{ } wikicite

Citations for the sum of all human knowledge

Arguably the *most important ingredient of open knowledge*, sources and references have ironically received little technical attention in the Wikimedia movement up until now. Before Wikidata, attempts failed to address the issue of representing citations and source metadata in a well-structured, machine-readable format due to both the lack of mature technology and sufficiently well-organized community efforts. With **WikiCite 2016** – an event hosted by the Wikimedia Foundation and Wikimedia Deutschland, and generously supported by the Alfred P. Sloan Foundation, the Gordon and Betty Moore Foundation, and Crossref – we seeded the development of a vision to build a centralized bibliographic database of source and citation metadata in Wikidata to serve Wikimedia projects and, eventually, the sum of all human knowledge.

D. Taraborelli, J. Dugan, L. Pintscher, D. Mietchen, C. Neylon (2016) *WikiCite 2016 Report.*
doi.org/10.6084/m9.figshare.4042530 • commons.wikimedia.org/wiki/File:WikiCite_2016_report.pdf CC BY

WIKIMEDIA FOUNDATION

WIKIMEDIA DEUTSCHLAND

Alfred P. Sloan FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

Crossref

WikiCite 2016 Report

Table of contents

- Meeting Report
  - Overview
  - Background
- Impact and Outcomes
- Workgroups
- Survey Results
- Financial report
- Organizing Committee
- List of Participants
- List of Organizations
- Ongoing Initiatives
  - Accomplishments
  - Events
  - Documentation
  - Outreach
  - Grant proposals
  - Data releases
  - Code releases
  - Research

Meeting Report

Overview

WikiCite 2016 was a two-day event in **Berlin, Germany,** from **25-26 May 2016.** WikiCite was held at **GLS Campus** in the Prenzlauer Berg district in Berlin. The Wikimedia Foundation and Wikimedia Deutschland co-hosted the event. Crossref, the Gordon and Betty Moore Foundation, and the Alfred P. Sloan Foundation generously supported the event. The Wikimedia Foundation Board of Trustees approved the funding to cover WikiCite's cost.

A diverse and unique group of approximately 55 participants from 48 organizations – which included universities, libraries, and open data stakeholders – met to discuss and design solutions for citations and source metadata on Wikimedia projects. The focus was on using the semantic backbone of Wikipedia – Wikidata – as a repository and mechanism with which to automate and support standardizing and implementing best practices for citations.

During the morning of the first day, submitted **proposals** were discussed by group consensus, and the attendees created several breakout **workgroups**. Starting after lunch on the first day, these groups worked on each subject area. In the afternoon of the second day, all the groups came together to present their findings and engage in a group discussion of next steps.

As a result of this coordinated work, several efforts were started, continued, and rejuvenated during and after the conference (see Impact and Outcomes below).

WikiCite work continues with the **WikiCite-discuss mailing list** as a main conduit of information and discussion and a new umbrella page for the initiative on Metawiki. There is ongoing discussion for a follow-up meeting in 2017 to continue the structuring work on source and citation data and coordinate people and efforts on connected projects.

Background

Citations are a simple, critical interconnection mechanism for all modern knowledge in the digital, Internet-connected world. Each time we assert knowledge and share it in the scholarly realm, we cite sources. Despite their critical importance, most citations today are usually

expressed as free text, inaccessible in open-licensed databases, and difficult to organize and assess by researchers who wish to understand our current knowledge. Not only is this true in the literature, it's also true for source metadata and citations in Wikimedia projects.

Wikipedia is one of the top-ten most visited global websites, and holds the largest and most complete set of general reference data available. Currently, the English Wikipedia (the largest out of more than 290 different language editions) includes approximately 5.2 million articles with over 29 million citations. Wikipedia's Verifiability Policy requires inline citations for any material challenged or likely to be challenged, and for all quotations, anywhere in article space. Citations are an incredibly important part of how Wikipedia works, and their use is deeply integrated into the process of group knowledge acquisition. Like citations in other knowledge areas, the citations on Wikimedia projects (including all Wikipedia editions) are currently not stored as structured data, but rather included as marked, free text that users manually edit and write onto Wikipedia pages.

Over the last three years, Wikimedia Deutschland, in collaboration with the Wikimedia Foundation, has built a new project, called Wikidata, to host and store structured data. Any data, expressed within any data model, can be stored and shared openly on Wikidata. Additionally, these data can be integrated into the knowledge expressed within the content of web pages across Wikimedia projects. For example, the population of a city (stored and updated as an integer) on Wikidata can populate the Wikipedia page about that city. This enables a wide diversity of automation, error checking and verifiability to the sum knowledge shared across a Wikimedia project.

Over the last two years, several contributors have developed Wikidata models to express and store source metadata (the bibliographic data) for the sources cited in Wikimedia projects. This data includes such things as *journal name*, *publication date*, *author names*, *page number*, etc. The next step is to implement similar structured models for the citations contained within Wikimedia projects. This is part of a longer process in a major, behind-the-scenes transition that will place both the bibliographic source data and the individual, specific references on pages currently in Wikimedia projects into the data models and structured data in Wikidata. To do this, we need robust, widely accepted models for how to express and use citations in a structured way, and then build tools and software that mine existing Wikipedia citations and express them in Wikidata.

To build a system to structure and share citations successfully, one goal of this meeting was to integrate the methods we use and the tools we build with existing tools and systems that create, use, and share citations today. We also want to align leading Wikipedia contributors and tool developers to the needs and benefits of structuring and sharing citation data within Wikidata. This integration included many face-to-face discussions and alignment with a diverse set of people, motivating the need for the conference.

Impact and Outcomes

> "Open scholarly communication infrastructure needs to shift from a document-centric to a knowledge-centric approach"
>
> – Sören Auer,

Despite being – arguably – the most important ingredient of open knowledge, sources and references have ironically received little technical attention in the Wikimedia movement up until now. Before Wikidata, attempts failed to address the issue of representing citations and source metadata in a well-structured, machine-readable format due to both the lack of mature technology and sufficiently well-organized community efforts. With WikiCite 2016, we seeded the development of a vision to build a centralized bibliographic database of source and citation metadata in Wikidata to serve Wikimedia projects and, eventually, the sum of all human knowledge.

