LD4-Wikidata Affinity Group web page
Google Drive Folder
WikiProject Practical Wikidata for Librarians

LD4 Wikidata Affinity Group

2021-03-09
9am PT / 12pm ET / 17:00 UTC / 6pm CET (Time zone converter)
Facilitator: Hilary Thorsen, Resource Sharing Librarian, Stanford Libraries
ld4-wikidata-coord@googlegroups.com

Recording

https://stanford.zoom.us/rec/share/nRQ_0INOCeoKWM0D8FmIjdHe-qNWvHj3MKzuLQT9hl3rFbw8GiP4H57aIby4HvoV.5iZi9r15XpcY0cB6?startTime=1615308641000

Presentation Slides and Other Materials

- Presentation Slides
- Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project

Attendance (if this is your first meeting, please write your name, Wiki* username, institution or affiliation, and interest in Wikidata if you'd care to share)

- Salgo60 Magnus Sälgö Stockholm, Sweden (Wikidata volunteer)
- Sarah Kasten (Univ. Notre Dame)
- Gwyn Johns (Penn State Special Collections)
- Stella Wisdom, British Library
- Lucy Hinnie, British Library, BLWikimedian
- Bob Kosovsky, New York Public Library || user:kosboot
- Georgina Burnett (Wikimedia Deutschland)
- Merrilee Proffitt, OCLC (User:Merrilee)
- Allison Bailund (San Diego State)
- Susan Radovsky (Harvard || Sradovsk)
- Melanie Polutta (Library of Congress)
- Elise Ramsey (Harvard University)
- Diane Shaw (Smithsonian Libraries & Archives)
- Kathleen Boyd (San Diego State)
- Megan Macken (Oklahoma State)
- Monika Soler Correa (University of North Carolina at Charlotte)
- Gina Solares (University of San Francisco)
- David Bigwood (Lunar and Planetary Institute)

- Hanna Bertoldi, Snite Museum of Art, University of Notre Dame
- Violet Fox (University of Nevada, Las Vegas)
- Cory Lampert, (University of Nevada, Las Vegas)
- Paul Bethke (Iowa City Public Library)
- Jeff Mixter (OCLC Research)
- Iman Dagher (UCLA)
- Prema Desai (Stanford University Libraries)
- Timothy Ryan Mendenhall (Columbia University, TrMendenhall)
- Melanie Wacker (Columbia University, mVajen)
- Mairelys Lemus-Rojas (IUPUI)
- Brandie Pullen (NLM)
- Colin Van Alstine (Smith College)
- Kalli Mathios (Teachers College, Columbia University) user: kind_data
- Alex Jung (University of Toronto Libraries)
- Loren Koenig (NYU)
- Jessie Keck (Stanford University)
- Mary Aycock (Texas State University)
- Irene Patrick (State Library of North Carolina)
- Soo Bae (National Gallery of Art Library)
- Jim Hahn (UPenn)
- Hsianghui Liu-Spencer (Carleton College)
- Bruce Washburn (OCLC Research)
- Adam Schiff (University of Washington Libraries)
- Crystal Clements (University of Washington Libraries)
- Huda Khan (Cornell)
- Elena Sillitti (Casalini Libri)
- Davidrlowe (New York Public Library)
- Benjamin Riesenberg (University of Washington Libraries)
- Christine Riggle (Harvard Business School)
- Stef Baldivia (California State University, Chico)
- Greta de Groat (Stanford) user: silentdivas
- John Riemer (UCLA)
- RJ Hardeman (MHLS)
- Ilkay Holt (British Library, UK)
- Rebecca O'Neill (Wikimedia Community Ireland)
- JShieh (Smithsonian Lib Arc)
- Eric Willey (Illinois State University)
- Steve Baskauf (Vanderbilt Heard Libraries), user: baskaufs
- Richard Naples (Smithsonian Libraries and Archives)
- Matthew Haugen (Columbia University Libraries)
- Dave Erlandson (Rice University, Fondren Library)
- Alla Roylance (NYU)
- Lucy Wang (Bridgewater State University)

- Ryan Finnerty (UC San Diego)
- Rhonda Super (UCLA) Semantic Web LibGuide: https://guides.library.ucla.edu/semantic-web
- Paromita Biswas (UCLA)
- Giovanna Fontenelle (WMF, GLAM and Culture team)
- Diane Vizine-Goetz (OCLC Research)
- Karen Li-Lun Hwang, Semantic Lab/Asian American Arts Centre
- Jessica Efron (State Library of NC)
- Cliff Landis (AUC Woodruff Library)
- Alexandra Wong (University of Toronto Libraries)

