encouraged of all NACO catalogers.

What it's like relating one entity to another

Wikidata's ontology is flexible and collaborative[42] and there are no restrictions on how one can describe an item with statements that connect entities as well as define the relationships between them using properties. For example, with regards to the application profile for UCLA faculty, Pilot team members discussed whether to use "educated at (P69)" and "academic degree (P512)" as separate properties in two statements, or use "academic degree (P512)" as a qualifier for the property "educated at (P69)."[43] (Figure 3)

Figure 3: The Wikidata item for "John Rahn (Q1565424)" contains two different methods of modeling academic degree information, one where "academic degree (P512)" is the main property, and one where "educated at (P69)" is the main property. The Pilot team ultimately decided to use "educated at" as the main property.

Oftentimes, the Pilot team made decisions by reviewing example application profiles, and/or using Wikidata's query service to determine the more common modeling choice. The Pilot team also encountered instances where reciprocal properties did not exist, and instead, the concept was created as an item, such as "has oral history of (Q106929368)", an inverse of the property "oral history at (P9600)."

As Pilot team members explored the Wikidata ontology further, they had more questions about the structure behind what might have initially appeared as a straightforward concept. Notably, the team members grappled with the difference between the "instance of (P31)" and "subclass of (P279)" properties.[44] For instance, particular university buildings have been identified as a "subclass" of an architectural element in Wikidata rather than as an "instance" of an architectural element. This example in particular represented to members how the knowledge graph of Wikidata is shaped by the way it is collaboratively understood and edited by its community of users.[45]

While constraints are available to guide the intended use of properties, and the violation of constraints displays a warning flag for the editor, there is no real mechanism for enforcing adherence to those constraints, or addressing the warning flags.[46] This makes it difficult to rely on a certain uniformity in the way items are described in Wikidata, and the Pilot team often faced difficult choices when describing items. For example, when describing UCLA Library collections, such as the *American Indian Relocation Project* oral history collection, one question

that arose for Pilot team members was whether the Wikidata item should describe this project as a corporate body "agent" or the collection of interviews as "works." Or, for archival digital collections, whether there was a need to create separate items for the print and the digital versions. Or, whether a serial title change can be captured within a single item for the serial with each title change qualified by a start and end times (ex: "The New York Times (Q9684)") or whether each title change should be represented by their own item in Wikidata using properties such as "Replaced by (P1366)" that can be mapped to the earlier/later title fields in MARC records (780/785) (e.g. items for the "Black Law Journal (Q100319362)" and its successor the "National Black Law Journal (Q15756635)").[47] In yet another instance, when describing certain university units such as libraries in Wikidata, Pilot team members questioned whether to describe the library as an architectural building, or a corporate body, or both. That is, should there exist two Wikidata items, or a single Wikidata item reflecting both these entity types? The *LC Subject Headings Manual*, from a bibliographic perspective, provides rules to conflate the architectural entity with the corporate body when they share the same name.[48] However, from another modeling perspective, buildings and corporate bodies are two different types of entities, each with different attributes. Did it make sense, therefore, to conflate the two? What were the implications for users of this data?

How duplicate entries are surfaced and dealt with

Merging items in Wikidata involves bringing together all of the data from multiple items into one single item and then redirecting the link from the obsolete items to the merged item (recipient item), rather than deleting the obsolete items.[49] This is separate from deleting items, which is more appropriate for items that may not conform to one of Wikidata's few policies,

21

such as its notability criteria.[50] Unlike in databases such as OCLC Connexion, where only PCC

member libraries can merge bibliographic records after extensive training and duplicate authority

records in LC's authority files can only be merged by LC, the merge process in Wikidata can be

done by any contributor with the help of the "Merge" gadget enabled in their account

Preferences or through Wikidata's "Special Merge" process.[51] Of course, given the ease of

merging, users may need to be careful of mistakenly merging items, particularly if they share the

same label. However, Wikidata provides guidance for unmerging incorrectly merged items.

While Wikidata provides a central page for those specifically looking for potential duplicate

items to merge, Pilot team members pursued merges primarily for duplicate items they came

across through the course of the Pilot projects.

Ease of finding & using documented best practices

With regard to Wikidata documentation, Pilot team members found it somewhat challenging to

navigate or search for Help pages within the Wikidata platform. While Wikidata has a link to its

list of tutorials on how to get started editing in Wikidata, its main search box prioritizes

searching for Wikidata items. Finding properties or other user pages (both of which the Pilot

team frequently searched for) wasn't particularly intuitive; Pilot team members had to go to the

Advanced Search page in Wikidata to search for these. Furthermore, the Advanced Search page

itself is not quite visible, and one has to click twice on Wikidata's main search box to get to this

page. Instead, Pilot team members found that using Google to search for Wikidata

documentation proved more efficient.

With regard to best practices, as mentioned in the previous sections, due to the flexible

nature of Wikidata's ontology and the fact there is no content standard or metadata application

profile guiding users to describe a resource in a particular way, the data models from PCC

participants proved to be the primary resource for the Pilot team; in particular, the Pilot team

greatly relied on application profiles created by Stanford University and University of

Washington. This documentation served as learning resources to create the team's own data

model. They represented examples of similar items or projects that the library community has

worked on and helped catalogers accustomed to working with relatively well-defined cataloging

rules, practices, and standards become more comfortable in editing in Wikidata. Equally, they

helped standardize description expectations within the Pilot.

Even as Pilot team members sought to adapt existing application profiles, however, there

were still areas of discussion when it came to deciding what application profiles to adapt from,

and which specific aspects of application profiles to adapt versus creating our own guidelines to

accommodate our specific collections. Notably, one area the Pilot team discussed were the ethics

of describing persons in an open linked data platform, which was particularly important for the

faculty project. Pilot members debated whether availability of information on a faculty member's

website amounted to an approval by the faculty member for using this information in their

Wikidata item, especially if it involved personal details such as date of birth, ethnic group, or

citizenship.

Assessing Productivity and Quality Assurance (QA) Tools

As part of the Pilot project, the UCLA Library Pilot team experimented with a number of batch

tools related to creating and editing Wikidata items, as well as querying Wikidata using

SPARQL.

Batch searching

Wikidata's search function offers a flexible keyword search experience, with an Advanced Search that allows for filtering by other content on a Wikidata page, as well as filtering by type of page, such as "Main" (items), or "property." Results typically provide the item's label, natural language description, and metadata regarding the number of statements present and the date of the last revision.

Because Wikidata does not require unique labels for its items, Pilot team members found that the human-readable short description of each item became all the more important when reviewing the results list, as items may share the same label, such as the various entities with the name "Hagia Sophia." However, when short descriptions are nondescript or when statements are absent, Pilot team members found it challenging to disambiguate items with the same label from each other. (Figure 4)

Figure 4: A search for "Hagia Sophia" brings up several items with the same label; reading the short description helped the Pilot team differentiate between the items.

Pilot team members particularly encountered this when working on the UCLA faculty project, and found that sometimes there were not enough statements to disambiguate one person from another, such as researchers whose items were created via ORCID data feeds that do not include more than the researcher name as public data. In such cases, Pilot team members had to do additional research to try and associate a faculty member's ORCID with an article that may include institution affiliation, or defaulted to creating a new item, thereby running the risk of creating a duplicate item.

Conversely, some items had only non-English labels, and thus would show up as an item lacking a label in the English-language interface. For catalogers used to working with English as the language of cataloging, this was yet another area where catalogers needed to be careful to not

create duplicate items. That is, without guidelines that ensured an English label would always exist, Pilot team members encountered situations where there was more uncertainty about the completeness of search results in Wikidata.