The meeting was an overwhelming success (see results from a participant survey). The event exceeded the simple goals of convening a diverse group of interested stakeholders and holding focused workgroup sessions on structuring and sharing citations. The meeting brought together several different projects already underway in science citations, and catalyzed work on existing efforts on Wikimedia citations. Now, 10 months later, several ongoing projects are in active development. We expect these projects to continue through 2017, and with ongoing efforts to spawn more, similar projects.

Highlights of initiatives that started or were significantly accelerated by WikiCite include:

- The ingestion into Wikidata of all references with an identifiable PMCID from English Wikipedia as well as the bibliographic metadata and the citation graph of all open access review articles from the biomedical literature of the last 5 years.
- The creation of a complete bibliographic corpus and citation graph on the Zika virus literature in Wikidata.
- A set of initiatives, in concert with the OpenCitations project and Open Access publishers, exploring strategies to accelerate the distribution and availability of citation data for scholarly works under open licenses.
- The cross-pollination of technical efforts around automated citation extraction between the Wikidata and DBpedia communities.
- The development or improvement of tools and algorithms for automated fact extraction from the literature, such as WikiFactMine or StrepHit.
- The design of a proof-of-concept application generating Wikidata-driven scholarly author profiles, entirely powered by linked open data and SPARQL endpoints.
- A series of high-profile presentations on WikiCite, targeted at different audiences and venues: the scholarly link open data community (VIVO '16 closing keynote), the Open Access scholarly publishing community (COASP '16 technology and innovation panel), the biocuration and medical research community (NIH Data Science lecture series), and the Wikimedia movement (September 2016 Wikimedia Monthly Metrics and Activities meeting).

Wikipedia is today the largest online reference work and one of the world's top ten sources of traffic to the literature: its success depends on the ability to provide readers and contributors with resources to check and verify information against reliable sources. We believe the work we seeded with WikiCite will have a lasting impact on the quality and reliability of Wikimedia projects and benefit their readers and contributors alike. We also believe that partnerships established at WikiCite (with organizations such as OCLC, Crossref, OASPA, the University of Chicago Knowledge Lab, OpenCitations, libraries and scholarly publishers) will help dramatically improve the availability of open citation data.

Workgroups

The following is a list of workgroups at the event in Berlin, cross-linked to the full report.

Group 1: Modeling bibliographic source metadata

Goal: Discuss and draft data models to represent different types of sources as Wikidata items

Summary: The group conducted in-depth discussions of various approaches of modeling different types of sources in Wikidata. Examples included blog posts, newspaper articles, journals, and different kinds of books. The approaches have different ramifications and concerns, including problems of automation, accessibility for new users, reusability, and required maintenance. The group discussed previously existing modeling attempts used by the Wikidata community (for example, regarding books), and reviewed the list of properties for different types of documents (articles, books, etc.). Choosing a consistent approach is important both to the Wikidata community and third party users. The discussion focused on many kinds of sources and important edge cases.

Group 2: Reference extraction and metadata lookup tools

Goal: Design or improve tools to extract identifiers and bibliographic data from Wikipedia citation templates, look up and retrieve metadata

Summary: This group conducted a wide-ranging discussion on various existing tools use for data extraction and structuring. Many examples can be seen in the raw notes from the discussion. Each member shared with the group the tools they worked with, and how their efforts meshed with the needs seen by others. Two main conclusions from the discussions included:

- Existing tools to extract bibliographic metadata and references should be improved further and more tailored to Wikidata.

- Further technical discussions, especially between people involved in Zotero translators and people involved in Citoid/WikiCite, seem fruitful for both sides.

Group 3: Representing citations and citation events

Goal: Discuss how to express the citation of a source in a Wikimedia artifact (such as a Wikipedia article, Wikidata statements, etc.) and review alternative ways to represent them

Summary: This working group discussed the use cases and data needs for structured citations. We defined the terminology of discussion, created a recommendation for the data structure for citation instances, and explored how the existing infrastructure and community needs would support a transition from the current systems for representing citation to a more structured approach leveraging source metadata stored in Wikidata.

Group 4: (Semi-)automated ways to add references to Wikidata statements

Goal: Improve tools for semi-automated statement and reference creation (e.g., StrepHIT, ContentMine)

Summary: This working group created a new RFC on Wikidata, located here to obtain community approval to expand and polish the feature set on the "Primary Sources Tool". This is an ongoing effort now within the Wikidata community.

Group 5: Use cases for source-related queries

Goal: Identify use cases for SPARQL queries involving source metadata. Obtain a small open licensed bibliographic and citation graph dataset to build a proof-of-concept of the querying and visualization potential of source metadata in Wikidata. Includes work on the Zika virus corpus.

Summary: Wikidata will serve as a centralized, highly structured, repository capable of representing the densely networked nature of the scholarly sources that support the knowledge archived across all Wikimedia projects. This signals an unprecedented opportunity for not only scientists and scholars but also society at large to explore the complex landscape of human knowledge. Yet, it is not clear what such an exploration would look like. What kinds of questions can be asked of such a system? We established a working group to not only envision concrete use cases for scholarly source-related questions in Wikidata, but also to determine whether the technical foundations required to effectively express those questions as intelligent, efficient, and systematic queries are in place. Where these technical foundations are lacking but needed, the working group tasked itself with developing proposals for overcoming such limitations.
This group focused on discussing and prioritizing use cases for Wikidata queries involving source metadata. The assumption is that we already have or have access to all the required data. In addition, we worked to obtain a small, open licensed bibliographic and citation graph dataset to build a proof of concept of the querying and visualization potential of having this data stored in Wikidata and exposed via SPARQL.

