{} wikicite

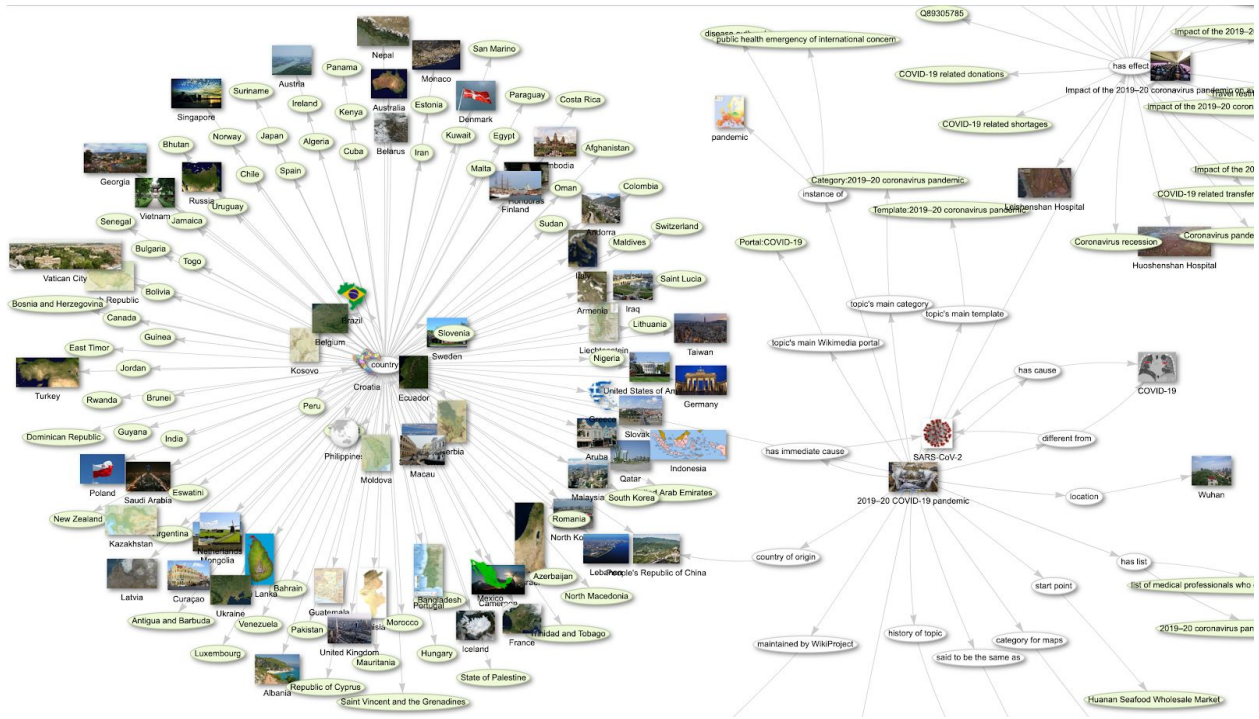*Citations for the sum of all human knowledge.*
Annual Report, 2019–20.

L. Wyatt – *Program manager*
P. Ayers, M. Proffitt, D. Mietchen, E. Seiver, A. Stinson, D. Taraborelli – *Steering committee*
C. Virtue, J. Tud, J. Curiel – Wikimedia Foundation staff

Partial neighbourhood graph of the COVID-19 Pandemic in Wikidata, April 2020.
By Daniel Mietchen. As visualised by the Wikidata Query Service. CC0

About

WikiCite is an initiative aiming to build **a comprehensive knowledge base of sources**, to serve the sum of all human knowledge. This report examines the impact, key milestones, and reach the WikiCite community has achieved over the course of the 2019-2020 fiscal year. Previous annual reports can be found at the WikiCite homepage on Meta Wiki.

The Alfred P. Sloan Foundation generously supported the 2018 WikiCite conference and WikiCite activities through 2019 and 2020. Additional financial and logistical support provided by the Wikimedia Foundation.