This uncertainty was also present when Pilot team members used the SPARQL query language, an RDF query language, to query Wikidata, whose data is structured using RDF. SPARQL offers a powerful method to query and retrieve data from the Wikidata knowledge base, without worrying about whether fields are indexed for search, as is the case with OCLC Connexion or a local ILS. Even qualifiers are retrievable via SPARQL, which was helpful, for instance, when Pilot team members wanted to know what items of interest lacked references for statements.

However, as discussed earlier, because Wikidata is a collaborative knowledge graph generated by numerous communities with no consistent data modeling, Pilot team members needed to build increasingly complex queries that accounted for a wide variety of data models. For instance, when recording researchers' academic degrees, there is more than one way to model this data, as mentioned in the previous section about relating entities to one another. Pilot team members quickly found that queries that endeavored to return comprehensive results would need to take into account these various modeling decisions. And SPARQL queries themselves can take time to write and refine, particularly given that this was a new query language for the Pilot team. This uncertainty was combined with the uncertainty of the completeness of the Wikidata database, both in terms of what items exist, as well as how well existing items are described; Pilot team members could not be sure that items didn't show up in a result set because of the query, missing item statements, or because the items didn't exist in Wikidata in the first

place. That is, Pilot members found that they needed to do additional manual review of the results for quality assurance.

Additionally, the broad scope of Wikidata - where agents, subjects, works, and more are all identified as "items" - can make understanding query results difficult. In the MARC cataloging environment, LC authority files take a "pre-filtered" approach by sorting entities into specialized lists for names, genres, demographic groups, subjects, and the like, and every label is formatted according to strict standards. This can make searching these sets of data more efficient by having this established facet with carefully formatted strings. With Wikidata's massively broader knowledge base and no requirements for unique labels, search results can pull in different types of entities, with many entities that may not be relevant making it into query results. The Pilot team relied on the presence of the statement "on focus list of Wikimedia project (P5008)" "WikiProject PCC Wikidata Pilot/UCLA (Q100999455)" to limit query results to items team members were working on; without this statement, Pilot team members would likely need to add further limiting parameters to queries.

Batch loading

The Pilot team also experimented with tools to batch load new and/or edit existing Wikidata items. The Pilot team first started with the Wikidata tool QuickStatements, a tool that can edit Wikidata items via simple text commands with defined syntax.[52] QuickStatements proved to increase efficiency of creating new items with the same statements. For instance, when creating new UCLA faculty items in batch, several statements related to place of employment and department were replicated for all faculty items in the batch; QuickStatements made it easy to quickly create these items with identical starting statements reflecting the faculty members' place

of employment; once the item was created, a team member could then further enhance the item manually with more unique statements. Members working on the buildings project realized a similar efficiency for statements related to UCLA buildings. (Figure 5)
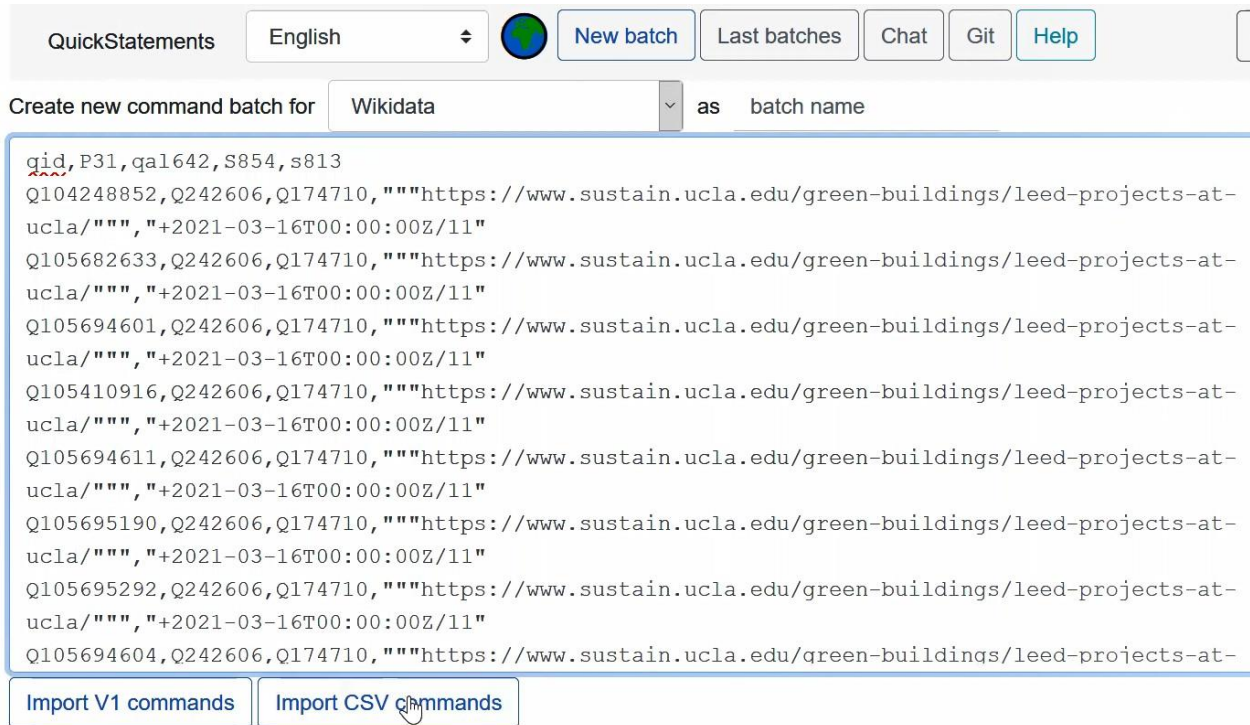
Figure 5: QuickStatements improved production time when UCLA team members needed to add the same statement to a number of items. In this example, team members were able to add an "instance of (P31)" statement, complete with a qualifier statement, as well as references, using the QuickStatements browser interface.

However, because QuickStatements relies on identifiers and specific punctuation to parse, typographical errors could often lead to failed imports. And just like manually editing a Wikidata item, Wikidata item range values need to exist in the Wikidata knowledge base before the statement can be created.

Following QuickStatements, the Pilot team also explored the use of OpenRefine to batch

load new items, and add new statements to existing items. OpenRefine, a popular tool used for

cleaning and manipulating datasets, comes with a built-in Wikidata extension that allows users to

transform tabular data into Wikidata statements.[53] OpenRefine was more well-suited than

QuickStatements for use with existing spreadsheets of data. Unlike in QuickStatements, the cell

values can remain as human-readable labels instead of Wikidata identifiers. Users can run the

Wikidata reconciliation service column by column on a spreadsheet; the service matches values

in the spreadsheet to existing Wikidata items, largely based on shared labels and aliases. Users

can set parameters such as only reconciling against Wikidata items with certain statements, such

as "instance of (P31)" "human (Q5)". Users can also set a confidence level at which OpenRefine

will automatically match a value; otherwise, manual matching is needed. To then import

reconciled values into Wikidata, the user creates a schema to determine how the structure of the

spreadsheet translates to Wikidata statements, including statements with qualifiers.

Similar to QuickStatements, OpenRefine increased the efficiency of creating new items

that have the same statements. However, the reconciliation process and manual matching that

was inevitably part of the workflow took more time than expected, and, depending on the size of

the dataset, did not necessarily increase productivity. The reconciliation process could take the

better part of an hour, even with a relatively small dataset of less than 100 rows, and as Pilot

team members experienced, given Wikidata's proliferation of entities with the same label,

reconciliation against labels resulted in more instances than not of needing manual review and

intervention after the reconciliation process had run.

