



Wikidata

Open Linked Data for Library Publishing

Mairelys Lemus-Rojas -  @mlemusrojas  Q50376985

Jere Odell -  @jaireeo  Q37371644

Ted Polley -  Q59158547

Indiana University-Purdue University Indianapolis -  Q1433199



The Unstoppable Force

What is Wikidata?

- A collaboratively edited knowledge base
- Stores structured linked data under a CC0 license
- It can be read by both humans and machines
- Newest project of the Wikimedia Foundation led by Wikimedia Deutschland
- Multilingual central repository with over 56 million items and 6 thousand properties
- It includes a growing collection of open citation data
 - more than 22 million publications and 174 million citations as of January 2019
(<http://wikicite.org/statistics.html>)

Wikidata Basics

- Entities in Wikidata are either **Items** or **Properties**
- Items represent topics, concepts, and objects
- Properties are used to describe data values
- All items and Properties have a unique identifier
 - **Items** identifiers start with a **Q** prefix + a number (called QID)
 - Example: [Q50332388](#) is the unique identifier for the journal item **Chronicles of Health Impact Assessment**
 - **Properties** identifiers start with a **P** prefix + a number
 - Example: [P123](#) is the unique identifier for the **publisher** property
- Items in Wikidata should be:
 - [Notable](#)
 - [Unique](#)
 - [Linked to other items](#) -- this is important!

Wikidata Basics (cont.)

- Items might contain
 - Statements (in the form of property-value pairs)
 - Qualifiers (to provide context and scope)
 - References (to provide sources)
- Properties link *items* to *values*
- Data stored in triples (Subject > Predicate > Object)
 - **Subject** > **Predicate** > **Object**
 - **Item** > **Property** > **Value**
 - **Library Publishing Coalition** > **instance of** > **organization**
 - **Q54911750** > **P31** > **Q43229**

Wikidata Data Model

The image shows a Wikidata entry for Marie Curie (Q7186) with various components annotated. The entry includes a label, a description, and a list of statements. The 'award received' statement is expanded to show qualifiers and references.

label — Marie Curie (Q7186) — **item identifier**

description — Polish-French physicist and chemist
Maria Skłodowska-Curie | Maria Salomea Skłodowska | Madame Curie — **aliases**
▶ In more languages

Statements

property — award received — **value** — Nobel Price in Physics

rank — [Rank indicator]

statement group — [Group indicator]

qualifiers — point in time: 1903
together with: Henry Becquerel, Pierre Curie
prize money: 35,339 Swedish krona

opened references — 2 references

reference URL	http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/index.html
retrieved	3 August 2015
publisher	Nobel Foundation
language of work or name	English
title	The Nobel Price in Physics 1903 (English)

