Proposal

Shared Citations

Prepared by Liam Wyatt / Wittylama
2020
Agenda

- Opportunity space
- How it helps users
- How it supports our priorities
- What we might do
- How we might do it
- Potential resources needed
Opportunity Space

- Citations
- The problem
- Current state
Citations are the core of **anti-disinformation** and **knowledge integrity** in our movement. Our **Verifiability** policy has become the backbone of the reliable web.

“Citations are a simple, critical interconnection mechanism for all modern knowledge in the digital, Internet-connected world ... arguably the most important ingredient of open knowledge, sources and references have received little technical attention in the Wikimedia movement up until now.”

–2016 WikiCite report
The Objective

“Make citations easier for the editor, more useful for the reader, and more efficient for our architecture”
The Problem

Wikimedia’s citations are one of our greatest assets, but by storing them as raw “inline” text in each content page, they are also one of our biggest burdens.

Our citations are high in maintenance, technical complexity, and duplication of effort.

This burden is shouldered by repetitive, manual, volunteer effort which is disproportionately felt by smaller communities.

Participants at WikiCite 2018 debating this problem. We’ve known this is an issue for a long time.
References in MediaWiki (WP)

Closures and cancellations

Through the first quarter of 2020, arts and culture sector organisations around the world progressively restricted their public activities and then closed completely due to the pandemic. Starting with China, East Asia, and then worldwide, by late March most cultural heritage organisations had closed, and arts events were postponed or cancelled, either voluntarily or by government mandate. This included galleries, libraries,[1] archives,[2] and museums[3][4][5] (collectively known as GLAMs), as well as film[6] and television productions,[7] theatre[8] and orchestra performances,[6] concert tours,[10] zoos,[11] and music[12] and arts festivals.}[9][13]
References in Wikipedia

<ref name = "koppen">
{{Cite journal
| title = Updated world map of the Köppen-Geiger climate classification
| year = 2007
| journal = Hydrology and Earth System Sciences
| volume = 11
| pages= 1633–1644
| date =<!!- 01/30/2016>
| access-date=30 January 2016
| doi = 10.5194/hess-11-1633-2007
| first1 =M C| first2 = B L| first3 = T A| last1 = Peel| last2 = Finlayson| last3 =McMahon
| doi-access = free}}
</ref>
## References in Wikibase (WD)

Reference fields are stored and edited independently. This is even when the same information is identical, on the same property, on the same item.

### Example:

<table>
<thead>
<tr>
<th>ID</th>
<th>Date</th>
<th>Reference URL</th>
<th>Title</th>
<th>Publisher</th>
<th>Language of work or name</th>
</tr>
</thead>
<tbody>
<tr>
<td>212</td>
<td>August 2014</td>
<td><a href="https://wikimediafoundation.org/wiki/Template:STAFF-COUNT">https://wikimediafoundation.org/wiki/Template:STAFF-COUNT</a></td>
<td>Template:STAFF-COUNT (English)</td>
<td>Wikimedia Foundation website</td>
<td>English</td>
</tr>
</tbody>
</table>

Source: [Wikidata: Wikimedia Foundation (Q180)](https://wikidata.org/wiki/Q180)
Re-used references

**Specific Source templates**
Local templates for individual, frequently reused works

- **English WP:** 452 templates for “rail transport books” e.g.

  ```
  {{Anchen-Iron Roads | page=136}}
  ```


- **French Wikipedia:** References Namespace. e.g.


  6. † Homère, *Iliade* [détail des éditions] [lire en ligne [archive]], IV, 2-3.