Overall, the Pilot team found these batch tools worked well for quickly creating the same

set of statements for a specific batch of Wikidata items. With regard to attempts at working with

a more diverse set of statements, particularly with OpenRefine, the time needed for the schema

creation, data cleanup, and reconciliation process did not necessarily result in production time

gains for the size of dataset the Pilot team worked with.

Notifications of changes made to our prior contributions

Wikidata allows contributors to set notification preferences in their profile; one such notification

feature is the "Watchlist", which allows users to watch Wikidata pages, and be alerted through

Wikidata of any changes to items, or whether any new links have been made to the item.

Notifications that some Pilot team members received include alerts for bot-added statements to

items, particularly statements in another language, and items the team created for UCLA faculty

members being linked to scientific articles they authored.

Pilot team members found that the ability to monitor all the changes made to items they

had created and/or edited via the Watchlist was a useful quality control feature to protect these

items from incorrect data or vandalism, as also highlighted by Landis, Wiseman, Smith, and

Stephens.[54] Additionally, Wikidata's notification features are easier to set up and more easily

accessible to individual users than notifications of changes made to, for example, bibliographic

records in OCLC's databases. For the latter, updated records are available when changes are

made to titles in an institution's knowledge base in Collection Manager or to titles in an

institution's WorldCat holdings.[55] However, the setting up and receiving of notifications for

record updates is primarily done at an institutional level; typically, the person/s overseeing an

institutional OCLC account are set up to receive these alerts rather than every individual

cataloger. For authority records in OCLC Connexion, users can see previously superseded

versions; however, there is no instant notification feature to alert users that a record has been

updated. Currently, possible workarounds to the situation include the users' own ILS, which could be configured to capture which fields have changed.

While Pilot team members did not necessarily receive notifications that required taking action on their watched items, the fact that Wikidata provides such a feature perhaps speaks to the transparency with which Wikidata operates with regard to the activity taking place in its knowledge graph.

Quality assurance tools that can bring to light problematic data/Detection of data conflicts when we attempt to use data

Several of the monthly meetings of the PCC Wikidata Pilot[56] held during the Pilot were dedicated to introducing data quality tools, such as ORES, a Wikimedia web service that uses machine-learning to aid in detecting data quality.[57] By and large, however, the Pilot team's experiences with data quality came from constraint flags within item pages, as well as SPARQL queries to detect missing statements in items of interest.

While working on any single item, Pilot team members encountered a number of constraint flags over the course of their Wikidata work. For instance, in addition to the "citation needed constraint (Q54554025)" mentioned earlier, Pilot team members, when developing the application profile for describing UCLA Library oral history collections, discussed the use of the property "recording location (P8546)" to capture where an oral history was recorded. However, in using it with oral history items, the statement resulted in a "subject type constraint (Q21503250)" warning, as the oral history was not considered in the classes "audiovisual work (Q2431196)", "musical release (Q2031291)," "release group (Q108346082)," or "audio track (Q7302866)."[58] Another constraint that assisted with determining potential duplicates was the

"distinct-values constraint (Q21502410)", which identified unique identifiers in use on more than one Wikidata item. (Figure 6)
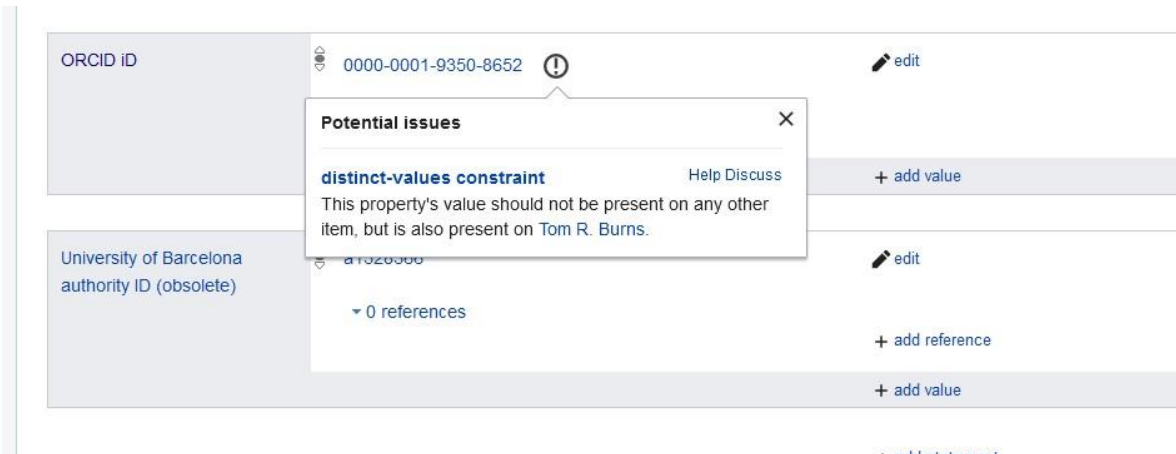
Figure 6: This item for "Tom R. Burns (Q2440014)" contains an ORCID ID that appears on another item with a similar label. For these cases, Wikidata will display a warning indicator that will give a brief explanation of the type of issue - in this case, an ORCID ID should not be found in more than one item.

SPARQL querying was also a method the Pilot team used to identify gaps in descriptions in items of interest. For instance, the team members from the faculty project built queries to identify faculty who may be missing department statements in order to add those statements manually.

Advisable Training Resources

The enormous diversity of informational resources on Wikidata, as well as resources created by peer institutions participating in the PCC Wikidata Pilot, helped Pilot team members get onboarded, increase familiarity and comfort, and find a source of guidance for issues related to

designing application profiles, running queries, and more. The following outlines some of the

resources Pilot team members found helpful throughout the Pilot.

Wikidata Basics. In learning the basics of Wikidata, one of the most valuable resources for the

Pilot team was the LD4 conference held annually every year. Wikidata tutorials by Will Kent

and Robert Fernandez and a presentation on adding references by Hilary Thorsen during the

2020 LD4 Conference provided an important foundation for the team's Wikidata work.[59] The

team also found useful training materials from Wikidata itself. Wikidata provides substantial

resources to onboard new editors to the process of adding items to its knowledge platform. For

example, Wikidata's main page includes links to its community portal, referred to as "complete

starters guide."[60] Among other helpful resources provided by Wikidata were guidelines for

describing living people including alerting users on properties and data that could lead to

violations of privacy as well as information on copyright and licensing information for Wikidata

contributors.[61]

Additional resources the Pilot team members found useful to get started in Wikidata

included training modules for becoming a "Wikidata Professional" from the Wiki Education

Foundation.[62] Resources about training in Wikidata created by librarians themselves were also

particularly useful for the PCC Wikidata library community. These include *The Practical

Wikidata for Librarians* page started in 2019 as a space for gathering resources that would help

librarians edit in Wikidata.[63] MARC catalogers in the Pilot team also found it useful to look

through the mappings between the various fields in MARC 21 authority format and Wikidata

properties created by the PCC Standing Committee on Applications.[64] This mapping provides an

impressive granularity for describing persons in Wikidata and includes relationship designators

that relate agents to one another, and identifies relationships between works, and between work and agent. It helped catalogers in the Pilot team get a sense of how statements in Wikidata compared to MARC fields when describing similar entities.

Wikidata Tools. Once Pilot team members became familiar with editing items in Wikidata, they moved to learning about tools to increase production in Wikidata. For gadgets and user scripts, a presentation by Hilary Thorsen helped Pilot team members more efficiently work on single items in Wikidata.[65] When it came to batch work, Pilot team members consulted Wikidata's documentation on QuickStatements and OpenRefine in order to run collaborative learning sessions with the rest of the Pilot team.[66] And for learning how to query Wikidata, presentations on using the Wikidata Query Service from the LD4 Wikidata Affinity Group, tutorials on Youtube, as well as sample SPARQL queries from Pilot participants - for example, by the University of Nevada, Las Vegas - all helped the Pilot team gain familiarity in using SPARQL.[67] In addition to tools for Wikidata editing, the Pilot team also learned from a presentation by Robert Fernandez on using the Listeria bot to generate lists of items created or edited, [68] which was useful for tracking items created during the Pilot.

