

Wikidata, Wikibase, and a federated ecosystem of structured knowledge for open science

Daniel Mietchen
University of Virginia
@EvoMRI

Dario Taraborelli
Wikimedia Foundation
@readermeter

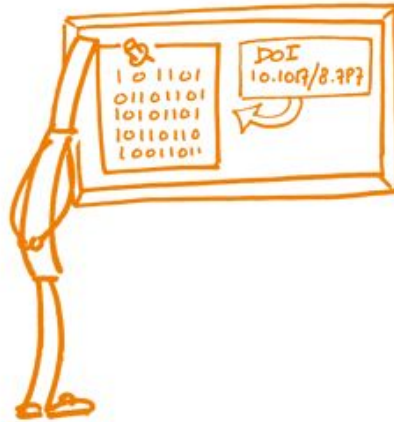
FORCE 2018 • Montreal, 11 October 2018

FAIR data platforms

FAIR DATA PRINCIPLES



FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE



WIKIDATA

A hands-on introduction to Wikidata and WikiCite

Friday, 3:00pm

Friday, October 12 • 3:00pm - 3:30pm



A hands-on introduction to Wikidata and WikiCite

In this session we'll give a hands-on introduction to Wikidata and the creation, curation and extraction of structured data about scholarly works, authors, institutions (part of the WikiCite initiative). You'll learn about the Wikidata data model and make your first edits. You'll learn how to populate a Wikidata entry from a DOI or an ORCID and how to retrieve and visualize data in SPARQL via the Wikidata Query Service. No prior knowledge of Wikidata or Wikimedia projects is required.

Speakers



Daniel Mietchen

Researcher/Scholar/Scientist, Data Science Institute, University of Virginia

- Integrating research workflows with the Web - Engaging the research community and the public with open research workflows - Using open research workflows in educational contexts



Dario Taraborelli

Director of Research, Wikimedia Foundation

Dario is a social computing researcher and open knowledge advocate based in San Francisco. He is the Director, Head of Research at the Wikimedia Foundation, the non-profit organization that operates Wikipedia and its sister projects. His research interests focus on online collaboration... [Read More](#) →

Wikidata is to data what Wikipedia is to text

- All data CC0
- Anybody can contribute
- Covers all domains of knowledge
- Fully version controlled and collaborative
- Integrated with the semantic web via RDF / open APIs
- High performance query engine (SPARQL)
- Stable. Not tied to short-term funding cycles
- Actively developed, full stack is open source
- Active community. Fastest growing Wikimedia project



50 million
entities

550M statements • 760M edits

[as of October 2018]