collapsed reference — Willard Gibbs Award

point in time	1921
---------------	------

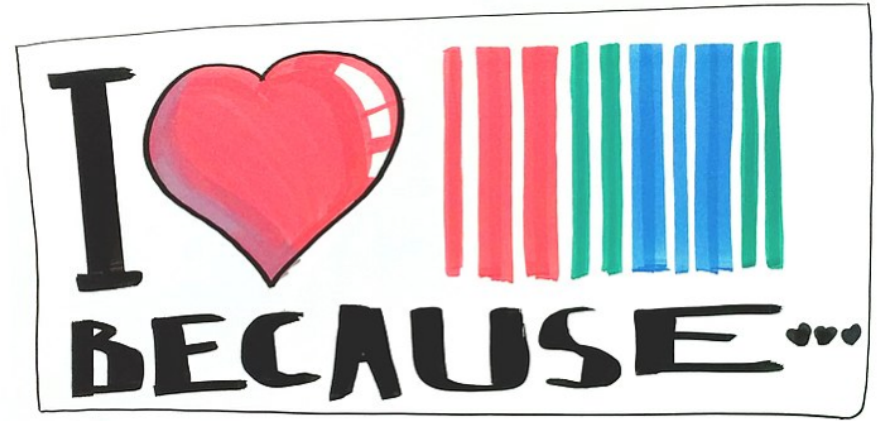
▶ 1 reference

+ add value

File: w.wiki/32q

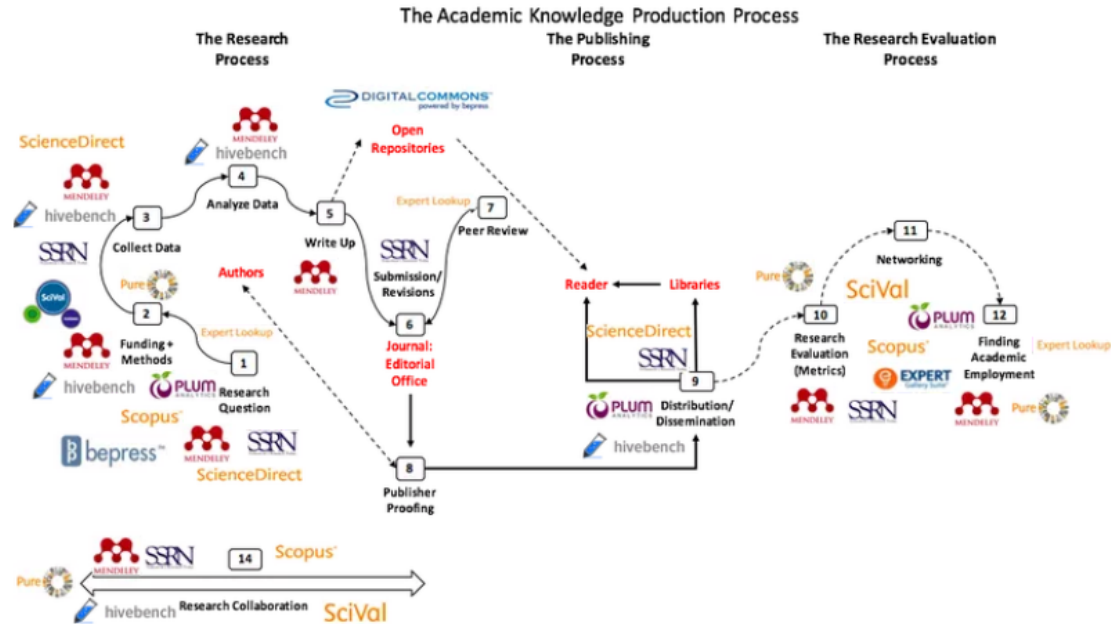
Wikidata and Libraries?

- Common interest in documenting and preserving our history
- Collect, describe, and manage data that would be captured in siloed and paywalled systems
- Aligns with other library-supported projects
- Library data can be reused and enhanced with data in the “wild” (geographic information, popular topics, multilingual sources)
- Join the movement: Set your data free; give it to the commons
- ***IT'S FUN!***



Why should library publishers care?

- **Economic Relevance:** Scholarly publishing is moving from a subscription marketplace to a data marketplace
- **Public Data:** Own the data or feed the monopolies?
- **Open Citations:** Low-barrier, crowdsourced citation network



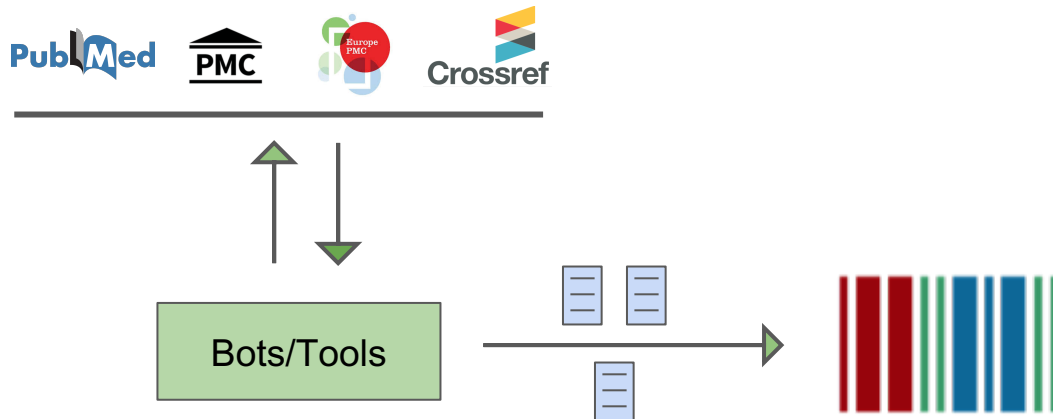
Posada, A., & Chen, G. (2018). Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. In L. Chan & P. Mounier (Eds.), ELPUB 2018.

<https://doi.org/10.4000/proceedings.elpub.2018.30>

Where do most Wikidata bibliographic items come from?

Indirectly through common sources of open publication metadata

- Crossref metadata distribution
- [Crossref open reference distribution](#)
- PubMed, PubMed Central (PMC), and Europe PMC



How can we contribute?