Application Profiles and Data Models. Pilot team members found application profiles shared by PCC Wikidata participants on WikiProject pages to be one of the most valuable learning resources throughout the Pilot. These application profiles provided an opportunity to see how library peers navigated the oftentimes confusing Wikidata ontology to suit the goals of projects coming out of the library world. In fact, the importance of application profiles for participants is demonstrated on the project page of Rice University in their statement acknowledging that a fair amount of time will be spent in researching data models.[69] As mentioned earlier, the Pilot team

found the application profiles from the University of Washington and Stanford University particularly helpful, and modeled some of their own application profiles following the data models of these institutions. These application profiles outlined basic, extended, and core descriptions that helped Pilot team members estimate the amount of time needed to be invested in creating items.

Workflow Examples. Pilot team members did not look to other institutions' workflow examples during the Pilot phase; however, several institutions' project pages outlined their workflow that could be helpful for people embarking on similar projects to get a sense of resources that need to be devoted to Wikidata work. Examples include a timeline of tasks from the National Gallery Art Library and workflows from NYU for projects on different genres of items from manga to special collections to digital libraries.[70]

The Pilot Community. Beyond documentation, the Pilot team found that engaging with the PCC Pilot community through an active listserv and monthly meetings provided a valuable learning and training space for Pilot participants. Initial meetings helped Pilot participants become familiar with creating project pages in Wikidata and with the culture of Wikidata; as the Pilot progressed, participants began to present on projects they were involved in, for example: presentations on NACO Lite by the Library of Congress; on using bots to edit in Wikidata (Vanderbot from Vanderbilt University); mapping serials in Wikidata (Penn's backfile on serials project); creating items for electronic theses and dissertations (University of Washington) or Christian hymns (Wheaton College), and also challenges in working with Wikidata (UCLA). Adam Schiff's presentation on creating properties in Wikidata that featured a live demo of

creating a property proposal was also extremely valuable to the Pilot team when deciding to create properties.[71]

To summarize, availability of excellent and numerous training resources on Wikidata as well as engagement and guidance from the broader PCC community helped the Pilot team to locally establish a workflow for the execution of the Pilot project. This involved setting up an internal Confluence wiki for project documentation, creating a local email group, and holding bi-weekly meetings for discussing topics collaboratively determined by the team. These meetings also served as training sessions on Wikidata tools and data modeling and often involved a team member volunteering to learn a tool or Wikidata feature ahead of the session, and then running a collaborative learning session. Starting from the first sessions related to the basics of creating and editing a Wikidata item manually, training meetings covered increasingly complex topics, such as using QuickStatements and OpenRefine, querying the Wikidata knowledge base, and discussing the complexities of creating data models and application profiles for Wikidata items.

Participating in the Community

The Pilot team had the opportunity to engage with the broader Wikidata community primarily via property proposals. The Pilot team can therefore only speak to a few of the learning objectives related to participating in the community: what it is like to propose a new property and when it seems warranted, enrich mappings between Wikidata properties and corresponding lists on loc.gov, persistence of the data, editing data when underlying policies do not match, and how much library goals align with the library community.

What it is like to propose a new property and when it seems warranted

Properties form the backbone of Wikidata's ontology, and are key to defining how an item can be described in Wikidata. Each of the properties created by the Pilot team - "UCLA Space Inventory LocID (P9794)"; "Densho Encyclopedia ID (P9938)"; "issuing agent of work (P9901)"; and "Online Archive of California finding aid ID (P9592)" - sought to fill a particular gap in Wikidata's knowledge graph. Pilot team members involved in these property proposals learned about the process of proposing a property, such as articulating the "motivation" section in the property/discussion or talk page[72] and indicating the domain(s) to which this property can be applied (for example, for the property "issuing agent of work (P9901)" created by the Pilot team, the property discussion page specifies it can be used for "periodicals" or other serial-type items). Pilot team members also experienced the open review by the wider community, during which reviewers may comment, and property creators may address comments by modifying the proposal. Eventually, if a property garners more votes in favor than in opposition, or there appears to be consensus among the responses in this direction, the property is approved and created by a Wikidata administrator, after which the property becomes available for use. The entire process, depending on the support or discussion a proposal generates, can be quite lengthy, or can be wrapped up within a week.

The above is in contrast, for example, to the time-intensive process of approval required for a new Library of Congress Subject Heading proposal, which is decided upon by a relatively small group of individuals in the LC Policy, Training, and Cooperative Programs Division (PTCP). The same is true of the approval process for new MARC values, which must be proposed and justified via the MARC Advisory Committee (MAC), a small committee with

members appointed by the American Library Association's Core division leadership. Whereas membership in Wikidata is open to everyone, and any Wikidata member can comment or vote on a property proposal, membership on bodies such as MAC or PTCP is limited to those selected or appointed to serve on that particular committee or hired by the particular organization.

Can we enrich mappings between Wikidata properties and corresponding lists on loc.gov? What to do when properties do not match

As the "Wikidata and MARC21 authority mappings" created by the PCC Standing Committee on Applications shows[73], not all MARC fields in an authority record had a corresponding Wikidata property at the time of the Pilot. In encountering these gaps, Pilot team members could seek out alternative properties that may convey a similar meaning, choose to not create a statement at all, or go through the property proposal process, as outlined in the section above. Such was the case in the serials project, where Pilot team members found that an important RDA work property - "has issuing agent"[74] - was not yet represented in Wikidata, and went through the proposal process to create the property "issuing agent of work (P9901)" in Wikidata.

Even in cases when a Wikidata property could be identified for a MARC authority field, it is not always a one-to-one mapping. For instance, the RDA relationship designator "music arranger" may be mapped to the Wikidata properties "adapted by (P5202)" or "contributor to the creative work or subject (P767)". Or conversely, many specific relationship designators in RDA may only map to a more generic Wikidata property, such as the relationship designator "dancer" to the Wikidata property "participant (P710)" or "performer (P175)". Wikidata contributors would then need to make a decision about whether this mapping suffices for their needs, or propose a more specific property.

Persistence of the data, in light of the Notability policy & shared editing policy

One of the concerns for the Pilot team about working in an open platform like Wikidata was whether statements that team members created would be removed or changed in any way by other contributors, or even whether whole items would be flagged by other contributors as failing to meet Wikidata's notability policy.[75] The structure of Wikidata includes an open log of changes to Wikidata items, and this transparent version control also allows contributors to track changes and reach out to other contributors if needed should issues arise. Fortunately, as of this article, Pilot team members have not encountered instances of either incorrect statements being added to created items, removal of existing statements, or existing items created as part of the UCLA Library Pilot getting flagged as violating Wikidata's notability criteria. However, this was possibly the result of Pilot team members refraining from adding items that would have failed the notability criteria. For example, even though local cataloging policies at UCLA Library dictate creating a name authority record for any individual appearing in a digitized item, even though they may not be well known, Pilot team members did not create Wikidata items for these lesser known entities because they would likely not meet Wikidata's notability criteria.