  Links to manually compiled listing of bibliographic details of all major editions of Pindare’s *Odes*; Homer’s *Iliad*

**Cite Q**
Lua module to display a Wikidata items as a reference

- **English WP:** Used in 43,000 pages
  (mostly inside infobox templates)

  ```
  {{Cite q|Q15625490|page=42|access-date=18 May 2017|quote=lorem ipsum}}
  ```


- **Proof that shared citations is technically possible, and the challenges they still face**

  Case studies on English Wikipedia:

  - “South Pole Telescope”
  - “Suffix Automaton”

Source: [https://fr.wikipedia.org/wiki/H%C3%A9b%C3%A9](https://fr.wikipedia.org/wiki/H%C3%A9b%C3%A9)
"There are 4.5 million unique sources in the datasets. While on average, every source is cited 3.5 times, the vast majority of sources in this dataset are used less than 500 times across wikis. Nine “super publications” are used more than 10,000 times."

– What are the ten most cited sources on Wikipedia? Let’s ask the data. [WMF Blog](https://www.wired.com/story/wikipedia-most-cited-authors-no-idea/) (2018)

[https://www.wired.com/story/wikipedia-most-cited-authors-no-idea/](https://www.wired.com/story/wikipedia-most-cited-authors-no-idea/)
[https://www.nature.com/articles/d41586-018-05161-6](https://www.nature.com/articles/d41586-018-05161-6)
Citation duplication

According to that 2018 research, the most reused citation is used 2,830,341 times:


2.8m citations @ 500 characters per citation = 1.4 billion characters

= 1.4GB for a single citation.

This ~500 characters is hand curated by volunteers, often with semi-automated tools.
How it helps users (user stories)

1. Knowledge creators
2. Knowledge consumers
Knowledge creators
(Editors / Patrollers / Researchers)

want to easily access & curate citations,
to improve content quality & integrity
Knowledge creators

As a Wikimedian editor, I would like to

- Quickly and easily add references to my article by re-using references from another language Wikipedia
- Benefit from the formatting done by others, so I can spend my time on research and writing, not on templates
- Train new-users to add footnotes in one session, in a way they can continue without intense support
- Create ‘redlists’ for Wikidata of authors and publications which are frequently cited but don’t have a Wikidata item
- Have my work on disambiguating authors on Wikidata cascade through to their references on other Wikiprojects

As a content patroller, I would like to

- Tag all citations to an instance of dis/misinformation or a retracted publication across all projects
- See, when someone is adding the same link across many different projects in quick succession, and revert
- Identify and track any citations to predatory journals

As a professional writer, I would like to

- Have academic ‘impact factor’ reports include Wikipedia citations to my scholarly work
- Check when Wikidata is referencing my work to ensure its findings are accurately represented
- Be notified when a Wikipedia article cites my journalism and be able to share it on social media
Knowledge consumers
(Readers / Writers / Reusers)

want to easily understand and access citations, and to have confidence in them.
Knowledge consumers

As a reader, I would like to

- Understand if I can trust the citation I am reading
- Be suggested other topics in Wikipedia which also reference this same author/book
- See dictionary definitions which prioritise usage examples published in my country’s vernacular
- Generate a list of primary, or newspaper, or local sources used on this topic, for my high-school homework

As a researcher, I would like to

- Be able to extract parts of the citation corpus and analyse it, without massive pre-processing
- Know how many references are behind paywalls, how much they cost, and if there are alternatives
- See how many references are about the language/location, but published in another language/location

As a ‘big tech’ company, I would like to

- Be able to answer the question “says who?” when a customer asks me to verify a fact just given them
- Train my algorithm to show more reliable sources for languages/topics where I have limited other data

As a library, I would like to

- Track that Wikimedia external links to our collection from Wikimedia are well maintained
- Notice which books on our key subjects are cited and ensure our library has holdings
How it supports our priorities

1. Strategic directions
2. OKRs
3. WMF teams
“This will help us achieve epistemological decolonisation”
- João Peschanski (Wiki Movimento Brasil)

“If libraries could look at global Wikimedia citations, it would help break a self-reinforcing cycle of certain sources’ popularity in library holdings”
- Phoebe Ayers (MIT Libraries)