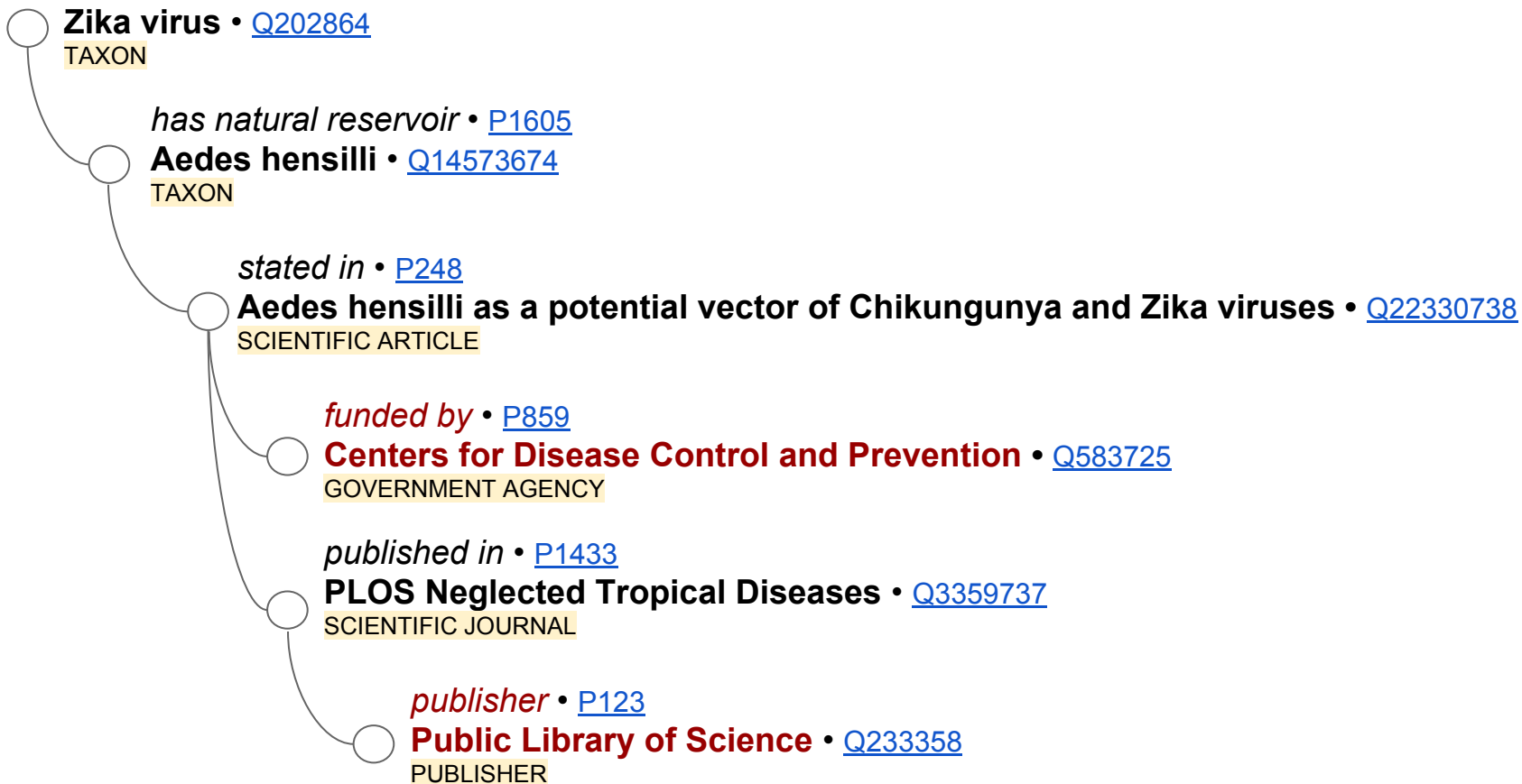
Types of content in Wikidata

- People
 - Places
 - Taxa
 - Buildings
 - Organisations
 - Artworks
 - Events
 - Astronomical Bodies
 - ...
- Chemicals
 - Processes
 - Theorems
 - Concepts
 - Creative works
 - Journals
 - Publishers
 - Meta-items
 - ...

Wikidata for Research



https://www.wikidata.org/wiki/Wikidata:WikiProject_Wikidata_for_research



Data provenance of a Wikidata statements by outlet, publisher and funder

Sample of current biomedical content in Wikidata



- All human, mouse genes and proteins (swissprot)
- All Gene Ontology terms
- All Human Disease Ontology terms
- All FDA approved drugs
- 109 reference microbial genomes

[Mitraka et al \(2015\) *Semantic Web Applications for the Life Sciences*](#)
[Burgstaller-Muelbacher et al \(2016\) *Database*](#)
[Putman et al \(2016\) *Database*](#)

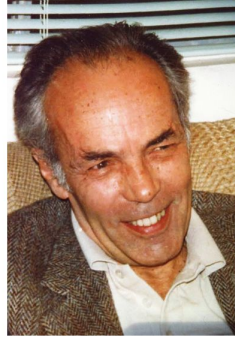
Biologists with Canadian citizenship



Severn Cullis-Suzuki
[commons:Severn Cullis-Suzuki.jpg](#)



Boucar Diouf
[commons:Boucar Diouf \(Salon du livr...](#)



Guy Courtois
[commons:Dr Guy Courtois.jpg](#)



Vladimir Hachinski
[commons:V hachinski.jpg](#)



Jack Szostak
[commons:Nobel Prize 2009...](#)



Anne Vallée
[commons:Anne Vallée.jpg](#)



Suzanne Morrow
[commons:Suzanne Morrow 194...](#)