Editing data when underlying policies do not match

For Pilot team members coming from a cataloging environment where policies, trainings, and best practices are central to ensuring a high level of quality in bibliographic and authority records, working in Wikidata was a culture shock due to its dearth of policies and best practices. Members sought to harmonize and align with peer contributors in their data modeling and application profiles as much as possible. In editing existing Wikidata items, however, it was clear that some items were created with choices that did not necessarily align with the Pilot

team's application profiles - for instance, when serial title changes were recorded in one Wikidata item instead of having two separate items linked together by a property such as "followed by (P156)"[76] Despite this alternative modeling, the Pilot team decided that unless data were truly inaccurate, or very clearly an error, such differences in modeling choices did not warrant removal of statements, nor did it preclude Pilot team members from modeling data according to the team's application profiles.

How much library goals align with community and can be advanced

The PCC launched the Wikidata Pilot to experiment with the idea of identity management in this open platform. Within the time period of the Pilot, team members found Wikidata to be an incredibly large hub for identifiers, even larger than a registry such as VIAF, able to map identifiers for a particular person or organization from domains beyond libraries, museums, and archives. The data contained in a Wikidata item may indeed go beyond what catalogers capture in the records they create, and the possibilities of incorporating that data into OPACs and discovery layers is exciting.

However, Pilot team members also found that disambiguation could still be a problem in Wikidata. As mentioned previously, when an item does not contain any statements, it can be hard to distinguish between it and an item with the same label, and similar lack of statements. For instance, faculty members who share the same names and also have set their ORCID data privacy to only display their ORCID identifiers resulted in Wikidata items that were incredibly difficult to disambiguate, as the only information available was a shared name, and an ORCID identifier that on its own, did not assist in disambiguation.

Additionally, the notability policy mentioned in the previous objective may lead to people and organizations being removed from Wikidata, meaning that there may always be some subset of bibliographic information that cannot be represented in Wikidata.

Post-Project Remarks and Reflections

 The UCLA Library Pilot team found participating in the PCC Wikidata Pilot to be an incredibly rewarding experience. From the opportunity to learn new skills, to doing research into entities such as UCLA buildings, or contributing information about UCLA Library's special collections to a more public platform for greater discovery, the Wikidata Pilot offered something of interest for all participants. The Pilot experience provided an opportunity for team members to reflect on the implications of working with community sourced data, its effect on discovery, the viability and appropriates of Wikidata serving as a complementary or additional tool for cataloging work, and also ethical issues of entering data into an open platform such as Wikidata. While not directly tied to the learning objectives, the authors feel these reflections can provide additional perspectives on the use of Wikidata as a cataloging tool. For example, a recurring concern that the Pilot team had revolved around the sparse rules, standards, and guidelines around creating and editing Wikidata items. While there was freedom in being able to create application profiles that fit the Pilot team's projects, this was also a major area of uncertainty. In particular, Pilot team members questioned the implications for metadata management and discovery in working alongside contributors from non-cataloging backgrounds, who may model relationships differently, and generally engage in different practices that can affect the ways that catalogers may be able to use the data. This concern also points to the role of the comprehensive cataloging training often needed in the cataloging community to contribute to a shared database, which

builds trust in the data being created. Of course, there is the possibility that quality metadata contributed to Wikidata by the library community in substantial amounts might influence other users to generate similar high-quality metadata for this platform.

Another area of concern was the idea of Wikidata as an additional registry. As enjoyable as working in Wikidata was, Pilot team members found working in Wikidata and NACO simultaneously to be duplicative, and therefore time-consuming. Pilot team members found themselves questioning whether the same entities should reside in both Wikidata and library authority files, and when might one registry be more appropriate to use than another. And if contributing data in Wikidata becomes part of the future of cataloging, would NACO-level expertise be needed for creating items in Wikidata?

Like many of the Pilot participants, the UCLA Library Pilot team was interested in the potential for increased discovery of UCLA Library's unique collections and agents associated them. However, assessing the impact Wikidata items can have on making library collections more discoverable was beyond the scope of this Pilot. , Whereas many studies[77] about the library catalogs' role in information retrieval or resources discovery have been done, the benefits of linked data for discovery are still being explored by library communities. For example, OCLC partnered with several libraries in a CONTENTdm Linked Data Pilot project to investigate the methods — and feasibility — of transforming metadata into linked data to improve the discoverability of digitized cultural materials.[78] The PCC Share-VDE (SVDE) PCC Data Pool Evaluation Task Group[79] attempted to review the accuracy and completeness of available PCC bibliographic data in the Share-VDE catalog, a shared discovery environment based on linked data.[80] The authors of this article hope future research can likewise focus on the impact of the Wikidata Pilot for user communities' search and discovery experience.

.

Finally, while Wikidata offers some solutions to issues the cataloging community has been grappling with, such as offering flexibility with labels and aliases to be more responsive to changes in terminology, the Pilot team also realized that linked open data poses its own ethical issues, issues that the Pilot team grappled with over the course of the Pilot without necessarily coming to firm conclusions. Data on Wikidata are open in a way that information in the LC NAF traditionally has not been. Just because information is made public, even by the person themselves, does not imply they consented to having that information put into a linked open database that can then be pulled into and changed by unknown downstream systems. How can metadata professionals ensure that their Wikidata work is informed by ethical principles that do not bring unintentional harm to people?

Conclusion

Looking back at the Pilot experience through the lens of the PCC learning objectives, the authors believe that even though the UCLA Library Pilot was not initially designed around these objectives, its assessment against the objectives can be valuable for the PCC's decision to use Wikidata as a viable tool for identity management in the context of cataloging information resources.

As catalogers experimenting with this platform, it was valuable to assess Wikidata with regards to other registries and databases catalogers are used to working with; it supports justifying continued resource allocation for this work. For the UCLA Library, the Pilot experience proved it was relatively easy and fast to get started in Wikidata given the multiple learning resources available on the web in addition to the resources put up by the PCC

community itself. At the same time, there were aspects of Wikidata that required training, and slowed down the creation or editing of Wikidata items. Working with structured data was different from creating a MARC record, where textual data can simply be keyed in; as a result, the Pilot team found that there was often a need to create additional entities to describe an item, which could add significant amounts of time to the production rate. Similarly, while the ability to use batch editing tools in Wikidata provided an option to speed up production, especially when creating items of a similar nature such as faculty items, this was not necessarily the case when describing distinct items in our library collections. Additionally, contributors to Wikidata are not necessarily from galleries, libraries, archives, and museums (GLAM); thus, the data contributed may not necessarily be consistent with library metadata needs.

With regard to participating in the community, the experience of the Pilot was beneficial in multiple new ways. Pilot team members engaged with colleagues both across RAMS and outside of RAMS, often for the first time. The Pilot was also possibly the first time that catalogers had broken out of the traditional silos often associated with technical services and started working in a knowledge database where their practices of metadata description were constantly being challenged and modified by a larger community of users. When Pilot team members submitted property proposals, for instance, they had to consciously make sure that terminology was understandable and applicable for all users, not just catalogers or metadata professionals, as proposals are vetted by the larger community. Working in Wikidata also showed that descriptions in an open environment can be constantly edited and added to by a wider community because anyone, including bots, can edit Wikidata. While the Pilot team did not encounter issues with vandalism or items being populated with problematic statements, given the lack of control over an open knowledge base, the unprotected and not necessarily reliable

data was a generally concerning issue for NACO catalogers used to protected and trustworthy data. Nevertheless, these constantly evolving connections and diverse and inclusive data are what allow for richer linkages of data than a siloed catalog or name authority file/registry normally would.

On the other hand, as mentioned earlier, Wikidata's notability policy appeared in some cases to serve as a barrier for providing visibility to certain entities. As one Pilot member noted, well cataloged online collections with in-depth description of their subjects "can become unique doorways to information about underrepresented individuals and cultural groups." However, such information seems risky to add into Wikidata; items created for these little-known entities would likely be deleted because of Wikidata's different notability standards.