Three additional workgroups formed spontaneously on the second day of the event:

Group 6: Wikidata as the central hub on license information on databases

Goal: Add license information to Wikidata to make Wikidata the central hub on license information on databases

Group 7: Using citations and bibliographic source metadata

Goal: Merge groups working on citation structure and source metadata models and integrate their recommendations

Group 8: Citoid-Wikidata integration

Goal: Extend Citoid to write source metadata into Wikidata

Survey Results

The organizing committee collaborated with a team at Carnegie Mellon University to conduct a survey of participants in WikiCite 2016, as part of a Sloan-funded project to "*enhance the sustainability of free and open source software by understanding how engagements with code build community, and disseminating knowledge and tools that will allow stakeholders to plan and conduct successful engagements to build strong, cohesive open source communities that will maintain and enhance the software they use.*"

Event participants were polled online and interviewed in June 2016, and the results were analyzed in the following months. A complete presentation of the research methods and summary tables of the results can be found on the wiki pages describing the research.

Satisfaction indicators

The below table presents aggregate feedback data about participants' satisfaction with various aspects of event organization. Items are on a 5-point scale, from Strongly Disagree to Strongly Agree, and 3 representing a neutral response.

Overall, results show participants were satisfied or very satisfied with most aspects of organization. Facilities showed a score slightly below neutral (2.91). However, qualitative feedback suggests this to be associated with major network stability issues that were attributed to the venue, rather than event organization.

Question	Number of responses	Mean	Standard deviation	Minimum value	Maximum value
Help for any problems	21	4.05	0.8	3	5
Communication by organizers	22	4.45	0.96	2	5
Facilities	22	2.91	1.31	1	5
Refreshments	22	4.23	0.81	2	5
Accommodation	21	4	0.84	3	5
Outings	21	4.24	0.83	3	5
Session variety	21	4.24	0.83	2	5
Session quality	22	4.32	0.84	2	5
Overall organization	22	4.41	0.67	3	5
Event duration	22	4.09	0.87	2	5

We attribute the low score on facilities to a network breakdown on day 1, making coding very difficult, but which had the surprising benefit of encouraging much more conversation between participants.

Multi-item scale results

The below table presents aggregate results of psychometric variables examined, as well as outcomes of the event. All items (except "New connections made") are on a 5-point scale, from Strongly Disagree to Strongly Agree, and 3 representing a neutral response. Overall results suggest participants were somewhat satisfied with the outcomes of the event, and the process of working together. Individuals made over 3 new connections on average with whom they may start new collaborations. Overall, groups reported a participative or highly participative environment, and some use of brainstorming techniques to source ideas from all group members. Individuals also reported being somewhat satisfied with goal clarity.

More detailed inferential statistics will be made available via an open access publication, currently under submission.

Question	Number of responses	Mean	Standard deviation	Minimum value	Maximum value
Satisfaction with Outcome	22	3.86	0.74	1.86	5
Satisfaction with Process	21	3.65	0.68	2.25	5
No. New Connections Made	21	3.48	1.21	1	6
Perceived Participation	21	4.37	0.52	3	5
Brainstorming	21	3.33	0.51	2.33	4.17
Goal clarity of session/group	22	3.49	1.17	1	5
Software use Self-efficacy	21	3.65	0.7	2.5	5

Financial report

The Wikimedia Board of Trustees approved the WikiCite initiative as a recipient of restricted grants from funders. Dario Taraborelli and Jonathan Dugan, co-PIs on the proposal, managed the grants, in coordination with the organizing committee and disbursed by the Wikimedia Foundation. As of October 19, the grant has been used as follows:

Total funding	$35,000	Grant from the Alfred P. Sloan Foundation, the Gordon and Betty Moore Foundation, and Crossref
Total spent	*$29,131*	See cost breakdown below
Balance	$5,868	

Cost breakdown

Travel grants	*$18,908*	We issued travel scholarships to allow 18 out of 55 participants with no additional sources of funding to attend the event.
Venue	*$4,288*	We obtained a 50% discount from the final invoice due to major network breakdown, which resulted in additional costs for securing connectivity.
Dinners	*$2,611*	Dinners for 55 participants
Other costs	*$3,324*	Wi-fi hotspots and administration costs; outreach travel expenses.

Unused funding from the grants will be used for travel costs related to outreach on the 2016 initiative and (pending WMF board approval) towards funding of the 2017 event.

Organizing Committee

- Jonathan Dugan
- Daniel Mietchen (National Institutes of Health (NIH))
- Cameron Neylon (Curtin University)
- Lydia Pintscher (Wikimedia Deutschland, Wikidata)
- Dario Taraborelli (Wikimedia Research)

List of Participants

- Thomas Arrow (ContentMine)
- Adam Becker (*Open Journal*, Freelance Astrophysicist)
- Patrice Bellot (Aix-Marseille Université - CNRS - LSIS / OpenEdition Lab)
- Terry Catapano (Plazi Verein / Columbia University Libraries)
- Scott Chamberlain (rOpenSci)
- Cristian Consonni (Wikimedia Italia, Università degli Studi di Trento (University of Trento))
- Karen Coyle (KarenCoyle.net)
- Marin Dacos (CNRS - OpenEdition Lab)
- Antonin Delpeuch (Dissemin)
- Eamon Duede (Knowledge Lab @ University of Chicago)
- Katie Filbert (Wikimedia Deutschland, Wikidata)
- Konrad Förstner (Universität Würzburg (University of Wurzburg))
- Marco Fossati (Fondazione Bruno Kessler (FBK))
- Susanna Giaccai (Wikimedia Italia)
- Aaron Halfaker (Wikimedia Research)
- James Hare (WikiProject X, Wikimedia DC)
- Lambert Heller (Technische Informationsbibliothek (TIB) (German National Library of Science and Technology))
- Erika Herzog (Wikimedia New York City)
- Markus Kaindl (Springer Nature)