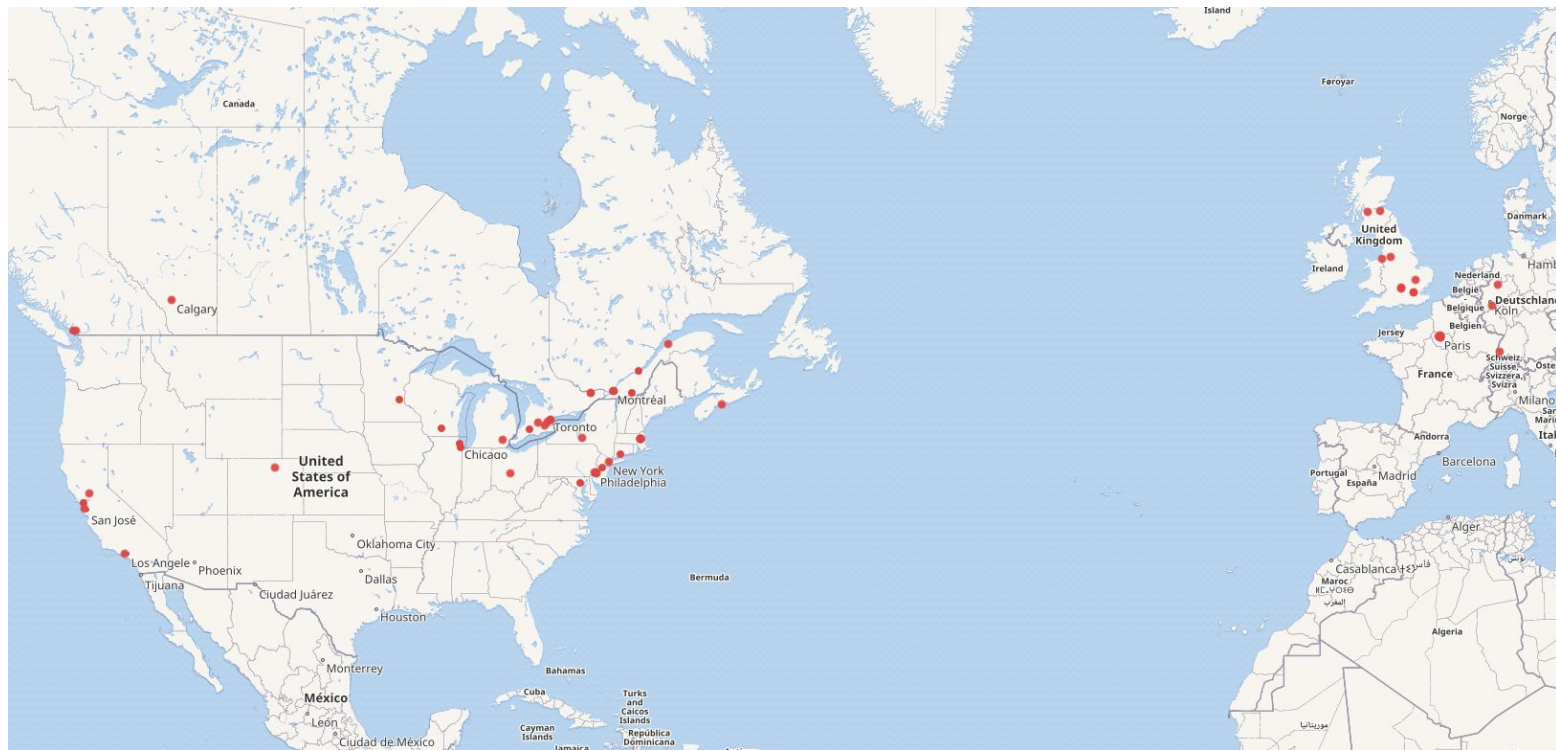


Sidney Altman
[commons:Sidney Altman crop.jpg](#)

