Lessons from representing library metadata in OCLC Research’s Linked Data Wikibase prototype

Karen Smith-Yoshimura
OCLC Research
About Project Passage

- Ran from December 2017 through September 2018
- 16 OCLC Member libraries participated
- Objective: Evaluate a framework for reconciling, creating, and managing bibliographic and authority data as linked data entities and relationships
- Pilot Wikibase instance included 1.2M entities from matches among Wikidata, VIAF, and WorldCat

Project Passage Ecosystem
Step 1: Search for matches in Wikidata, VIAF, and FAST

<table>
<thead>
<tr>
<th>Wikidata</th>
<th>VIAF</th>
<th>FAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elizabeth Gould</td>
<td>Gould, Elizabeth, 1804-1841, Personal name</td>
<td>Gould, Elizabeth, 1804-1841, Person</td>
</tr>
<tr>
<td>English artist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>illustrator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lithographer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elizabeth Gould</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American neuroscientist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elizabeth Gould Davis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American writer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 2: Retrieve and edit data

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Elizabeth Gould</td>
<td>American neuroscientist</td>
<td>Gould, Elizabeth</td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalan</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 3: Create a new item entity

Elizabeth Gould

American neuroscientist

Also known as

Gould, Elizabeth

In more languages

In English

Description

American neuroscientist

Also known as

Gould, Elizabeth

In French

Description

No description defined

Also known as

Gould, Elizabeth

In German

Description

No description defined

Also known as

Gould, Elizabeth

In Hindi

Description

No description defined

Also known as

Gould, Elizabeth

All entered languages

English, French, German, Hindi

Statements

employer

Princeton University

1 reference

warcru item ID: Q2582896

official website

http://www.web.wisc.edu/
Passage Pilot Partners Use Cases
Project Passage Use Case: Translations

Translations of Sein und Zeit
Being and Time (German: Sein und Zeit) is a 1927 book by the German philosopher Martin Heidegger, in which Heidegger seeks to analyse the concept of Being. This has fundamental importance for philosophy, he thought, because since the time of the Ancient Greeks, philosophy has avoided this question, turning instead to the analysis of particular beings. Heidegger seeks a more fundamental ontology through understanding being itself. He approaches this through seeking understanding of beings to whom the question of being is important, i.e. ‘Dasein’, or the human being in the abstract. Although written quickly, and though Heidegger did not complete the project outlined in the introduction, Being and Time remains his most important work. Being and Time has profoundly influenced 20th-century philosophy, particularly existentialism, hermeneutics, deconstruction, and the enactivist approach to cognition. The book is dedicated to Edmund Husserl “in friendship and admiration.”

From dbpedia

Philosophy book by Martin Heidegger

Also known as: Sein und Zeit
Use Case: Postcard Depicting a Person

“Map of Concord, Massachusetts, 1852 by H.F. Walling”
instance of: map
title: Map of the town of Concord, Middlesex County, Mass. (English)
cartographer: Henry Francis Walling
surveyor: Henry David Thoreau
scale: 18,400
coordinate location: 42°27'37"N, 71°20'58"W
depicts: dwelling owned by: Ralph Waldo Emerson
depicts: Obelisk Monument

“civil engineer in Boston, ca. 1852”
instance of: person
occupation: philosopher, poet
death date: 1862

instance of: monument
production date: 1837
location: Concord, Massachusetts

Use Case: Musical Work for an Event

"Nuper rosarum flores (Latin)"
instance of: motet
language: Latin
date of first performance: “1436”
composer: Guillaume Dufay
has part: Terribilis est locus iste
commemorates:

"Choral musical composition"

"Netherlandish Renaissance composer"
instance of: person

instance of: Gregorian chant

"Consecration of Florence Cathedral"
inference:
instance of: dedication
inception: “25 March 1436”
location: Florence Cathedral
commemorates: Brunelleschi’s dome

"Dome of Florence Cathedral"
part of:

consecration -> ceremony -> event

"Santa Maria del Fiore (Cathedral : Florence, Italy)"
architect: Filippo Brunelleschi

Consecration of Florence Cathedral

About

Consecration of Florence cathedral, 25 March 1436, on the occasion of the completion of the dome built under the instructions of Filippo Brunelleschi

Also known as: Dedication of dome (Florence Cathedral) | Dedication of Florence Cathedral | Consecration of dome (Florence Cathedral)