Directly through these methods:

- Method 1: Unique IDs (DOI, PMCID, PMID) → [Source MetaData](#) tool → [QuickStatements](#)
- Method 2: Use the [CSV to QuickStatements tool](#)
- Method 3: Create manual entries directly in Wikidata

What have we done?

- Semi-automatic process using the [Source MetaData](#) tool
 - 332 new journal articles from 7 journals of 9 journals selected
- Manual process
 - IUPUI Open Access Journals
 - 4 new journal items created
 - [Chronicles of Health Impact Assessment \(Q50332388\)](#)
 - [Indiana Libraries \(Q56326567\)](#)
 - [Metropolitan Universities \(Q58618179\)](#)
 - [Journal of Civic Literacy \(Q56653804\)](#)
 - 7 articles ([w.wiki/3Zk](#))
 - 18 References in 1 Article ([Q61986829](#))
 - 5 existing journal items enhanced



Citation graph for [Q61986829](#)

Benefits

- Enables open scientometric and bibliometric research
 - Most Cited Female Authors (w.wiki/3h8)
- Inclusion: Many library-published works are less likely to be included in PubMed and Crossref (common sources for the Wikidata bots and tools)
- Dissemination: One more point of discovery for our publications & a query tool for reporting & bibliographic studies



ex_publisherLabel

- undefined
- American Physical Society
- American Statistical Association
- Armand Colin
- BioMed Central
- Cambridge University Press
- Elsevier
- Emerson College
- Ernest Leroux
- IOS Press
- Lawrence and Wishart
- Lippincott Williams & Wilkins
- MIT Press
- Nature Publishing Group
- Pensoft Publishers
- Public Library of Science
- SAGE Publishing
- Springer Science+Business Media
- Taylor & Francis
- Universitat Autònoma de Barcelona. Departament de Filosofia
- University of Chicago Press
- Walter de Gruyter
- Wiley-Blackwell

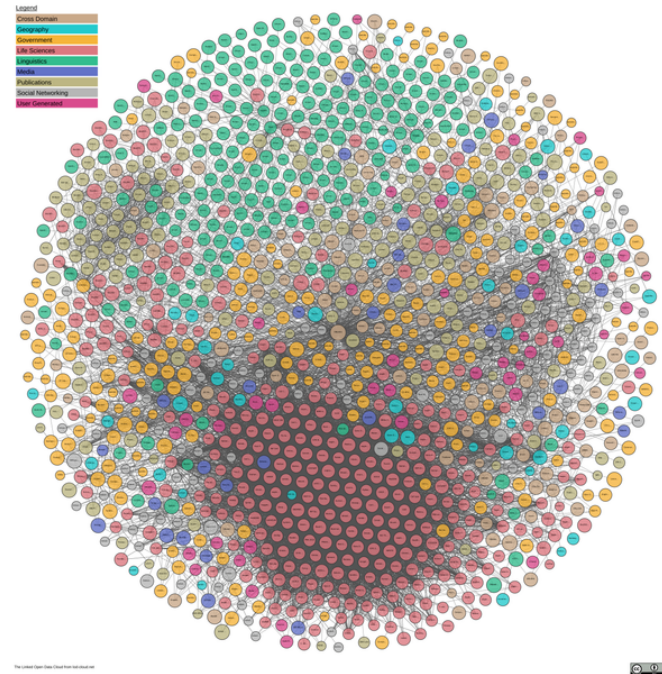
Example Query:

Journals with editors on Twitter
(size by label: Publisher)

w.wiki/3h7

Challenges


- “Wiki” - huh?
- Article reference lists can be messy--what do we do about errors in the published article?
- Is this too much work? Does it “scale”? Is it worth the investment of our time?
- Do commonly used journal publishing systems generate metadata exports in formats that would facilitate reuse in other platforms, like Wikidata? (Would someone please develop this for OJS?)





What's Next

- Developing a workflow to actively contribute to Wikidata as our journals publish new issues
- Investigate the value of contributing related items from our journals (editorials, letters, authors, references, etc.)
- Encouraging other LPC members to contribute their journal metadata to Wikidata and indicate their membership in LPC
 - e.g. **IUPUI University Library** is a **member of Library Publishing Coalition**

IUPUI University Library (Q5975426)...

University Library of Indiana University–Purdue University Indianapolis  edit

Indiana University–Purdue University Indianapolis University Library

member of	 Library Publishing Coalition ...	 edit
	0 references	+ add reference

Q19279214!