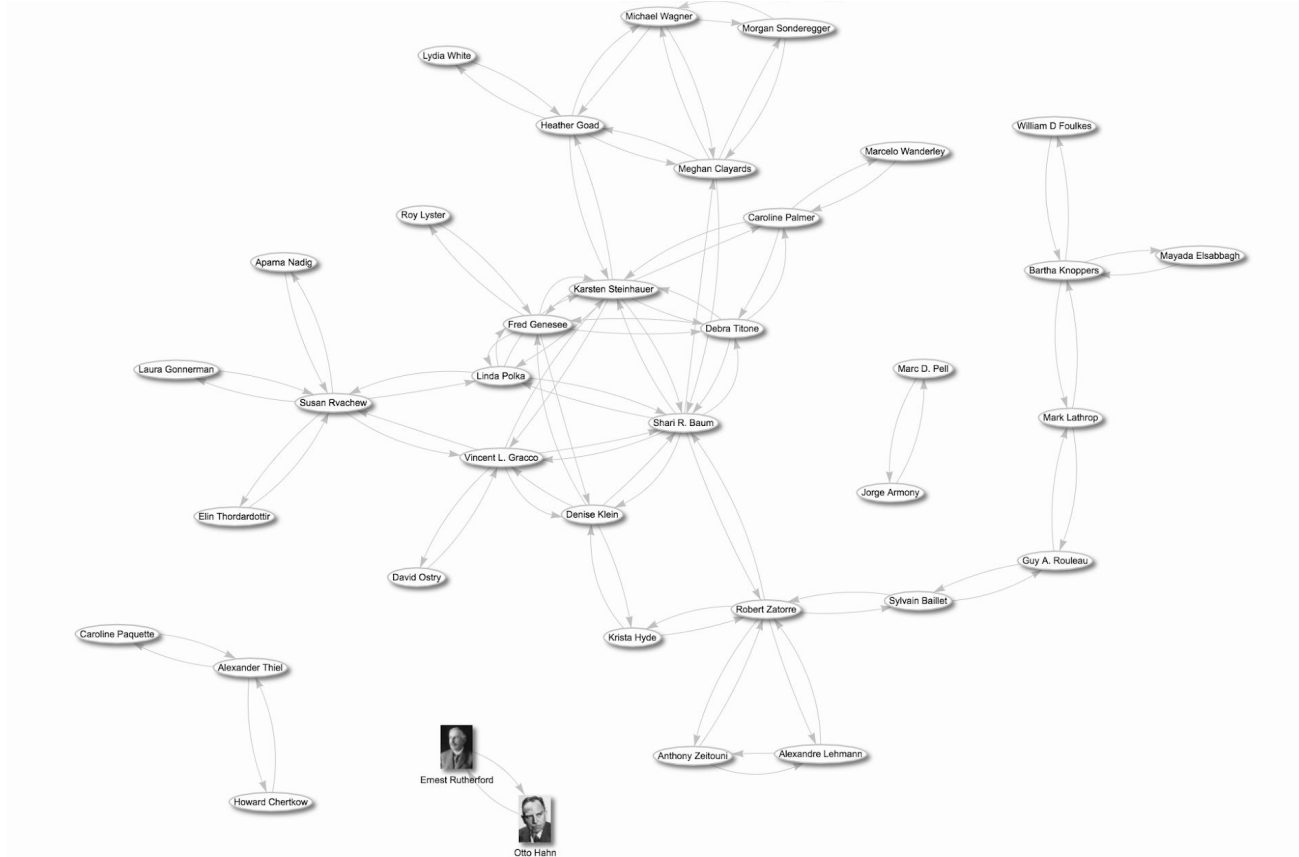


Roberta Bondar
[commons:Roberta Bondar2.jpg](#)

Institutions where Canadians got their PhD



Co-author graph of McGill-affiliated authors



[[SPARQL query](#)]

Award recipients affiliated with McGill

Count	Award	Recipients
56	Officer of the Order of Canada	Louis Dudek, Margaret Lock, Carl Goldenberg, Carl Goresky, Daniel Ling, David McCurdy Baird, Mark Wainberg, Henry Mintzberg, John Jonas, Eric Kierans, Don Johnston, Alan B. Gold, Alice Vibert Douglas, Balfour Mount, Bartha Knoppers, Clarke Fraser, Michel Chrétien, Nahum Sonenberg, Irwin Cotler, Bruce Trigger, Richard Goldbloom, Roy Heenan, Saul Hayes, Stanley Hartt, Stephen Toope, Thomas Cameron, Thomas Chang, Paul Davenport, Oliver Jones, Antoine Hakim, George Karpati, Heinz Lehmann, Jean Davignon, Desmond Morton, Donald Baxter, Rémi Quirion, Samuel O. Freedman, Theodore Sourkes, Thomas J. Hudson, Ronald Melzack, Warren Allmand, Krešimir Krnjević, Leo Yaffe, Albert Aguayo, Arthur Vineberg, Barry Posner, Gregory Baum, Marian Engel, Frederick Andermann, George Gray Bell, Heather Munroe-Blum, John Dossetor, Jonathan Larmonth Meakins, Michael Kelway Oliver, Moyra Allen, Myer Horowitz
24	Fellow of the Royal Society	Brenda Milner, Ernest Rutherford, William Osler, Donald O. Hebb, Charles Philippe Leblond, Arnold Burgen, Wilder Penfield, Charles Scriver, Michel Chrétien, Nahum Sonenberg, Otto Maass, Frank Dawson Adams, Edgar William Richard Steacie, Boris Babkin, John William Dawson, V. C. Wynne-Edwards, John George Adami, Allan Hay, Archibald Macallum, Arthur Stewart Eve, Frederick Soddy, Juda Hirsch Quastel, Henry Bovey, John Stuart Foster
19	Canadian Medical Hall of Fame	Mark Wainberg, Brenda Milner, William Osler, John McCrae, Donald O. Hebb, Charles Philippe Leblond, Wilder Penfield, Clarke Fraser, Charles Scriver, Michel Chrétien, Phil Gold, Richard Goldbloom, John Browne, Antoine Hakim, Heinz Lehmann, Ronald Melzack, Maude Abbott, Albert Aguayo, Henry Friesen
16	Companion of the Order of Canada	Kenneth Hare, Brenda Milner, Hugh MacLennan, Charles Philippe Leblond, Wilder Penfield, Charles Scriver, Paul-André Crépeau, Phil Gold, Richard Cruess, F. R. Scott, Victor Goldbloom, Henry Friesen, Raymond Klibansky, Charles Taylor, Juda Hirsch Quastel, Gerald Le Dain
16	Officer of the National Order of Quebec	Guy A. Rouleau, Luc Vinet, Margaret Lock, François Tavenas, Mark Wainberg, Brenda Milner, Henry Mintzberg, Michel Chrétien, Paul-André Crépeau, Irwin Cotler, Bruce Trigger, Richard Cruess, Ronald Melzack, Jean-Marie Dufour, Frederick Andermann, Heather Munroe-Blum

Federation

Wikidata's identifier mappings

Roberta Bondar (Q258385)

Canadian astronaut

Roberta Bondar

► In more languages

Statements

instance of

human

► 1 reference

image



Roberta Bondar2.jpg

371 × 463; 44 KB

Identifiers

VIAF ID

13114961

► 2 references

Library of Congress authority ID

nr95035760

► 1 reference

GND ID

119302152

► 1 reference

ISNI

0000 0000 7372 7287

► 1 reference

FAST ID

364028

▼ 0 references

Encyclopædia Britannica Online ID

biography/Roberta-Bondar

▼ 0 references

Photographers' Identities Catalog ID

388743

▼ 0 references

SNAC Ark ID

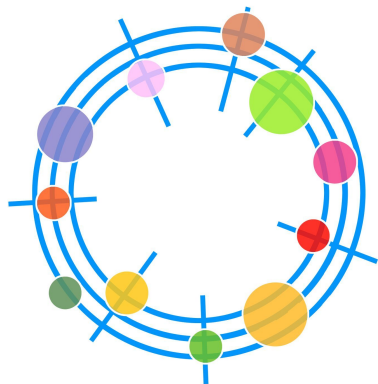
w6f21bsc

▼ 0 references

From Wikidata to Wikibase



Why Wikibase?