Agenda

- **Announcements**
    - If interested, subscribe to the LD4 Community list by sending an email to:

      ld4-community+subscribe@googlegroups.com

      You will receive a confirmation email with a link to a subscription page. We hope to be able to use this list for community discussions but also to keep the community informed of community activities and updates, conference information, opportunities, etc. The list is open to all and you will be able to self-subscribe. Please forward this message to anyone you think will be interested.
    - Friday, March 19, 2021 at 10:00am PT / 1:00pm ET / 17:00 UTC / 6:00pm CET (Time zone converter)
    - Wikidata Gadget Usage
    - New browser for Wikidata: Sling Knowledge Browser
    - Wikidata and Correspondence Archives by Yann Ryan
    - Shared Citations--proposal for central reference management database for all Wikimedia projects
    - Wikidata Weekly Summary for March 1st
    - Wikidata Weekly Summary for March 8th
- **Today's Call Topic:** Crossover event with the Discovery Affinity Group! Bruce Washburn and Jeff Mixter from OCLC will talk about their approach to building out the CONTENTdm Wikibase data model outlined in Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project
    - Come back next Tuesday March 16th at 9am PT / 12pm ET / 16:00 UTC / 5pm CET to hear about the project's experimental applications to increase discoverability at the Discovery Affinity Group call
    - https://cornell.zoom.us/j/177198322?pwd=U1N1bVp3K0ZZU3NsaUQzSURkT0JhQT09 (password: pizza)

    - Agenda: https://docs.google.com/document/d/1QgjgfuCaSOtYvxoYNxmXmtaNHon1ponujNRp_7BiExs/edit?usp=sharing
- **Overview of the pilot project**
    - CONTENTdm--digital content management service by OCLC used by libraries, archives, museums
        - Everyone has their own instance and their own collections where they can manage their collections in OCLC's cloud
        - Started in September of 2019 and ended in September 2020
        - OCLC research and 5 partner institutions--The Huntington, U Miami Libraries, Cleveland Public Library, Minnesota Digital Library, Temple University Libraries
    - Areas of investigation
        - Divergent practices and collection assessment
        - Shared data models for diverse collections
        - Machine learning and human intervention
        - Bring in subject matter expertise
        - New discovery tools for finding connections?
        - Paradigm shift of moving to linked data
- **Developing the project's data model**
    - CONTENTDM data very heterogeneous
        - Variety in vocabularies/data dictionaries
    - Single model needed for the project
        - Decided not to work with existing data model--wanted to test hypothesis of sampling current CONTENTdm to find patterns and convergence
            - Model driven by data; could avoid speculation
            - Recreating data model poses own challenges for interoperability
                - Diligent to connect properties and classes creating to existing ontologies
                    - Provide context to other data models being used on the web
    - Metadata analysis began with inventory of CONTENTdm mapping of local data elements to Dublin Core--looking at localized labels institutions applied to those elements
        - Looked for most frequently encountered classes and relationships
        - 13 million records were analyzed to find most widely-used properties for creative works
        - Most frequently used created the taxonomy
        - Data model turned out very similar in size and structure to existing ontologies found on the web
    - Data Modeling Challenges
        - Describing "type" of creative work at 3 levels
        - Distinguishing between instances of concepts and ontological classes
        - Decided to re-evaluate "typiness" of things
            - High level had restricted list--DCMI types to say something a type of

        - Benefits: Easy to do, easy and intuitive high level faceting
    - For more granularity created property classification used--acceptable values were set of 25 classifications--based on work done by Linked Art project derived from Getty AAT vocabulary
    - Created property called “process or format”--add any other classification you wanted to
    - Example--type is image; classification use is Prints; process or format is Postcards
- Distinguishing between instances of concepts and ontological classes
    - Libraries have history of creating controlled vocabularies
        - See terms that can be used as subject headings and classifications
    - Depicts dogs
    - In ontology can have things that are dogs--Type classification of “Dog” for a specific Dog
    - Looked for ways to make the distinction
    - Found convention used in RDF--”isDefinedBy”
    - In description of “Dog” Class, the class itself “isDefinedBy” the Concept of “Dogs”
    - Worked on this with partners
    - How to represent things that blur class and instance of a concept
    - Steve Baskauf: kind of like foaf:focus
        - Jeff--connecting concept to Real World entity, but basically same principle how to connect things with same label, but represent different ontological concepts
- Questions
    - What properties or classes used to create associations between OCLC terms in development and external vocabularies
        - Equivalent class
        - Equivalent property
            - Used both to connect to outside vocabularies
    - Did you use Entity Schemas for better data quality?
        - Jeff Young wrote Wikibase gadget for checking constraints
            - Used SPARQL query to check constraints
            - Each property had their own constraint claims
            - Could use to drive user interface
    - Was the modeling effort complicated by need to make statements about physical item and digital surrogates?
        - Decided to focus on physical--didn’t want to blur the lines between physical and digital surrogate
            - RElying heavily on IIIF for digital technical metadata for items
            - In item metadata descriptions--did frequently see technical metadata as well as administrative metadata
            - Did create mechanism for storing metadata as JSON