Significant event of

- Florence Cathedral
cathedral
- Brunelleschi's dome
dome of Florence cathedral

Commemorated by

- Nuper rosarum flores
motet composed by Guillaume Dufay for the 25 March 1436 consecration of the Florence cathedral, on the occasion of the completion of the dome built under the instructions of Filippo Brunelleschi
Use Case: Photograph in an Archive

Content

- **instance of**: photograph
  - **photographer**: John W. Mosley
  - **production date**: 3 August 1965
  - **digital representation**: digital.temple.edu
  - **depicts**: event
  - **depicts**: Martin Luther King, Jr.
  - **depicts**: Cecil B. Moore

Institutional context

- **instance of**: collection
  - **photographer**: John W. Mosley
  - **has part**: Photograph

- **instance of**: collection
  - **photographer**: John W. Mosley
  - **industrial**: archive
  - **curator of**: John W. Mosley Photographs
  - **official website**: https://library.temple.edu/collections/blockson
  - **parent organization**: Temple University Libraries

**“Civil Rights demonstration at Girard College, August 3, 1965”**

- **instance of**: demonstration
  - **location**: Girard College
  - **subject**: desegregation
  - **subject**: civil and political rights
  - **sponsored by**: National Association for the Advancement of Colored People

- **“Americanclergyman, activist, and leader in the American Civil Rights Movement”**
  - **birth name**: Michael King (English)

- **“American lawyer”**
  - **instance of**: person

- **“photojournalist who extensively documented the everyday Activities of the African-American community in Philadelphia, Pennsylvania”**
  - **occupation**: photographer

- **“Charles L. Blockson Afro-American Collection”**
  - **instance of**: organization
  - **industry**: archive
  - **curator of**: John W. Mosley Photographs
  - **parent organization**: Temple University Libraries
Sun Yat-sen

孫中山

中华民国的主要创立者，中国国民党创党总理
孙逸仙 | 孙文 | 孙中山 | 孫逸仙 | 孫文

<table>
<thead>
<tr>
<th>语言</th>
<th>标签</th>
<th>描述</th>
<th>别名</th>
</tr>
</thead>
<tbody>
<tr>
<td>中文</td>
<td>孫中山</td>
<td>中华民国的主要创立者，中国国民党创党总理</td>
<td>孙逸仙</td>
</tr>
<tr>
<td>英语</td>
<td>Sun Yat-sen</td>
<td>Chinese physician, politician and revolutionary</td>
<td>Sun I-hsien</td>
</tr>
</tbody>
</table>
### Some other Wikidata labels

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amharic</td>
<td>እን ዓኝ ወን ወን</td>
</tr>
<tr>
<td>Arabic</td>
<td>سون يات سين</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Сун Ятсен</td>
</tr>
<tr>
<td>Bangla</td>
<td>সান ইয়াত-সেন</td>
</tr>
<tr>
<td>Tibetan</td>
<td>ཕུན་བུགས་པ།</td>
</tr>
<tr>
<td>Greek</td>
<td>ΣΟΥΥ ΓΙΑΤ-ΣΕΥ</td>
</tr>
<tr>
<td>Hebrew</td>
<td>יולו-וב-יו-יו</td>
</tr>
<tr>
<td>Hindi</td>
<td>सन यात-सेन</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Сун Ят-сен</td>
</tr>
<tr>
<td>Malayalam</td>
<td>സന്യാത-സെന്യ</td>
</tr>
<tr>
<td>Mongolian</td>
<td>Сун Ятсэн</td>
</tr>
<tr>
<td>Marathi</td>
<td>सून यात-सेन</td>
</tr>
</tbody>
</table>
Lessons Recap

Need for Retriever (import) and Explorer (discovery)

Structured data created with more precision than current library standards

Need to enhance the imported ontology – but how do we share the enhancements? Need interoperability

Multiscript support replaces “language of cataloging” and obviates the need for transliterations
Reflections

- **What entities matter to this object?** Traditional distinction between authority and bibliographic data disappears
- Need “best practices”: What statements, what properties?
- Still need interpretative context: interplay of “structured” and “narrative” data
- Would crowd-sourcing undermine librarians’ authoritative, quality data?
- Local Wikibase instances expedient, but how to interoperate with Wikidata, other instances?
Thank you!

Contact: Karen Smith-Yoshimura
smithyok@oclc.org
@KarenS_Y

oc.lc/linkeddataresearch

Together we make breakthroughs possible.