Linked Jazz

“Started thinking about how our data could live in Wikidata and started investigating feasibility of that possibility.

But we have very esoteric project data that doesn't seem appropriate to be in Wikidata so began looking at our own Wikibase instance.”

What's Wikibase

- [Wikibase Repository](#) - MediaWiki extension for structured, non-relational data in a central, collaboratively managed repository.
 - “writing RDF”
 - Revision control
 - FAIR by default
- [Wikibase Client](#) - MediaWiki extension for retrieving and embedding structured data from a central repository into a client wiki.
- [Query Service](#) that allows to query the contents of a Wikibase installation using [SPARQL](#)
- A set of [reusable components](#) that provide a foundation for tasks in the same domain.

<http://www.wikidata.org/entity/Q1>

.json

.php

.n3

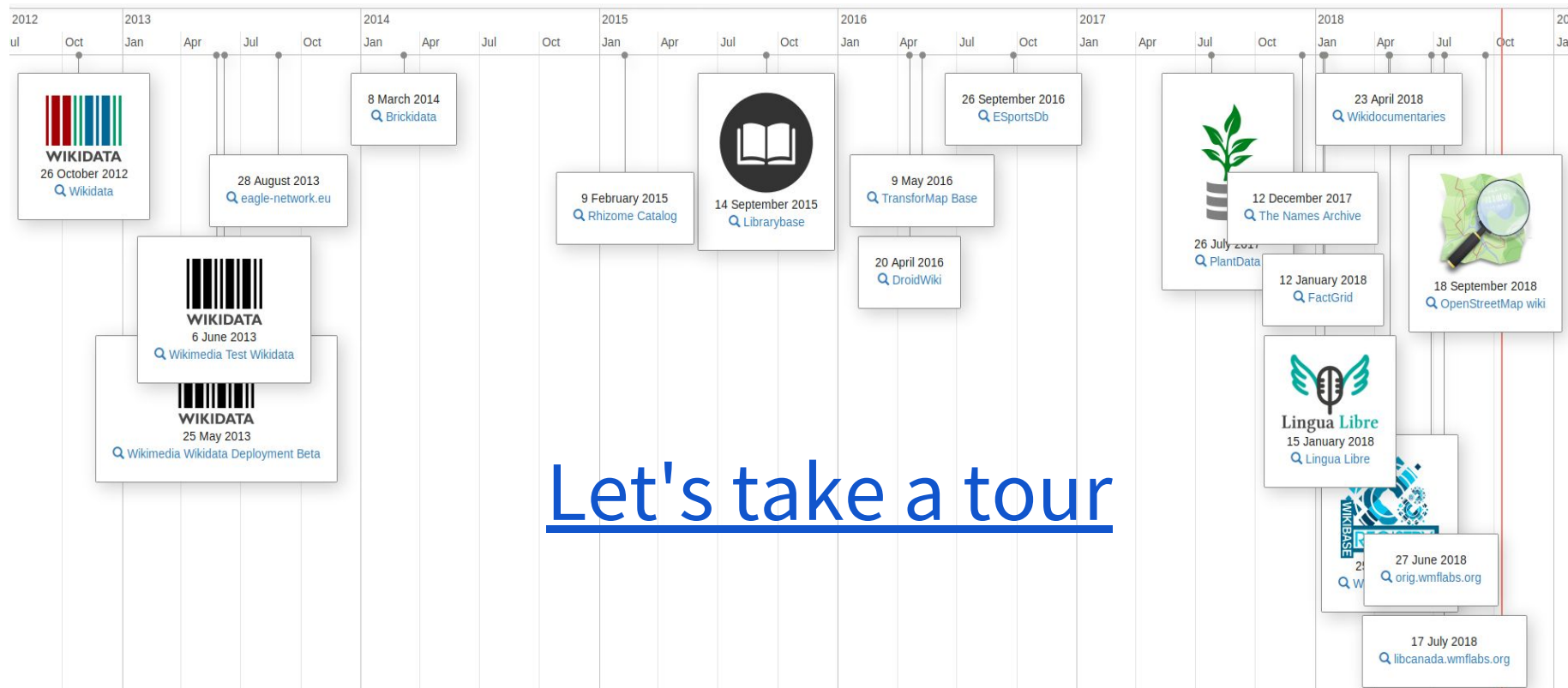
.ttl

.nt

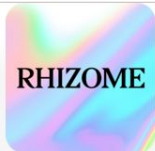
.rdf

Data formats in Wikibase
(versus wikitext)