                - JSONLD mechanism for including string literals
                - In RDF description could embed JSON record of technical metadata that didnt' want to describe in ontology
        - Intentional absence of having classes in Wikidata seems somewhat incompatible with more formalized defining of an ontology involving classes
            - Manageable challenge
            - Created OWL ontology
            - Go in with clear idea of how to model the data and how to manage it--had to be very mindful
        - Can the gadget be added to Wikidata?
            - Replaced sitelink real estate for embedding Mirador viewer (opensource software for loading IIIF manifest)
            - Would be useful to have
            - Where could it go in Wikidata without being disruptive? What would be the key for determining it should be shown?
            - Had a link to view it in Mirador--some implementation of Mirador--just having the link might be useful
-
-
- **Why Wikibase**
    - OCLC Research familiar with Wikibase from Project Passage
    - Great software for LD investigations
    - Why not Wikidata?
        - Pilot nature--wanted to test new properties and classes and knew would change minds, need adjustment
        - Time needed to propose properties in Wikidata would be too much in the pilot timeframe
        - Reluctance from participants to have data visible in Wikidata
            - Transformed metadata wouldn't be able to reviewed/maintained once shared in Wikidata
    - Debt owed to communities that developed and shared Wikibase with the world--made such a difference of ability to explore
    - Managing Data Model in Wikibase
        - Reused Wikidata templates for new Class and Property proposals
            - Handy for keeping track of prior decisions
        - Talk pages for managing collections
            - Used for entity that described a collection--used as way to link to separate OpenRefine project and list of all fields found in collection with transformation notes and example values
    - Wikibase Ecosystem Advantages
    - Implementing Authority Control

        - Can name something many different ways and many languages and have identifier always be the same number
    - Decreasing cataloging inefficiencies
        - No need to repeat data--can just link to item with description
    - Generating Data Visualizations
        - SPARQL query gives you visualizations of where photos taken in map--very easy
    - Adding Mirador Viewer
        - Make a gadget!
    - Showing contextual information
        - Push LD value proposition and find data elsewhere and bring back to your own item
    - Adding Context
        - Looked for context in Wikipedia and Wikimedia Commons
    - Adding Schema.org in CONTENTdm
    - Summing up
        - Reaffirmed prior lessons and new insights
        - Initial data model informed by other standards provided solid set of classes and properties
        - Model could be expanded as new types of entities and relationships emerged in CONTENTdm metadata
    - Moving forwards
        - New tools for working in CONTENTdm
        - Propose properties in Wikidata
        - Crossover gadgets
    - Jeff Mixter jeff_mixter@oclc.org
    - Bruce Washburn, bruce_washburn@oclc.org
    -
- **Next calls:**
    - Discovery Affinity Group Crossover: Tuesday March 16th at 9am PT / 12pm ET / 16:00 UTC / 5pm CET to hear about the project's experimental applications to increase discoverability at the Discovery Affinity Group call
        - https://cornell.zoom.us/j/177198322?pwd=U1N1bVp3K0ZZU3NsaUQzSURkT0JhQT09 (password: pizza)
        - Agenda: https://docs.google.com/document/d/1QgjgfuCaSOtYvxoYNxmXmtaNHon1ponujNRp_7BiExs/edit?usp=sharing
    - Wikidata Affinity Group Tuesday, March 23, 2021 at 9am PT / 12pm ET / 16:00 UTC / 5pm CET (Time zone converter) with the EthicsLD group on their work developing a checklist for ethics in linked data and forthcoming book

March 2021 LD4 Wikidata Affinity Group Calendar

- **Wikidata Working Hour**
    - Friday, March 19, 2021 at 10:00am PT / 1:00pm ET / 17:00 UTC / 6:00pm CET (Time zone converter)
- **Wikidata Affinity Group Call**
    - Tuesday, March 23, 2021 at 9am PT / 12pm ET / 16:00 UTC / 5pm CET (Time zone converter)
- **Wikidata Working Hour**
    - Monday, March 29, 2021 at 11:00am PT / 2:00pm ET / 18:00 UTC / 8:00pm CEST (Time zone converter)