There is indeed still much to explore with regards to Wikidata and cataloging work. For example, the Pilot team at a large research university dedicated considerable time and resources to participate in the PCC Wikidata Pilot. This might be challenging for smaller institutions with limited resources. A future research paper could probably address the distinction and compare the Pilot's assessment across the various institutions involved in the Wikidata Pilot and beyond.

However, there continues to be enthusiasm for the platform. In 2022, after the Pilot concluded, the PCC endorsed the LD4 Wikidata Affinity Group as "an important venue for Pilot participants to continue working in community for presentations and regular skill-building meetings."[81] During the Pilot, the PCC had also explored the idea of a SACO Wikidata Funnel, where members would create in Wikidata subject terms that could then be added to MARC fields in both bibliographic and authority records with a subfield $2 indicating Wikidata as the term source. These terms, referred to as SACO-Lite subjects, were meant to describe topics not

covered or adequately represented in LCSH or other LC vocabularies and when submitting a

SACO proposal is not feasible. While the project was suspended later on during the Pilot, the

idea of SACO-Lite, enabled by Wikidata, continues to hold promise.

Finally, with regard to collaborations beyond RAMS, Pilot team members found

opportunities to work on novel Wikidata projects through the Pilot experience. Such projects

include engaging with fellow Middle East Librarians Association (MELA) members to use

Wikidata for Middle Eastern-related topics, which resulted in two conference presentations;

enhancing Wikidata items related to medieval Islamic technology work, and creating links

between translations and works; and joining the Wikidata Religion & Theology Community of

Practice.[82]

As this analysis of the PCC learning objectives with regards to the UCLA Library

Wikidata Pilot demonstrates, Wikidata provides an exciting opportunity for catalogers to engage

with a wider community of stakeholders in creating metadata that is dynamic, open to the

possibility of constant changes, and representative of an increasingly diverse future. Wikidata

has the potential to be used as an identity management tool, especially in its agility in making

library data more responsive to community needs; however, the fluidity in its data model, ability

for its data to be easily edited and a somewhat narrow definition of notability needs to be taken

into account when considering its authoritativeness as an information source. Perhaps a PCC

instance of Wikibase.cloud, a cloud-based service that allows users to run an instance of the

technology underlying Wikidata, with more control over the data itself, could potentially address

some of these concerns. In the meantime, at UCLA Library, Pilot team members have continued

to incorporate Wikidata in cataloging work, such as adding Wikidata identifiers to LC name

authority records and adding LC name authority identifiers to Wikidata, and have also engaged

in national projects that explore the use of Wikidata in library work. Yet, this engagement is inconsistent among Pilot team members depending on their existing workload following the Pilot's conclusion. The authors hope the cataloging community will continue to work in Wikidata and look toward the PCC for setting the next steps for further development in this direction.

Acknowledgments

---

[1] Program for Cooperative Cataloging, "PCC (Program for Cooperative Cataloging) Strategic Directions January 2018-December 2021 (Extended to December 2022)," accessed September 27, 2023, https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2022.pdf.

[2] PCC Identity Management, "Wikidata Pilot," LYRASIS Wiki, accessed September 19, 2023, https://wiki.lyrasis.org/display/pccidmgt/Wikidata+Pilot.

[3] Throughout the article, the phrase "Pilot participants" generally refers to participating institutions of the PCC Wikidata Pilot, and the phrase "Pilot team members" generally refers to UCLA Library participants specifically.

[4] Program for Cooperative Cataloging, "Vision, Mission, and Strategic Directions (January 2015-December 2017)," accessed September 27, 2023, https://www.loc.gov/aba/pcc/about/PCC-Strategic-Plan-2015-2017.pdf.

[5] Program for Cooperative Cataloging, "PCC (Program for Cooperative Cataloging) Strategic Directions January 2018-December 2021 (Extended to December 2022)."

[6] Participants in the PCC's NACO program contribute authority records for names, uniform titles, and series to the Library of Congress/NACO Authority File. Library of Congress-Program for Cooperative Cataloging, "About NACO," accessed May 22, 2023, https://www.loc.gov/aba/pcc/naco/about.html.

[7] Program for Cooperative Cataloging, "PCC ISNI Pilot Home," LYRASIS Wiki, accessed September 19, 2023, https://wiki.lyrasis.org/display/PCCISNI

[8] PCC Identity Management, "URIs in MARC Pilot," LYRASIS Wiki, accessed September 19, 2023, https://wiki.lyrasis.org/display/pccidmgt/Wikidata+Pilot.

[9] MARC or Machine-Readable Cataloging, is a technical standard for the representation of bibliographic and related data; it is developed and maintained by the Library of Congress in consultation with various user communities. See MARBI, "The MARC 21 Formats: Background and Principles", Library of Congress, accessed May 25, 2023, https://www.loc.gov/marc/96principl.html.

[10] PCC Task Group on Identity Management in NACO, "PCC Identity Management Home," PCC Identity Management - LYRASIS Wiki, accessed March 29, 2023, https://wiki.lyrasis.org/display/pccidmgt/PCC+Identity+Management+Home.

[11] See, SD2, Section 2.2, Program for Cooperative Cataloging, "PCC Strategic Directions."

[12] Wikidata, "Main Page," accessed September 27, 2023, https://www.wikidata.org/wiki/Wikidata:Main_Page

[13] Program for Cooperative Cataloging, "Wikidata:Wikiproject PCC Wikidata Pilot," accessed March 29, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot. The NACO-Lite approach prioritizes the creation of identifiers over the construction of text strings, thereby freeing up resources needed to train catalogers to construct unique headings. See John J. Riemer, "Unpacking the Meaning of 'NACO Lite,'" *Technicalities* 38, no. 3 (2018), https://escholarship.org/uc/item/3ss1t4xx.

[14] Riemer, John and Durocher, Michelle. "The Program for Cooperative Cataloging Wikidata Pilot." Webinar from LD4 Wikidata Affinity Group, virtual, June 16, 2020. https://docs.google.com/document/d/1z1SSAp4c4tftOGW3BbJ6Fxfd8oRIhfzveh0zjebDhkk/edit?usp=sharing

[15] PCC Task Group on Identity Management in NACO, "Wikidata Pilot Meeting Notes, Slides & Recordings," PCC Identity Management - LYRASIS Wiki, accessed March 29, 2023, https://wiki.lyrasis.org/pages/viewpage.action?pageId=187174065.

[16] Program for Cooperative Cataloging, "PCC Joint Operations Committee Meeting (OpCo), 2022 - Library of Congress," accessed March 29, 2023, https://loc.gov/aba/pcc/documents/OpCo-2022/Agenda-OpCo-2022.pdf.

[17] Participating in the PCC's Subject Authority Cooperative Program allows member institutions to submit proposals and changes to the Library of Congress's various subject and genre/form vocabularies as well as classification schedules. Library of Congress. Program for Cooperative Cataloging, "SACO - Subject Authority Cooperative Program," accessed May 22, 2023, https://www.loc.gov/aba/pcc/saco/index.html.

[18] Gary Strawn, "Authority toolkit: create and modify authority records," accessed May 22, 2023, https://files.library.northwestern.edu/public/oclc/documentation/.

[19] University of California, Los Angeles, "Wikiproject PCC Wikidata Pilot/UCLA," Wikidata, accessed March 29, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/UCLA.

[20] The annual LD4 Conference has a separate track dedicated to Wikidata. See conference schedules available at, "Conferences," LD4 Community Site, accessed March 29, 2023, https://sites.google.com/stanford.edu/ld4-community-site/conferences.

[21] University of Washington Libraries, "Wikiproject PCC Wikidata Pilot/University of Washington," Wikidata, accessed March 29, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/University_of_Washington.