- Alex Kalderimis (RefMe)
- Sebastian Karcher (Qualitative Data Repository / Zotero, Citation Style Language (CSL))
- John Kaye (Jisc)
- Chris Keene (Jisc)
- Daniel Kinzler (Wikimedia Deutschland, Wikidata)
- Jonas Kress (Wikimedia Deutschland)
- Nettie Lagace (National Information Standards Organization (NISO))
- Rachael Lammey (Crossref)
- Mairelys Lemus-Rojas (University of Miami Libraries)
- Luca Martinelli (Wikimedia Italia)
- Jens Nauber (Die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) (Saxon State and University Library Dresden (SLUB)))
- Finn Årup Nielsen (Danmarks Tekniske Universitet (Technical University of Denmark))
- Jake Orlowitz (Ocaasi) (The Wikipedia Library)
- Merrilee Proffitt (OCLC Research)
- Laura Rueda (DataCite)
- Diego Sáez-Trumper (Eurecat)
- Sébastien Santoro
- Till Sauerwein (Universität Würzburg (University of Wurzburg))
- Tobias Schönberg (talk) (Wikidata)
- Elizabeth Seiver (Public Library of Science (PLOS))
- Adam Shorland (Wikimedia Deutschland, Wikidata)
- Mike Showalter (OCLC)
- Chiara Storti, (Wikimedia Italia, Rete bibliotecaria di Romagna e San Marino)
- Jon Tennant (Imperial College London, ScienceOpen)
- Katherine Thornton (University of Washington)
- Marielle Volz (Wikimedia Foundation) (attending remotely)
- Andra Waagmeester (Micelio)
- Joe Wass (Crossref)
- Chris Wilkinson (*eLife Sciences*)
- Andrea Zanni (Wikisource) / Aubrey
- Jan Zerebecki (Wikimedia Deutschland)
- Philipp Zumstein (Universitätsbibliothek Mannheim (Mannheim University Library))

List of Organizations

We brought together Wikidata editors, Wikipedians, developers, data modelers, open access publishers, and information and library science experts from various organizations, as well as academic researchers from groups with experience working with Wikipedia's citations and bibliographic (linked open) data in general. This is the list of organizations represented at the event.

- Aix-Marseille Université
- Centre national de la recherche scientifique (CNRS)
- Columbia University Libraries
- Content Mine
- Crossref
- Citation Style Language (CSL)
- Danmarks Tekniske Universitet (Technical University of Denmark)
- DataCite
- Die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) (Saxon State and University Library Dresden (SLUB))
- Dissemin
- École Normale Supérieure
- *eLife Sciences*
- Eurecat
- Fondazione Bruno Kessler (FBK)
- Gene Wiki
- Imperial College London
- Jisc
- Knowledge Lab @ University of Chicago
- Laboratoire des Sciences de l'Information et des Systèmes (LSIS)
- Micelio
- National Institutes of Health (NIH)
- National Information Standards Organization (NISO)

- OCLC
- OpenEdition Lab
- *Open Journal*
- Plazi Verein
- Public Library of Science (PLOS)
- RefMe
- rOpenSci
- ScienceOpen
- Springer Nature
- Technische Informationsbibliothek (TIB) (German National Library of Science and Technology)
- Università degli Studi di Trento (University of Trento)
- Universität Würzburg (University of Wurzburg)
- Universitätsbibliothek Mannheim (Mannheim University Library)
- University of Manchester
- University of Miami Libraries
- University of Pittsburgh
- University of Washington
- Wikidata
- Wikimedia DC
- Wikimedia Deutschland
- Wikimedia Foundation
- Wikimedia Italia
- Wikimedia New York City
- Wikimedia Research
- WikiProject X
- Zotero

Ongoing Initiatives

We published the first overview of initiatives that took place after WikiCite 2016 in Berlin, spawned by the activities and connections that were created or accelerated at the event.

Accomplishments

The Zika corpus

In February, the World Health Organization declared a public health emergency over the Zika virus outbreak and its links (then suspected, by now confirmed) to microcephaly and Guillain-Barré syndrome. By that time, around 150 scholarly articles had been published about the virus since its discovery in 1947, and the majority of these articles had already been assigned Wikidata items.

Since then, the literature on the topic has grown about tenfold, and the Wikidata coverage has mostly kept pace, with a typical time lag of less than a week. While not complete, this corpus covers most PubMed-indexed English-language articles reporting or reviewing original research about the Zika virus and the infections it can cause in mosquitoes, humans and animal models, as well as about approaches to prevention, diagnostics, therapy, or surveillance.

The ZIka corpus served as a nucleus for creating a citation graph on Wikidata (see below) and for exploring co-author networks and similar information on Wikidata. It is now slowly expanding to encompass literature about related subjects, e.g., flaviviridae and mosquito-borne diseases more broadly, epidemiological modeling or data sharing in public health emergencies.

All English Wikipedia references citing PMCIDs

All identifiable references mentioned in the English Wikipedia with a *PubMed Central identifier* (P932), based on a dataset produced by Aaron Halfaker using the mwcites library have been imported as individual bibliographic entries in Wikidata. As of today, there are over 150,000 items using this property.

Metadata of OA biomedical reviews (2011-2016) and their citation graph

James Hare has been working on importing open access review papers published in the last 5 years as well as their citation graph. These review papers are not only critical to Wikimedia projects, as sources of citations for Wikipedia articles and statements: a significant portion of these works also open license their contents, which will allow semi-automated statement extraction via text and data mining strategies. As part of this project, the property *cites* (P2860) created during WikiCite 2016 has been used in over half a million statements representing citations from one paper to another.

While this is a tiny fraction of the entire citation graph, it's a great way of making data available to Wikimedia volunteers for crosslinking statements, sources and the works they cite.

New Wikidata properties

The *Crossref funder ID* property (P3153) can now be used to identify funders that can be linked to particular works (when available) via the P859(*sponsor*) property. This will allow novel analyses on sources for Wikidata statements as a function of particular funders.

The *uses property* property (P3176), which Finn Årup Nielsen conveniently dubbed the "selfie property", can now be used to identify external works that mention specific Wikidata properties. The list of articles and papers with that grows.

The *OpenCitations bibliographic resource ID* property (P3181) can be used to specify the bibliographic resource identifier for any publication in WikiCite that is also included in the OpenCitations Corpus.