Introduction

Wikipedia is the world's largest, most widely used online encyclopedia. It is free and open, a vast body of knowledge anyone can contribute to. Wikipedia relies on policies that put a premium on verifiability of the information it holds, a commitment to citations, fact-checking, and accuracy.

How does the Wikimedia movement empower individuals to assess reliable sources and arm them with quality information so they can make decisions based on facts? How do we identify bias or distortions in the application of these verifiability policies? These questions are relevant not only to Wikipedia users but to consumers of media around the globe.

Over the past decade, the Wikimedia movement has come together to answer that question. Efforts to design better ways to support sourcing have coalesced around Wikipedia's sibling project Wikidata – the free knowledge base that anyone can edit. With the creation of a rich, human-curated, and machine-readable knowledge base of sources in Wikidata, the WikiCite initiative is crowdsourcing the process of vetting information and its provenance.

The number of items in Wikidata representing sources has grown dramatically since the first WikiCite conference in 2016, with almost 30 million items about individual publications in Wikidata and over 160 million known citations – where one of those known publications references another – making an ever more interconnected knowledge graph. With this growth has come a large number of innovative tools to work with this data. These tools and experiments have been developed by a community of volunteer Wikimedia developers, librarians, and linked data experts and enthusiasts, and together show the power and potential for this new source of open bibliographic data. These WikiCite-related projects have been presented and discussed all over the world. Since 2018, the library community has shown increasing interest in using Wikidata and Wikibase (the underlying software of Wikidata) in library-related linked data systems, with several experimental projects being developed, notably by several European national libraries. The WikiCite community uses an active mailing list (with 270 members) and social media platforms (with over 4,000 followers of the @Wikicite twitter account), as well as discussions on Wikidata itself.

WikiCite Satellite events 2020

From 2016 to 2018, an annual WikiCite conference was held to support this community: Berlin 2016, Vienna 2017, and Berkeley 2018. For the 2019-20 fiscal year, the steering committee behind the most recent edition of the WikiCite conference series decided to diversify the content and format of WikiCite events and to increase its accessibility and equity to a greater range of contributors.

WikiCite content and languages have remained focused on the Global North. The tools and best practices that Wikidata and the WikiCite community have developed are not known to many communities which need them. Additionally, participants of past events felt there was not enough time to work on their projects, suggesting a need for more events focused on workshops, hack days, and 'do-athons'. To expand our community and make it more diverse in both participant and subject area, we need to assist and facilitate the community to organise events around the world to foster interest in the project….

- 2020 Satellite events call for proposals

Consequently for 2019-2020, the steering committee decided that WikiCite would not be a single conference but a series of **"WikiCite Satellite events" around the world in the first half of 2020**. To achieve this: the funding from the *Alfred P. Sloan* and *Wikimedia* Foundations was repurposed into small grants; a call for proposals was made to the global Wikimedia community; and a *Program Manager* to administer the satellite grants was hired with a portion of the funds.

In mid-December 2019, the call for proposals for 2020 Satellite events was launched, with a total budget allocation of $38,000 split flexibly between two financial categories:

1. International events in association with professional conferences

    Maximum $10k per grant, deadline 1 February. Submissions were especially encouraged if they related to marginalised knowledge and/or were hosted at a location in the global south.

2. Local workshops and meetups

    Maximum $2k, open until funds are allocated. Priority was given to languages, regions and subject areas that are currently poorly served by Wikidata.

In total, 15 applications were received from all continents. Following committee review, **nine WikiCite Satellite event grants were approved** with a combined value of $35,278. The successful grants were, with descriptions from their own applications:

From category 1 –

- *World Library and Information Congress (WLIC)* hosted by the International Federation of Library Associations and Institutions (IFLA), **Dublin** Ireland, August. $10,000.