[22] See Aim and Scope of these libraries' respective Wikidata Pilot projects: Texas State University Libraries, "Wikiproject PCC Wikidata Pilot/Texas State University Libraries," Wikidata, accessed March 29, 2023,

https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Texas_State_
University_Libraries; San Diego State University, "Wikiproject PCC Wikidata Pilot/San
Diego State University," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/San_Diego_
State_University; Brandeis University, "Wikiproject PCC Wikidata Pilot/Brandeis
University," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/BrandeisUni
versity.

[23] Library of Congress, "Wikiproject PCC Wikidata Pilot/LC-BIBFRAME-Wikidata-Project,"
Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/LC-
BIBFRAME-Wikidata-Project; University of Utah, Marriott Library, "Wikiproject PCC
Wikidata Pilot/University of Utah," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/University_o
f_Utah; Harry Ransom Center, "Wikiproject PCC Wikidata Pilot/Harry Ransom Center,"
Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Harry_Ranso
m_Center; Casalini Libri, "Wikiproject PCC Wikidata Pilot/Casalini Libri. SVDE-
Wikidata Interaction," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Casalini_Lib
ri._SVDE-Wikidata_Interaction.

[24] Illinois State University, "Wikiproject PCC Wikidata Pilot/Illinois State University,"
Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Illinois_State
_University.

[25] IUPUI, "Wikiproject IUPUI University Library," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_IUPUI_University_Library.

[26] University of Delaware, "Wikiproject PCC Wikidata Pilot/Participants/University of
Delaware," Wikidata, accessed March 29, 2023,
https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Participants/
University_of_Delaware.

[27] Philadelphia Museum of Art, "Wikiproject PCC Wikidata Pilot/Philadelphia Museum of Art Library and Archives," Wikidata, accessed March 29, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Philadelphia_Museum_of_Art_Library_and_Archives.

[28] PCC Task Group, "Wikidata Pilot Meeting Notes."

[29] For a breakdown of survey results, see PCC Task Group on Identity Management in NACO, "Identity Management Advisory Committee & Wikidata Projects," PCC Operations Committee Meeting, May 6, 2022, https://www.loc.gov/aba/pcc/documents/OpCo-2022/Wikidata.pdf.

[30] The PCC provides specialized training, often in conjunction with the Library of Congress, for its member institutions participating in its programs such as NACO and SACO. See Program for Cooperative Cataloging/Library of Congress, "About the PCC," accessed March 30, 2023, https://www.loc.gov/aba/pcc/about.

[31] While items in Wikidata can be anything within the scope of human knowledge, from concepts to events to well-known figures, Wikidata items need to be "notable," determined through having a corresponding page on a Wikimedia site, being a clearly identifiable material or conceptual entity, or fulfilling a structural need within Wikidata's knowledge graph. Wikidata's guidance for creating and editing items are available through its various Help Pages. See "Help:Items," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Help:Items.

[32] In Wikidata's data model, a statement is composed of an item and a property-value pair; for example, the item for "Earth (Q2)" can have a statement with the property "highest point" (P610) having the value "Mount Everest" (Q513). See "Help:About Data," Wikidata, accessed October 2, 2023, https://www.wikidata.org/wiki/Help:About_data.

[33] Wikidata's guidance for creating and editing Items are available through its various Help Pages. See "Help:Items," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Help:Items.

[34] The 670 field in a MARC name authority record is used to justify the authorized access point and cross references present in the record. It may also contain additional information necessary to identify the entity covered by the record and its relationships to other entities

in the authority file. Program for Cooperative Cataloging, *NACO Participants' Manual*, 2020, https://www.loc.gov/aba/pcc/naco/documents/NACOParticipantsManual.pdf, p. 70.

[35] Property constraints are rules that guide how properties should be used in Wikidata. Certain properties, such as "date of birth (P569)", have a "citation needed constraint (Q54554025)" specifying that statements with this property should have at least one reference. "Help:Property Constraints Portal/Citation Needed," Wikidata, accessed May 23, 2023,

https://www.wikidata.org/wiki/Help:Property_constraints_portal/Citation_needed.

[36] For example, Mr. Ibrahembot, operated by a Wikidata user, adds descriptions in languages such as Arabic to items that may not even have labels. See revision history of Wikidata item Q12217557 showing that this bot added descriptions in eight languages. "Revision History of (Q12217557)," Wikidata, accessed May 23, 2023, https://www.wikidata.org/w/index.php?title=Q12217557&action=history. See also the bot's page in Wikidata, "User:Mr.Ibrahembot," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/User:Mr.Ibrahembot.

[37] "Wikidata-l," *Wikidata-l*, April 22, 2015, https://lists.wikimedia.org/hyperkitty/list/wikidata@lists.wikimedia.org/thread/WHR5ZY TUCNNZB6UOUH44P2MR4HWTCLKG/.

[38] Anchalee Panigabutra-Roberts, Steve Baskauf, and Iman Dagher, "Working with VanderBot to Add Multilingual Content (in English and Arabic) to Wikidata," in UT Libraries Faculty: Other Publications and Presentations, 2022, https://trace.tennessee.edu/utk_libfac/12.

[39] See, for example, the 6XX Notes fields in MARC authority records that contain a variety of unstructured information from biographical notes to general notes for the public. Library of Congress, "667-68X - Notes-General Information," MARC 21 Format for Authority Data: 667-68X, accessed September 20, 2023, https://www.loc.gov/marc/authority/ad66768x.html.

[40] "Help:Sources," Wikidata, accessed September 21, 2023, https://www.wikidata.org/wiki/Help:Sources.

[41] University of Washington Libraries, "Wikidata:WikiProject PCC Wikidata Pilot/University of Washington/Application Profiles/Faculty and Staff," Wikidata, accessed March 29, 2023,

https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/University_o
f_Washington/Application_Profiles/Faculty_and_Staff

[42] Piscopo and Simperl discuss how Wikidata does not have a predefined formal ontology and
that its ontology may change quickly, a result of the collaborative editing and
maintenance of its knowledge graph by its community of editors, both humans and bots.
Alessandro Piscopo and Elena Simperl, "Who Models the World? Collaborative
Ontology Creation and User Roles in Wikidata," *Proceedings of the ACM on Human-
Computer Interaction* 2, no. CSCW (2018), https://doi.org/10.1145/3274410.

[43] See for examples items created for UCLA faculty members as part of the UCLA Library
Wikidata project and outside of the project: "Janice L Reiff," Wikidata, accessed March
30, 2023, https://www.wikidata.org/wiki/Q102252631; "Aditi Lahiri," Wikidata,
accessed March 30, 2023, https://www.wikidata.org/wiki/Q357592.

[44] See "Instance Of," Wikidata, accessed March 30, 2023,
https://www.wikidata.org/wiki/Property:P31; "Subclass Of," Wikidata, accessed March
30, 2023, https://www.wikidata.org/wiki/Property:P279.

[45] As noted by Piscopo and Simperl, Wikidata's collaborative and loosely designed ontology
"means, among others, that Wikidata does not formally distinguish between items that are
classes, and items that are entities. One study used properties P31 (instance of) and P279
(subclass of) to explore common issues in the Wikidata taxonomy, attributable
principally to the misuse of these two Properties. " Piscopo and Simperl, "Who models
the world," p. 4. For example, see the item for "Assuria Hermitage High-Rise
(Q113996227)," Wikidata, accessed March 30, 2023,
https://www.wikidata.org/wiki/Q113996227. This item describes the Hermitage as a
building type (using "subclass of (P279)") instead of an instance of a building, although
the item also uses the property "instance of (P31)" indicating the item is an instance of a
residential tower, which is itself a subclass of residential buildings.