Events

WikiCite 2017

We're planning a follow-up event in May 2017 in Vienna to further the connections and projects ongoing within the community, and provide a dedicated time and location for people who need to collaborate and coordinate in person to meet. In addition to the intentions of the first meeting in 2016, we expect the 2017 event to also showcase recent results and include new participants to further the community utilization and goals of open, structured citations and source metadata and integrate it into open knowledge tools. We aim to reach a broader number of organizations including key stakeholders with the NIH, the OpenCitations project, the Gene Wiki project, more librarians, and more representatives from science publishing, both Open Access and subscription-based. We expect to co-host the event with the 2017 Wikimedia Hackathon in Vienna (May 19-21, 2017), which will give us access to a large number of volunteers as well as WMF and WMDE developers.

Documentation

WikiCite 2016 Report

The present report on the WikiCite 2016 meeting and ongoing activities is available on wiki.

Book metadata modeling proposal

Chiara Storti and Andrea Zanni – who attended the event in Berlin – posted a proposal with examples to address in a pragmatic way the complex issues surrounding metadata modeling for books. If you're interested in the topic, please chime in.

Wikidata Primary Sources Tool RFC

The open request for comment centralizes feature requests, technical issues and general discussion on the primary sources tool, namely a data curation facility with a focus on the addition of references to Wikidata claims.

Outreach

European Library Automation Group (ELAG '16)

On June 9, 2016, Karen Coyle gave a brief presentation on WikiCite at *ELAG '16*, the annual meeting of the European Library Automation Group.

Wikimania 2016

On June 24, 2016, Alex Stinson gave a talk at *Wikimania '16* titled "What do the Footnotes mean? The Implications of Wikipedia's Verifiability Policy", with a high-level overview of how sources came to be so important in various Wikipedias, recent research on the value and impact of our current citations, and community programs that focus on the importance of citations, such as the Wikipedia Library, its #1lib1ref campaign and Wikicite 2016.

WikiCite on Open Science Radio

On May 26, 2016, after the closing of WikiCite 2016, Konrad Förstner recorded a podcast interview for *Open Science Radio* with Lydia Pintscher and Dario Taraborelli on the event and the motivation behind it.

WikiCite at VIVO '16

On August 19, 2016, Dario Taraborelli delivered the closing keynote at the *VIVO '16* conference in Denver, CO. The keynote sparked a discussion on how Wikidata can help connect siloed research information systems and linked data repositories. A video and slides of the keynote are available.

WikiCite, Wikidata and Open Access publishing

On September 21, Dario Taraborelli gave an invited presentation (slides) on WikiCite in the *Technology and Innovation* panel at the *8th Annual Conference of the Open Access Publisher Society* (COASP 2016) in Arlington, VA. The presentation triggered a discussion on the availability of open citation data. In collaboration with Jennifer Lin (Crossref) we discovered that out of 999 publishers already depositing citation data to Crossref, only 28 (3%) make this data open. We urged publishers, particularly Open Access publishers and OASPA members, to release this data that's critical to initiatives such as WikiCite.

Linking sources and expert curation in Wikidata: NIH lecture

On September 23, Dario Taraborelli also gave a longer presentation at the National Institutes of Health (NIH) in Bethesda, MD, mostly focused on the integration of expert-curated statements (such as those created by members of the Gene Wiki project) and source metadata in Wikidata, as part of the *NIH Frontiers in Data Science lecture series*. (video, slides) This is a slightly modified version of the VIVO '16 closing keynote, targeted at the biomedical science community.

So what can we use WikiCite for?

Finn Årup Nielsen wrote a blog post showcasing different ways in which a repository of source metadata could be used. He also posted a list of possible use cases, comparing Wikidata to other research information/profile systems. Discussions triggered by his blog post led to the creation of Scholia – a proof-of-concept application generating Wikidata-driven scholarly author profile, entirely powered by open data and SPARQL endpoints.

WikiCite at WMF Monthly Metrics

On September 29, a short retrospective on WikiCite was presented during the September 2016 Wikimedia Monthly Activity and Metrics Meeting (video, slides)

WikiCite at the 2016 Crossref Annual Meeting

On November 2, 2016 Dario Taraborelli will give an invited plenary talk at LIVE16, Crossref's annual meeting in London, UK.

Grant proposals

Three proposals closely related to WikiCite applied for funding through Wikimedia Grants. As of October 15, funding decisions for WikiFactMine and LibraryBase have been published and both projects were successfully selected by the funding committee. The StrepHit grant renewal extension is still pending a funding decision.

WikiFactMine

WikiFactMine is a proposal by the ContentMine team to harvest the scientific literature for facts and recommend them for inclusion in Wikidata.

Librarybase

Librarybase is a proposal to build an "online reference library" for Wikimedia contributors, leveraging Wikidata.

StrepHit

The StrepHit team submitted a grant renewal application to support semi-automated reference recommendation for Wikidata statements. The main goal is to make the primary sources tool usable.

Data releases

First release of the Open Citation Corpus

The OpenCitations project announced the first release of the Open Citation Corpus, an "open repository of scholarly citation data made available under a Creative Commons public domain dedication (CC0), which provides accurate bibliographic references harvested from the scholarly literature that others may freely build upon, enhance and reuse for any purpose, without restriction under copyright or database law." The OpenCitation project uses provenance and SPARQL for tracking changes in the data.

Data on DOI citations in Wikipedia from Crossref

Crossref recently announced a preview of the Crossref Event Data user guide, which provides information on mentions of Digital Object Identifiers(DOI) across non-scholarly sources. The guide

includes a detailed overview of how the system collects and stores DOI citations from Wikimedia projects, and how this data can be programmatically retrieved via the Crossref APIs.

Code releases

Converting Wikidata entries to BibTeX

ContentMine fellow Lars Willighagen announced a tool combining Citation.js with Node.js, which allows, among other things, to convert a list of bibliographic entries stored as Wikidata items into a BibTex file.