  Organised by the official IFLA-Wikidata working group, a two-day, single-track event, with one day of invited workshops and one day of invited presentations, immediately following WLIC 2020 – the preeminent global conference for the international library sector. Its aims were to strengthen a community of practice focused on the support of open sharing of library metadata, and especially bibliographic and authority data.

- *WikiCite Satellite Cologne* hosted by organized by ZB MED - Information Centre for Life Sciences, GESIS – Leibniz Institute for the Social Sciences, TH Köln – University of Applied Sciences, and Verbundzentrale des GBV. **Cologne** Germany, May. $8,500.

  Independently organised by some of the most active WikiCite community members in Germany, and in the spirit of the 2015-18 annual conferences, a three-day event in German and English. The event aimed to connect local library institutions and academic researchers with the WikiCite community in their shared interest in an open infrastructure about open citations and linked bibliographic data for research and education.

- *Consortium of Academic and Research Libraries in Ghana (CARLIGH) International Conference* hosted by Global Open Initiative Foundation. **Accra** Ghana, July. $5,675.

  A booth during the main conference followed by a two-day training workshop, intended to establish a relationship between the WikiCite Community, the Wikimedia community in Ghana and in the West Africa region, and the attendees' affiliated institutions.

From category 2 –

- *2020 WikiCite ANZ*, hosted by Wikimedia Australia. **Melbourne** Australia, February. $1,370.

  Held in association with the annual national library conference *VALA*, a one-day event featuring the keynote speaker from the main conference, panel sessions and editing workshops on the topic of Australian and New Zealand indigenous languages, authors, academics and subject areas currently poorly served by Wikidata, Wikipedia and related projects. Participants were invited to work on ways to address these gaps. [Dashboard of editing outcome statistics; Slides and presentation documentation]

Participants of 2020 WikiCite ANZ. By Canley CC BY-SA 4.0

- *BibioWorkshop*, hosted by Wikimedia Czech Republic. **Prague** Czechia, March. $2,000.

  A two-day training camp for librarians – the first Czech event on the topic of collaboration between Wikidata and the Czech library network. Bringing together 12 experts working in the field of bibliographical open structured data, and Czech Wikidata contributors, with workshops to focus on syncing national and regional library authority records with Wikidata, and technical training on the current state-of-the-art import and data-query tools.
  [Event report (Czech)]

- *WikiCite Satellite Delhi*, Hosted by the Open Heritage Foundation. **New Delhi** India, April. $2,000.

  Independently organised by some of the most active WikiCite community members in India, and in the spirit of the 2015-18 annual conferences, a two-day event in Hindi and English. The event aimed to focus on the South Asian region and address the prevalent gap of Wikidata and Wikisource integration for works and their consecutive items.

- WikiCite Satellite Haiti, hosted by Community Wikimedia User Group Haiti. **Port-au-Prince** Haiti, April. $1,760.

  A one-day workshop for librarians of the capital region, aimed in particular to enrich the Wikidata elements that concern Haitian Studies (authors, works, achievements, awards, etc.) and Haitian universities and interconnect them, with Wikidata and Wikipedia pages. This workshop addresses the need to not only train Wikidata contributors in Haiti, but also to improve the presence of this country in Wikidata.

- Open Foundation West Africa, **Accra** Ghana, April-June. $1,980
  A one-day WikiCite workshop in April for attendees to contribute Ghanaian content in open citations and bibliographic metadata into Wikipedia and Wikidata. Following the workshop: a one-month competition, in association with the global 1Lib1Ref campaign, to add qualified citations for inline references and linked metadata especially focused on African-related content.

- Saami vocabularies trips, organised by Wikimedia Finland. **Sevettijärvi** Finland, April. $1,992.
  A two-trip expedition to the Northern Saami communities in association with The Sámi Archives, Siida the Sámi Museum, Yle Sápmi [broadcaster], Skolt Sámi Cultural Foundation, and the Sámi Giellagáldu parliamentary body. Trip one would focus on importing pilot Saami vocabularies to Wikidata followed by public workshops where participants engage in considerations for privacy, respect for traditional knowledge, and cultural protection on Wikimedia projects. Participants would also discuss cultural miss-use through commercial and privacy considerations. Trip two would encourage the public to join tagging content with the vocabularies added to Wikidata and complement them with their own knowledge.