[46] "Help:Property Constraints Portal," Wikidata, accessed March 30, 2023,
https://www.wikidata.org/wiki/Help:Property_constraints_portal.

[47] See items: "The New York Times," Wikidata, accessed March 30, 2023,
https://www.wikidata.org/wiki/Q9684; "The Black Law Journal," Wikidata, accessed
March 30, 2023, https://www.wikidata.org/wiki/Q100319362; "National Black Law

Journal," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Q15756635.

48 Per the LC Subject Headings Manual instruction sheet H 1334, "Generally, when a building is known by the same name as the corporate body that occupies it, a corporate name heading is established for the body and that name heading is used as a subject heading for works about either the organization or the building." Library of Congress, "Buildings and Other Structures - H 1334," in Subject Headings Manual, accessed March 30, 2023, https://www.loc.gov/aba/publications/FreeSHM/H1334.pdf.

49 "Help:Merge," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Help:Merge.

50 "Help:Requests for Deletions," Wikidata, accessed March 30, 2023, https://www.wikidata.org/w/index.php?title=Help%3ARequests_for_deletions%2Fen&oldid=1492187993.

51 "Help:Merge," Wikidata, accessed March 30, 2023.

52 "Help:QuickStatements," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Help:QuickStatements.

53 "Tools/OpenRefine/Editing/Tutorials/Basic Editing," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing/Tutorials/Basic_editing.

54 Cliff Landis, Christine Wiseman, Allyson F. Smith, and Matthew Stephens, "Ganch: Using Linked Open Data for Georgia's Natural, Cultural and Historic Organizations' Disaster Response," The Code4Lib Journal, February 10, 2021, https://journal.code4lib.org/articles/15576.

55 "About Worldcat Updates," OCLC , accessed May 23, 2023, https://help.oclc.org/Metadata_Services/WorldShare_Collection_Manager/WorldCat_updates/About_WorldCat_updates_through_Collection_Manager/About_WorldCat_update

56 PCC Task Group, "Wikidata Pilot Meeting Notes."

57 See for example, the presentation on June 9, 2021, by Lydia Pintscher and Amrutha Chandra on data quality in Wikidata available on the PCC Identity Management Wikidata Pilot site. PCC Identity Management, "Wikidata Pilot Meeting Notes."

[58] Subject type constraint specifies the item described by a property should be a subclass or instance of a given type. "Subject Type Constraint," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Q21503250.

[59] For a list of presentations on Wikidata at the LD4 2020 Conference, see "Wikidata - 2020 LD4 Conference on Linked Data in Libraries Schedule," accessed March 30, 2023, https://ld42020.sched.com/overview/type/Wikidata.

[60] "Community Portal," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Wikidata:Main_Page

[61] "Living People," Wikidata, accessed September 21, 2023, https://www.wikidata.org/wiki/Wikidata:Living_people; "Licensing," Wikidata, accessed September 21, 2023, https://www.wikidata.org/wiki/Wikidata:Licensing.

[62] Wiki Education, "Wikidata Professional Development Training Modules," Wikidata Professional Development Training Modules - Wiki Education Dashboard, accessed March 30, 2023, https://dashboard.wikiedu.org/training/wikidata-professional.

[63] "Wikiproject Linked Data for Production/Practical Wikidata for Librarians," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_Linked_Data_for_Production/Practical_Wikidata_for_Librarians.

[64] PCC Standing Committee on Applications, Wikidata and MARC21 authority mappings, accessed May 23, 2023, https://docs.google.com/spreadsheets/d/1d4w10m6ChNshX8fiaqackiLlQvNt0K2fSlkY9XuKgb0/edit#gid=1077502018.

[65] Hilary Thorsen, "Wikidata: Gadgets & User Scripts," LD4 Wikidata Affinity Group Calls, July 16, 2021, https://docs.google.com/presentation/d/1u2Jx5LbaTX6Q3_e9UBy6XAdkAGzKRZG2/edit#slide=id.p1.

[66] "Help:QuickStatements," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Help:QuickStatements; "Tools/OpenRefine/Editing/Tutorials/Basic Editing," Wikidata, accessed May 23, 2023,

https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing/Tutorials/Basic_editing.

67 University of Nevada, Las Vegas, "Wikiproject PCC Wikidata Pilot/UNLV," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/UNLV#SPARQL_Queries.

68 Robert Fernandez, "Listeria," LD4 Wikidata Affinity Group Calls, September 8, 2020, https://commons.wikimedia.org/wiki/File:Listeria_presentation_LD4_Wikidata_Affinity_Group_2020.pdf.

69 Fondren Library, Rice University, "Wikiproject PCC Wikidata Pilot/Fondren Library, Rice University," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Fondren_Library,_Rice_University.

70 National Gallery of Art Library, "Wikiproject PCC Wikidata Pilot/National Gallery of Art Library," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/National_Gallery_of_Art_Library.; New York University Libraries, "Wikiproject NYU Libraries," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Wikidata:WikiProject_NYU_Libraries.

71 "Wikidata Pilot Meeting Notes, Slides & Recordings - PCC Identity Management - LYRASIS Wiki." Accessed July 7, 2023. https://wiki.lyrasis.org/pages/viewpage.action?pageId=187174065.

72 See for example, the motivation discussion, "Property Proposal/UCLA Library Space Inventory LocID," Wikidata, accessed March 30, 2023, https://www.wikidata.org/wiki/Wikidata:Property_proposal/UCLA Library Space_Inventory_LocID.

73 PCC SCA, "Wikidata and MARC 21 Mappings."

74 "Has issuing agent," RDA Registry, accessed May 25, 2023, http://www.rdaregistry.info/Elements/w/#P10046

75 Per Wikidata's notability policy, an item can be included in Wikidata if it has a sitelink to another Wikimedia page, is mentioned in serious and publicly available references, or

fulfills a structural need in Wikidata's database of items. "Notability," Wikidata, accessed May 24, 2023, https://www.wikidata.org/wiki/Wikidata:Notability.

[76] See for example the Wikidata item for *The New York Times*: "The New York Times," Wikidata, accessed May 23, 2023, https://www.wikidata.org/wiki/Q9684.

[77] See for example the overview of library catalogs as they have changed over time in David Wells, "Online Public Access Catalogues and Library Discovery Systems," *Knowledge Organization* 48, no. 6 (2021): 457–66, https://doi.org/10.5771/0943-7444-2021-6-457.

[78] Bahnemann, Greta, Michael Carroll, Paul Clough, Mario Einaudi, Chatham Ewing, Jeff Mixter, Jason Roy, Holly Tomren, Bruce Washburn, and Elliot Williams. 2021. *Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project*. Dublin, OH: OCLC Research. https://doi.org/10.25333/fzcv-0851.

[79] Program for Cooperative Cataloging, SHARE-VDE PCC Data Pool Evaluation Task Group, accessed October 2, 2023, https://www.loc.gov/aba/pcc/taskgroup/SVDE-PCC-Data-Pool-Evaluation-TG-charge.pdf.

[80] Share-VDE: Linked data for libraries, accessed October 2, 2023, https://wiki.share-vde.org/wiki/Main_Page.

[81] Isabel Quintana, "FW: draft email to PCC list re: endorsement of LD4 Wikidata Affinity Group," PCC Listserv Archives, accessed May 25, 2023, https://listserv.loc.gov/cgi-bin/wa?A2=ind2205&L=PCCLIST&P=55858

[82] "Wikidata Religion & Theology Community of Practice Conversation Group," Atla, accessed May 23, 2023, https://www.atla.com/webinar/wikidata-religion-theology-community-of-practice-conversation-group/