Research

Finding news citations for Wikipedia

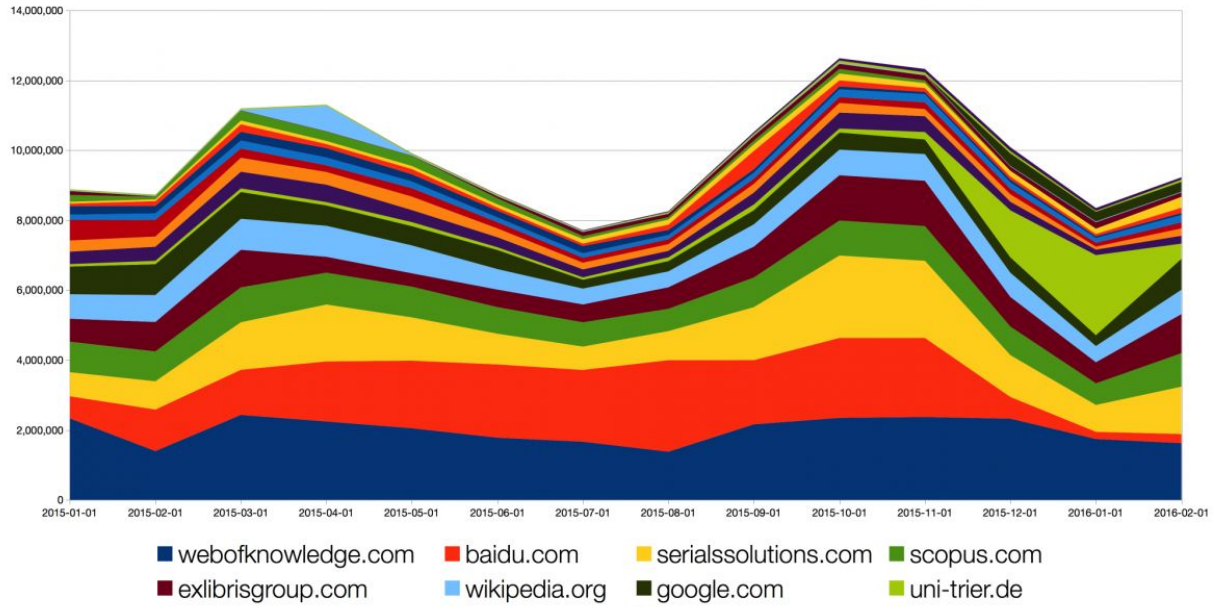
Besnik Fetahu (Leibniz University of Hannover) presented his research on news citation recommendations for Wikipedia at the Wikimedia Research showcase (slides, video). In his own words, "in this work we address the problem of finding and updating news citations for statements in entity pages. We propose a two-stage supervised approach for this problem. In the first step, we construct a classifier to find out whether statements need a news citation or other kinds of citations (web, book, journal, etc.). In the second step, we develop a news citation algorithm for Wikipedia statements, which recommends appropriate citations from a given news collection."

DBpedia Citation Challenge

Krzysztof Węcel (Poznań University of Economics and Business) presented his research (slides) in response to the DBpedia Citations and References Challenge, analyzing content in Belarusian, English, French, German, Polish, Russian, Ukrainian and showing how citation analysis can improve the modeling of quality of Wikipedia articles.

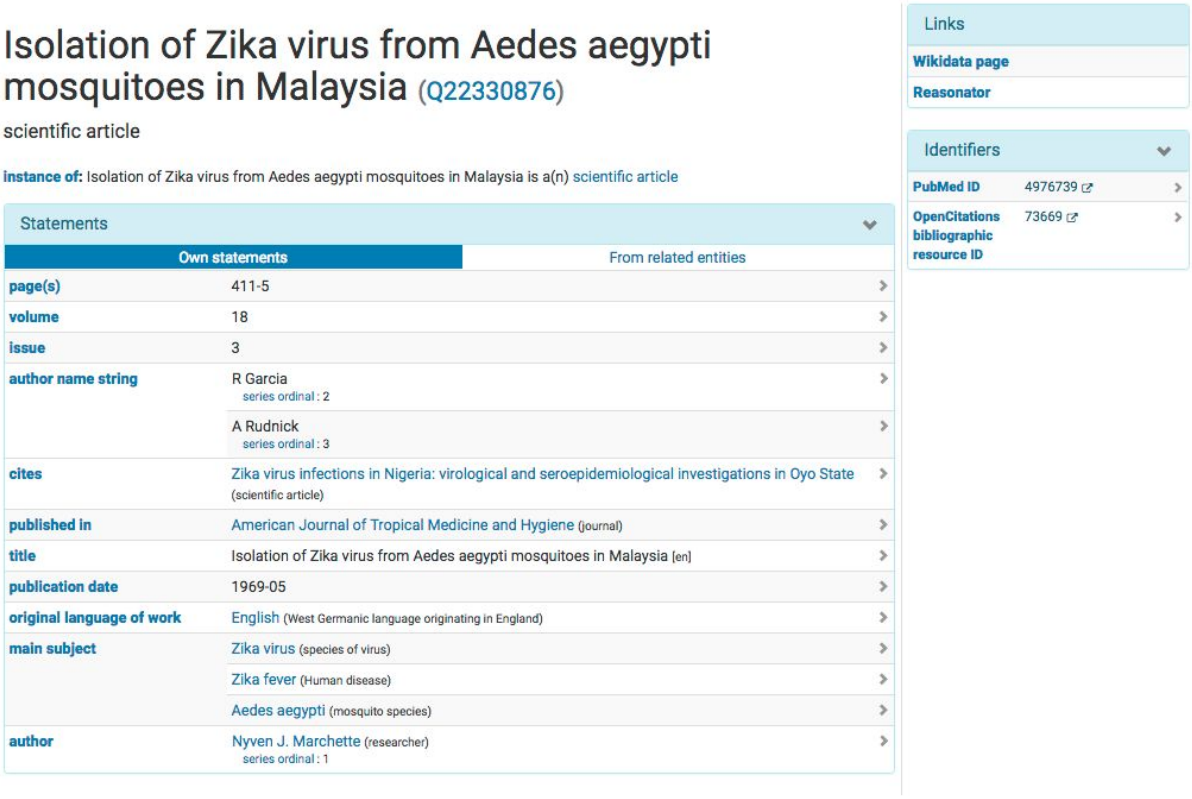
Appendix

Wikipedia as one of the top sources of DOI lookups



http://crosstech.crossref.org/2014/02/many-metrics-such-data-wow.html
http://blog.crossref.org/2016/05/https-and-wikipedia.html