COVID-19 changes

In mid-March [the Wikimedia Foundation sent notice](#) to all affiliates and grant recipients that, subsequent to the WHO declaration of a global coronavirus pandemic, all in-person events must be cancelled or postponed until *at least* mid-September. This was later expanded such that no *new* grants for in-person events would be approved until 2021.

Consequently, the organisers of all approved WikiCite Satellite events were requested to either: **cancel**, indefinitely **postpone**, or completely **adapt** their project to meet the original spirit of their proposal in a digital format. Fiscal sponsors the *Alfred P. Sloan* and *Wikimedia* Foundations were kept apprised of developments during this time. Two approved WikiCite events had already occured: Australia's *VALA* conference event, and Czech Republic's *BiblioWorkshop*. Consequently, they are the only two for which event-reports exist. Organizers of the latter event were forced to cancel day two as the new rules were promulgated with immediate effect in the middle of their event. The decision taken by the other approved events were:

Cancelled: Accra [category 1], Cologne, and New Delhi
Postponed: Accra [category 2] and Port-Au-Prince
Adapted:

- IFLA (Dublin) has opted to produce a series of long-form training videos, to be released at the end of 2020 and with a decreased budget of $9,000. As the grantee describes:

While the delivery formats of the event have changed, the theme of the event remains the same, that is to support and activate the professional library communities' efforts to foster knowledge equity through the development of open bibliographic data to increase the visibility of knowledge and intellectual traditions that remain marginalized on an internet dominated by certain cultures as well as decrease barriers to linked open data in libraries around the world.

The digital resources to be developed and shared will be for sharing 'best practices,' tools, and techniques which could further help communities share knowledge by overcoming technical and resource constraints and become contributors themselves.

Proposed activities keep close to the originally proposed activities, consisting of a mix of workshops and talks. However, we proposed to deliver content in an asynchronous format to better support participation of a distributed, international community. Packaging material this way also removes dependence on a stable, continuous internet connection.

- Wikimedia Finland (Sevettijärvi) has opted to prepare datasets in close collaboration with the original partner institutions/communities. As the grantee describes:
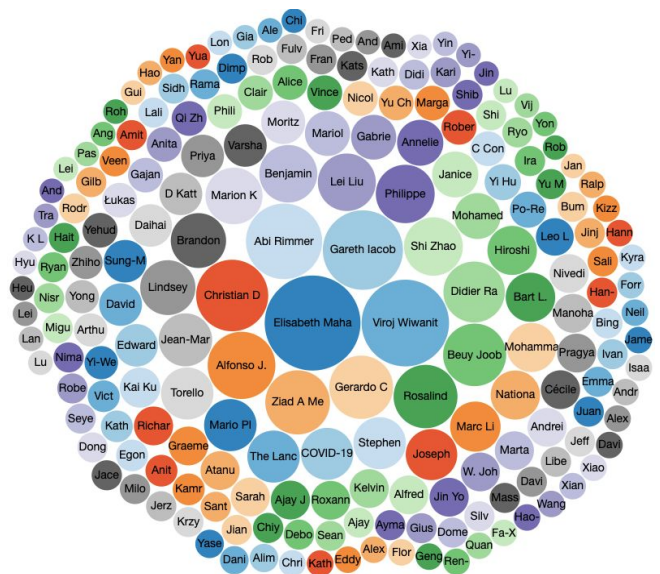
We will address this task though the concept of Traditional Knowledge, and the rights of the indigenous cultures to manage their online representation. The aim is to connect the work to an international discussion about the shortcomings of the copyright regime in defining the rights of sharing indigenous content. As a result we wish to identify practical working methods and tools for the Saami communities, GLAM organisations and the Wikimedia communities for sharing Saami content online in respectful and consensual ways.