Many faces of Wikibase



Properties on the Rhizome Wikibase



[Main page](#)
[Recent changes](#)
[Random page](#)
[Help](#)

Tools

[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)
[Page information](#)

In other languages

[Add links](#)

[Log in](#)

Page [Discussion](#)

[Read](#) [View source](#) [View history](#)



Indexes/Properties

[<](#) [Indexes](#)

id ↕	label ↕	aliases ↕	description ↕	inverse of ↕	equivalent property ↕
P2	equivalent property	owl:equivalentProperty	equivalent property in other ontologies (use in statements on properties, use property URI)		www.wikidata.org www.w3.org
P3	instance of	is a, is an, rdt:type	this item is a specific example and a member of that class		www.w3.org www.wikidata.org
P4	equivalent item		the same item in other ontologies (use item URI)		www.wikidata.org
P5	inverse of		links a property to its inverse property		www.wikidata.org
P6	reference URL	website, url, Uniform Resource Locator, webref, URL	should be used for internet URLs as references		www.wikidata.org
P7	has part	have part, composed of, contains, formed from, formed out of, assembled from, assembled out of, created from, created out of, amalgamation of, set of, consists of, has part, members, holonym of, includes, ingredient	object is a part of this subject	part of	www.wikidata.org
P8	part of	meronym of, section of, system of, subassembly of, sub-assembly of, contained within, assembly of, within a set	subject is a part of that object	has part	purl.obolibrary.org , www.wikidata.org
P9	imported from		source of this claim's value (use only in References section)		www.wikidata.org
P10	point in time	date, as of, time of event, at time, when, year, time	time and date something took place, existed or a statement was true		www.wikidata.org
P11	start time	from, starting, beginning, began, from time, since, from date, start date, building date, starttime	indicates the time an item begins to exist or a statement starts being valid, usually used as a qualifier		www.wikidata.org
P12	subproperty of	rdfs:subPropertyOf	all resources related by this property are also related that property.		www.wikidata.org , www.w3.org
P13	end time	to, to time, until, ending, enddate, end date, closed, fall date	indicates the time an item ceases to exist or a statement stops being valid		www.wikidata.org
P14	last modified		last known date of modification of the subject		
P15	subclass of	rdfs:subClassOf, hyponym of, is a, type of, is a type of, generalized by, generalization, superclass	this item is a class of that item	superclass of	www.wikidata.org , [http://www.w3.org/2000/01/rdf-schema#subClassOf] www.w3.org
P16	depends on	dependency, technical requirement, software dependency	the subject depends on this software		www.wikidata.org
P17	official website	website, official site, homepage, home page, official page, official web site, official url, official homepage, web address, url, blog	URL to the website of this item		www.wikidata.org , xmlns.com
P18	residence	lived in	the place where the person is, or has been, resident		www.wikidata.org
P19	country of citizenship	citizenship, nationality, citizen of	the object is a country that recognizes the subject as its citizen		www.wikidata.org

A French recording of the word “Canada”

Canada (Q26858)

From [Lingua Libre](#)

audio record - fra - GrandCelinien (GrandCelinien)

▾ In more languages [Configure](#)