A sample bibliographic entry in Wikidata



Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia (Q22330876)

scientific article

instance of: Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia is a(n) scientific article

Statements

Own statements	From related entities
page(s)	411-5
volume	18
issue	3
author name string	R Garcia series ordinal : 2
	A Rudnick series ordinal : 3
cites	Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State (scientific article)
published in	American Journal of Tropical Medicine and Hygiene (journal)
title	Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia [en]
publication date	1969-05
original language of work	English (West Germanic language originating in England)
main subject	Zika virus (species of virus)
	Zika fever (Human disease)
	Aedes aegypti (mosquito species)
author	Nyven J. Marchette (researcher) series ordinal : 1

Links

Wikidata page

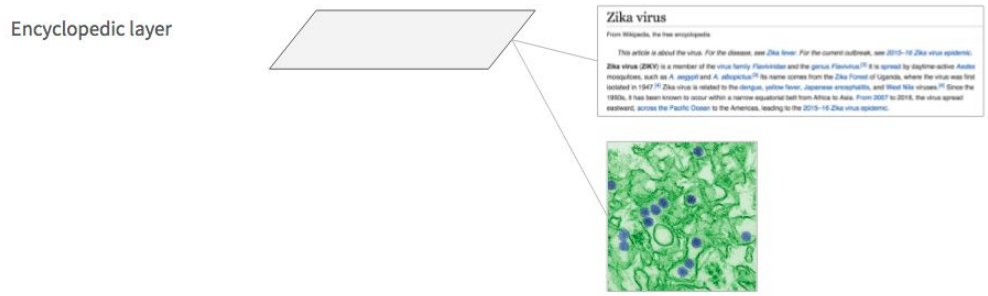
Reasonator

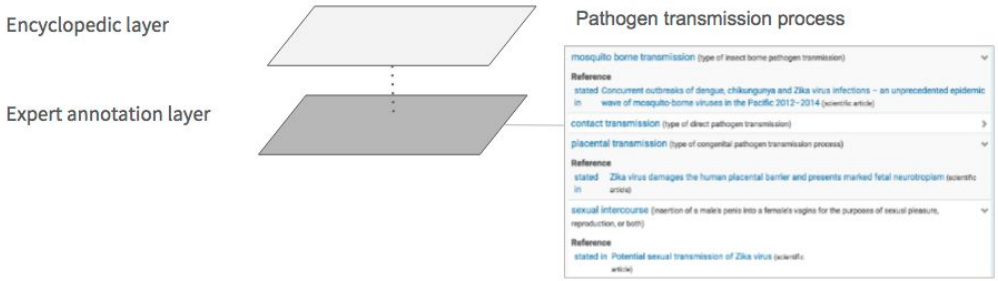
Identifiers

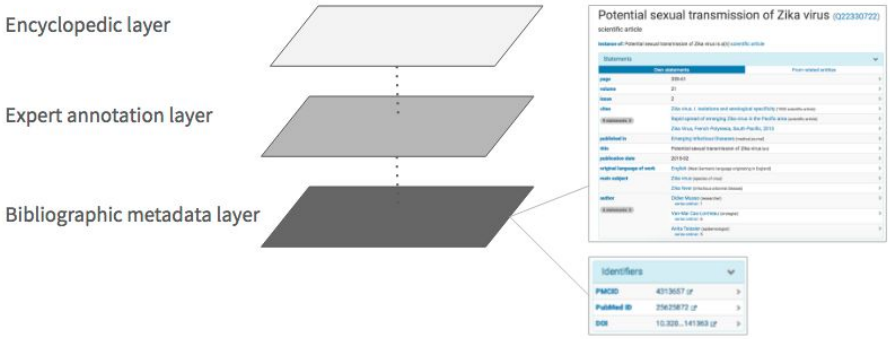
PubMed ID	4976739
OpenCitations bibliographic resource ID	73669

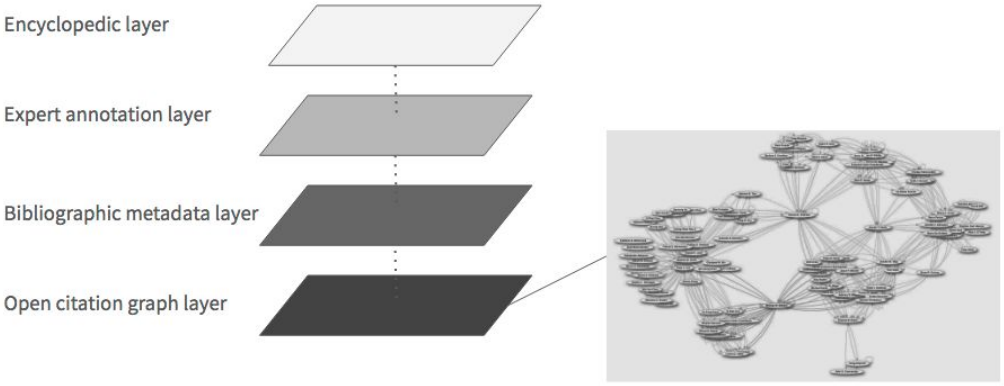
https://tools.wmflabs.org/sqid/#/view?id=Q22330876