The project is a collaboration between Saami cultural organisations and Wikimedia Finland. In the process we will go through steps to engage Saami community members in contextualizing archival material and identifying the conditions for sharing the knowledge publicly. The resulting terminology will be prepared for Wikidata to be used for multilingual tagging. We will create best practices for presenting Saami content in Wikimedia projects.

The current *Wikimedia Foundation* guidelines on non-essential travel and in-person events (for both itself and its grantees) extend through at least mid-September, and will be in effect until the World Health Organization declares an end to the global pandemic. Also, no *Wikimedia Foundation*-sponsored international in-person events will occur until 2021. Consequently, at the time of this report, the WikiCite "satellite event" grant program is being revised for the second half of 2020 as we adhere to restrictions on physical gatherings. Exemplified by our work with IFLA, which is now producing long-form training videos, we continue to support groups to produce online events and training materials that meet the ideals of the *Alfred P. Sloan* grant and the overall WikiCite initiative.
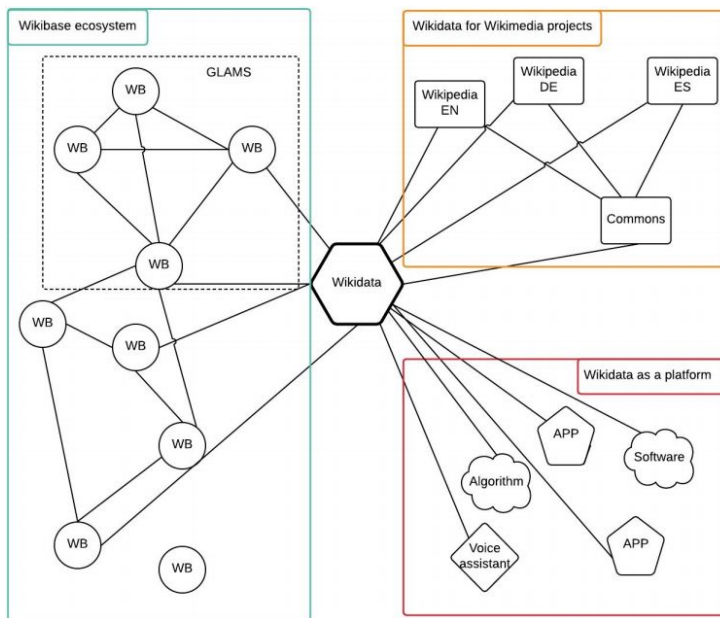
News from the movement

The WikiCite satellite grants are only a small part of the story. Over the past year, the WikiCite community, partners, and the professional and volunteer contributors to Wikidata (and its underlying software Wikibase) were part of a number of important initiatives, including strategic planning, events and conferences, and content creation or re-use projects.

Author score for SARS-CoV-2. *Visualised by Wikidata Query service, via Scholia.* CC0

Strategy

- [Wikimedia movement strategy](#). After several years of preparation, the consultation and drafting phases of the Wikimedia 2030 strategy process have concluded and the [final recommendations](#) published in May 2020. Built upon the strategic direction statements "Knowledge as a Service" and "Knowledge Equity," the 10 major outcomes will take significant time and commitment from all stakeholders in the Wikimedia movement to implement. They speak of a Wikimedia movement "ecosystem" – of technological platforms, processes, and organizational structures – that is built upon the principles of equity, inclusivity, collaboration, transparency, efficiency, and resilience. As the Wikimedia movement seeks to become the "essential infrastructure to the ecosystem of free knowledge" the purpose of WikiCite – to develop open citations and linked bibliographic data to serve free knowledge – becomes ever more crucial.