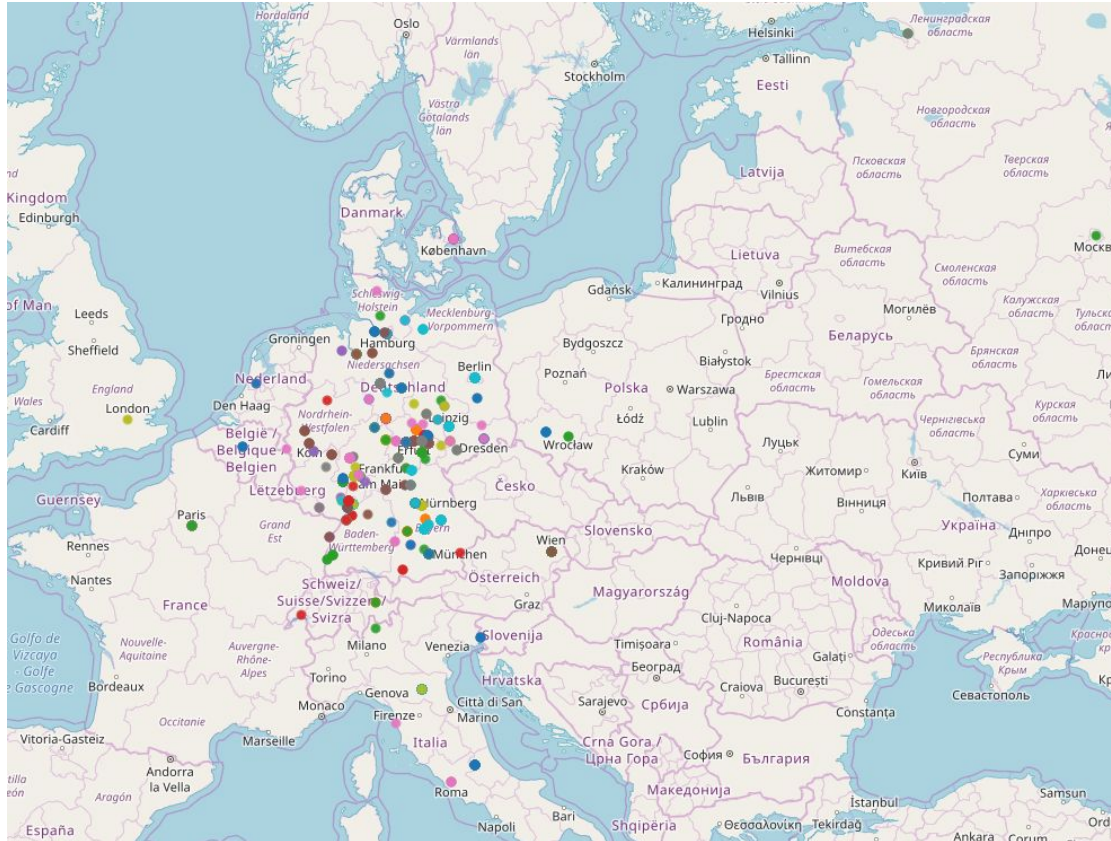
Language	Label	Description	Also known as
English	Canada	audio record - fra - GrandCelinien (GrandCelinien)	

Statements

instance of	record 0 references
subclass of	word 0 references
audio record	LL-Q150 (fra)-GrandCelinien-Canada.wav 0 references
language	French 0 references

From [Lingua Libre](#)

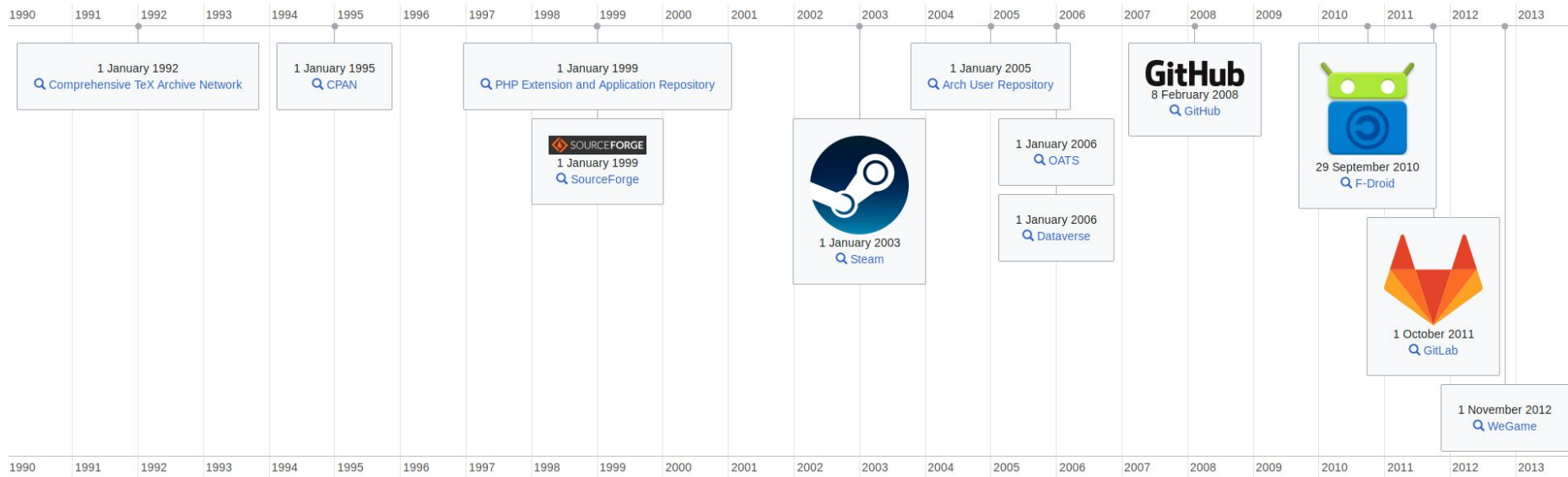
Letters sent by Illuminati



From [FactGrid's SPARQL endpoint](#)