The Zika corpus in Wikidata

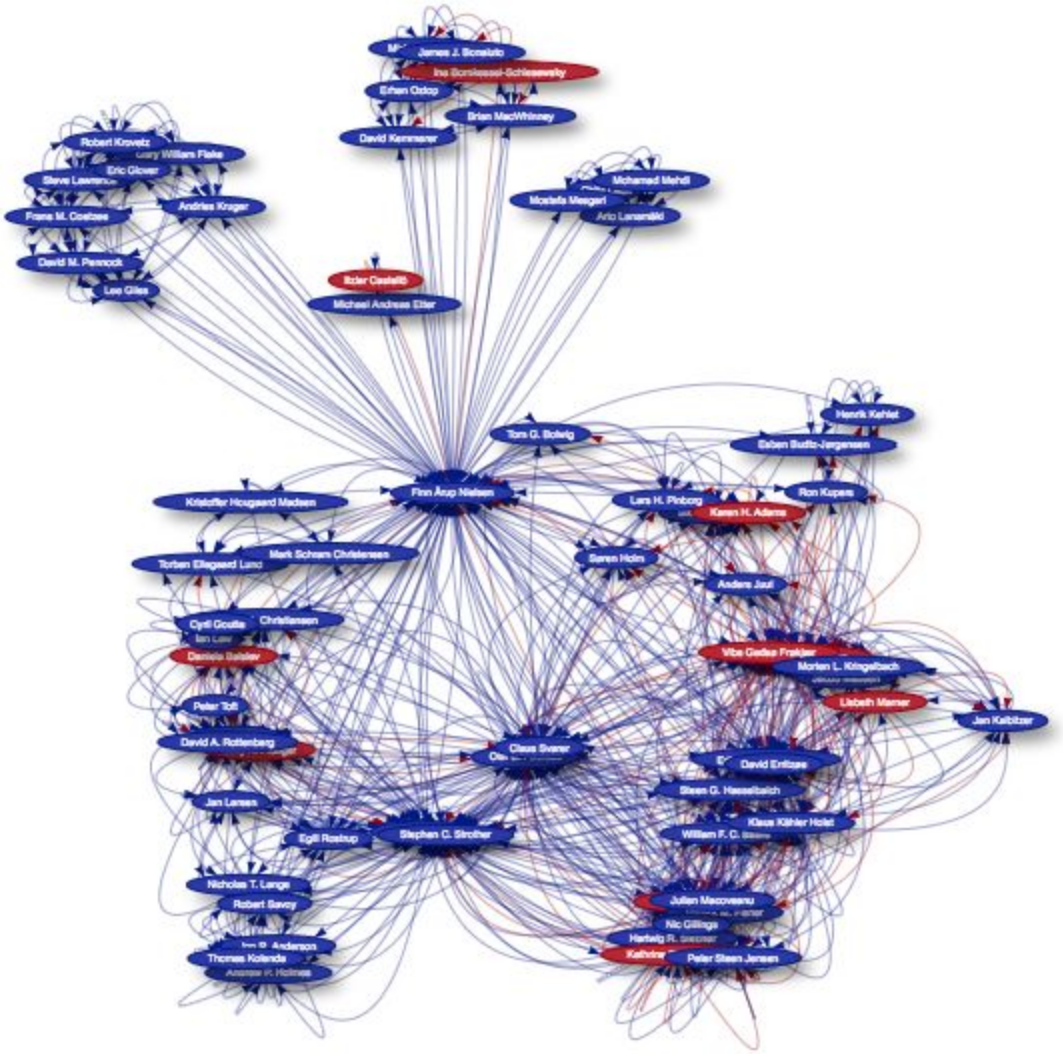






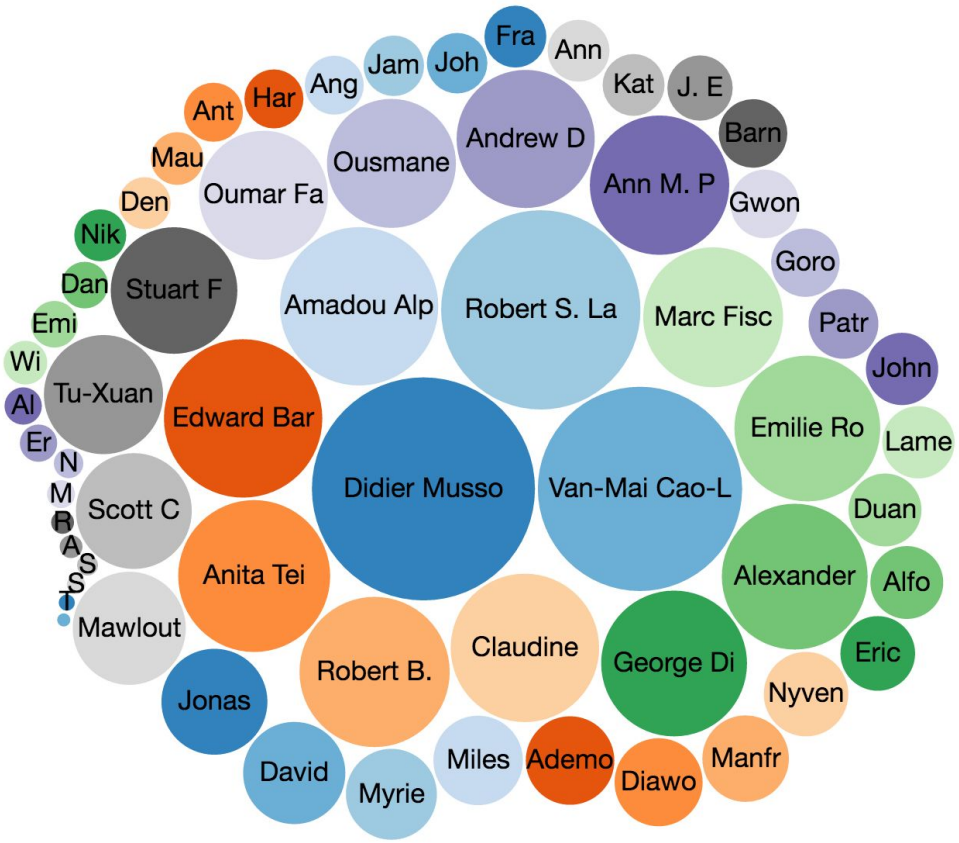


A coauthor graph generated from Wikidata



SPARQL query source: http://tinyurl.com/ztnov3r

Most cited authors in the Zika corpus in Wikidata



SPARQL query source: http://tinyurl.com/jb8da68