- In August 2019, the development team of Wikidata and Wikibase published four [strategic direction whitepapers](#): the Vision for Wikidata/Wikibase; the role of Wikidata for Wikimedia projects; the role of Wikidata as a platform in its own right; and the future of the Wikibase ecosystem. Working with libraries – both as a pragmatic test-case community, and also an important stakeholder in open data – features heavily in the plans in several aspects: Content quality and community growth; federated Wikibases through library catalogues; and authority control through data roundtripping. Consequently, the work of WikiCite in building partnerships (human, institutional, and data) is highly connected to this strategy.

The Wikidata and Wikibase ecosystem. *By Pintcher et al.* CC BY-SA 4.0

Conferences

- *Wikimania*, August 2019. Stockholm Sweden.

The annual *Wikimania* conference was held with a [programmatic focus](#) on the UN's Sustainable Development Goals. For the first time, Wikimania was designed around independently curated tracks of content of which "[libraries](#)" was one. [Wikidata/Wikibase featured](#) throughout all tracks of content (as well as bibliographic metadata issues), including during the pre-conference [hackathon](#), continuing the trend of Wikidata being the most popular field for submissions and acceptances. Following the main conference a special "[libraries & wikidata meeting](#)" was held, the first in-person gathering of all national libraries currently experimenting with Wikidata and/or Wikibase for metadata management. The program chair for Wikimania and the post-conference meetings is the WikCite program manager; the chair of the library track is a WikiCite steering committee member.

- *WikiDataCon*. October 2019. Berlin Germany.

The second biennial Wikidata conference, held to coincide with the project's seventh "birthday," featured many demonstrations, training sessions, discussions, and research presentations relating to bibliographic metadata and libraries. The [program outcomes](#), including a panel discussing the WikiCite satellite grant program itself, held a focus on knowledge equity and external content partnerships.

- *Wikimedia Remote Hackathon*. May 2020.

Originally scheduled for Tirana Albania, the annual community technical event was [successfully moved online](#). Featuring many of the WikiCite volunteer community working on projects relating to structured citation and metadata information, many of which resulted in [demonstrated prototypes](#)), the event was the first truly global virtual event in the Wikimedia community post-COVID-19, and will be emulated by others.

Remote hackathon 'group photo'. By Tohaomg. CC BY-SA 4.0

- *LD4 Conference on Linked Data in Libraries*. May 2019. Boston USA.

A major library-sector event with significant crossover community to WikiCite, the conference featured in the program a Wikimedian keynote, several Wikidata tutorials, and presentations on mapping to specific library metadata standards.

- *SWIB Semantic Web in Libraries*. November 2019. Hamburg Germany.

An annual event in Germany, its 2019 iteration featured sessions specifically on WikiCite; the use of Wikidata for vocabulary translation; authority control matching; and integration with tools such as OpenRefine.

Other major events that included WikiCite content, presenters, and/or program committee members, which were moved entirely online due to the global COVID-19 pandemic, included:

- WikiWorkshop 2020, originally scheduled for April in Taipei, Taiwan.
- OpenCitationsCon, originally scheduled for Bologna, Italy.
- LD4P conference, originally scheduled for Texas, USA.

Projects

The global volunteer and professional communities that contribute to and use Wikidata and Wikibase – both directly and through third-party platforms – continue to grow and so does the density and quality of data relating to citations. A few recent projects with connections to WikiCite worthy of note include:

- WikiProject COVID-19 on Wikidata. An incredibly active grassroots project to create and maintain a thorough, multilingual and up-to-date data model and corpus citations relating to the coronavirus.
- LD4 Wikidata affinity group. Hosting fortnightly public calls featuring expert demonstrations, the group's goal is understanding how libraries can leverage Wikidata as a platform for publishing, linking, and enriching library-linked data.
- OpenCitations.net, the infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data, announced its index of Crossref surpassed 700 million citations, and has published a COVID dataset.
- Scholia report on SARS-COV-2. Building on the success of the Zika virus research corpus, the Scholia visualisation platform collates Wikidata's content relating to the novel coronavirus including 3,400+ authors, 500+ publications, and 1,300+ venues.