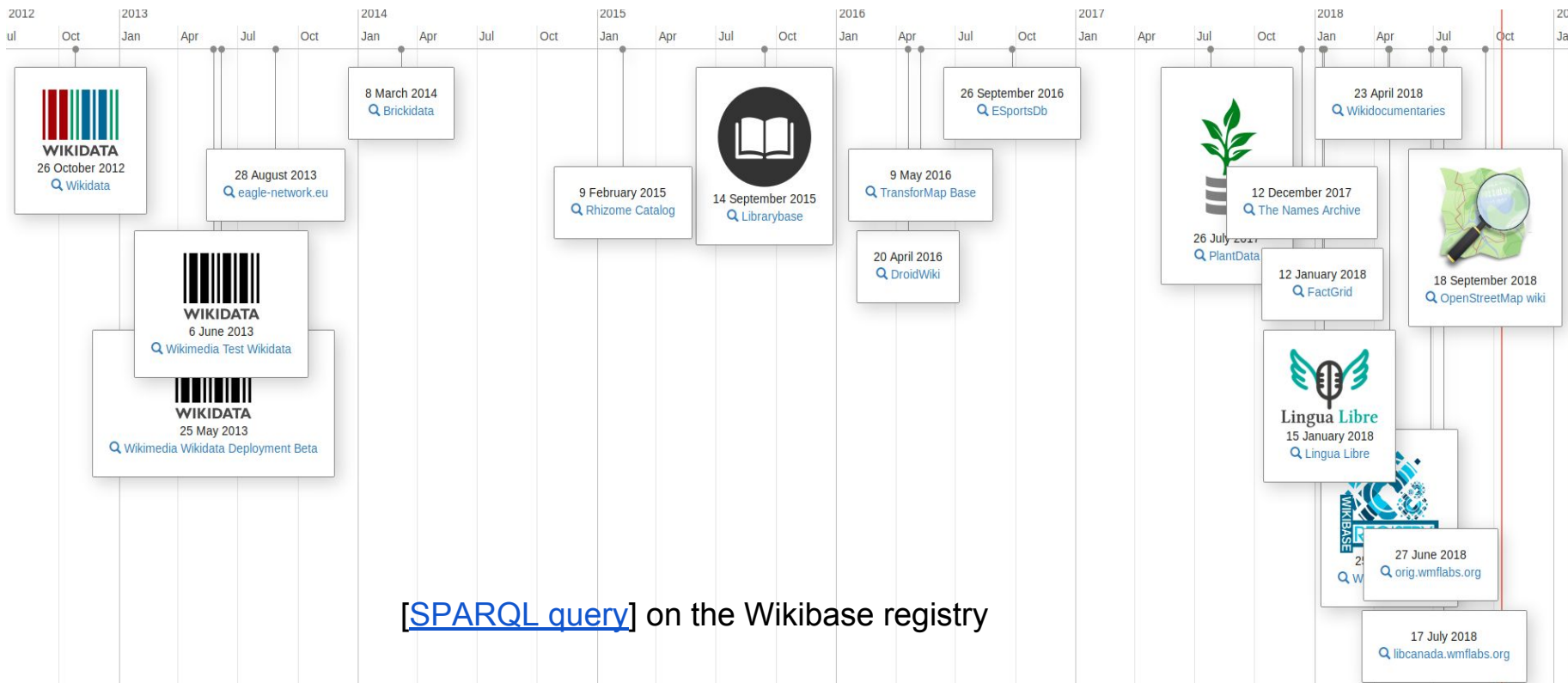
Colour indicates author

Timeline of software repositories



[[SPARQL query](#)] on Wikidata

Timeline of Wikibase instances



[\[SPARQL query\]](#) on the Wikibase registry

Wikibase and software repositories



Further notes

- Federation is [possible](#) to and from any SPARQL endpoint, not just Wikibase ones (and works fine on mobile)
- A Wikibase instance also has a [Mediawiki API](#)
- [Docker container](#) available to install Wikibase
- Wikimedia Commons is [moving to Wikibase](#)
- Ecosystem of [Wikidata tools](#), some being adapted to generic Wikibase instances
- Work on using [Shape Expressions](#) to share data models across instances
- Coordination through series of [workshops](#)
- Various non-public tests, e.g. [at OCLC](#)



TheWikipediaLibrary

@WikiLibrary

Following



Slide: "Wikibases in Canada?" Could an entire country use Wikibase(s) to protect data, distribute workflows, model worldviews, interact with other systems, and connect with libraries? @stacymallison asking the big questions at #WikibaseNYC

11:34 AM - 19 Sep 2018

Stacy Allison-Cassin (2018) [Wikibase & Indigenous Knowledge in the Canadian Context](#)

Wikidata or Wikibase(s)?

Wikidata community	Governance	depends
generic	Granularity	depends
CC0	Licensing	depends
stable	Funding	depends
stable	ID mappings	depends
many	Language(s)	depends

Thank you

D. Mietchen, D.Taraborelli (2018) Wikidata, Wikibase, and a federated ecosystem of structured knowledge for open science. *FORCE 2018*
doi.org/10.6084/m9.figshare.7195358 [CC BY]



Slides mashed up with contributions by
Andy Mabbett and Andra Waagmeester

growth by Fabio Rinaldi [CC BY], *research* by Minnie Pigeon [CC BY],
graph by Icon Lauk [CC BY] from the Noun Project