“Standardized templates and references benefit products designed for emerging markets, which need high interoperability of content form-factor, language, and projects.”
- Runa Bhattacharjee (WMF)

“Having systematic insight into where our knowledge comes from will help us to diagnose what kinds of sources, languages, voices are missing.”
- Ben Vershbow (WMF)
“Commons makes it easy to use the same image on different wikis. But to copy a citation is very hard. Everything is manual and slow.” - Amir Aharoni (WMF)

“Imagine being able to recommend useful footnotes to editors, readers.” - Sam Walton (WMF)

“This adds integrity to our citations - “a fortified citation layer.” - Chris Albon (WMF)

“Abstract Wikipedia articles will be far more useful if references are formatted from structured data instead of plaintext.” - Denny Vrandečić (WMF)
Why us? Why now?

OKRs
- **Platform Evolution**: Software platforms with structured data, reuse of code & content
- **Thriving Movement**: We will support diverse content creation
- **Worldwide Readership**: Substantially extend our core product experiences
- **Brand Awareness**: Amplify the unique characteristics of Wikimedia
- **Global Advocacy**: Build technical & community structures that exemplify our free knowledge policy agenda

Synergies
- We are prioritising development in **structured data**, product & platform integration
- addresses some existing needs directly (both in WMF & WMDE teams), lowering their project costs and extending their functionality
WMF use cases

Abstract WP – “if you can build the monitoring system, Abstract Wikipedia will be using it” – Denny Vrandečić

Architecture – “References as 1st class objects: This is consistent with a ‘software to systems’ model” – Kate Chapman & Moriel Schottlender

Campaigns – “Systematically leveraging relevant sources used elsewhere in our ecosystem to close knowledge gaps” – Felix Nartey

Citoid – “This could make creating better quality and more content-rich citations easier.” – Marielle Volz

Content Translation – “When adding refs in translated articles is like adding images, translators can focus on prose.” – Amir Aharoni

Okapi – “Many high volume reusers would like to be able to parse our citations consistently” – Lane Becker

ORES – “This would be a great training dataset, ‘Editors who add these kinds of footnotes do those kinds of actions’” – Chris Albon

Mobile – “This would make features like the ‘Featured reference’ prototype on iOS viable at scale” – Josh Minor

SDAW – “This integration would be directly applicable to SDAW and connecting reliable content across wikis.” – Amanda Bittaker

The Wikipedia Library – “We could have a field for Proxied URL allowing TWL cardholders direct access” – Sam Walton
What we might do

1. The proposed solution
2. The principles
3. Examples
Proposed solution

- A service product: central database of Wikimedia citations.
- Software integration: improvements to inter-Wiki monitoring and editing.

These two pillars would empower community-managed workflows and tools to address all the aforementioned user-stories.
Principles

**Add upon need**
Citation items are created by community members, when they are being used for a reference on a wiki. Not pre-created by an automated process.

**Enabled upon readiness**
Communities should be supported to be ready to it for it to be enabled on their Wiki, and benefiting from a network effect.

**Non-deprecation**
Preserve existing referencing systems. Structured references being enabled on a Wiki does not prejudice ‘traditional’ references.

**Style independence**
Local wikis determine the display format of a reference for the reader and can be modified by local tools and templates.

**Editorial independence**
Local wikis determine their content standards. The existence of a shared citation in the database does not prejudice another project’s policies.
Community Concerns

The two biggest pain-points expressed by the community, which hinder efforts to use structured-data in the Wikimedia ecosystem, are the current inability to easily:

**Monitoring**

Be informed of changes which affect how content is shown here, but was edited over there.

**Editing**

See (and ideally change) the content over there, without being required to leave here.

The status quo can be understood as incomplete ecosystem integration.
1. Monitoring Principles

**Granular**

Watchlist and Page-history must record and reflect every change that affects what is shown to a reader of that page, and only those changes.

**Arbitrary**

The citation information being changed and displayed on the client wiki (e.g. Wikipedia) might come from any Wikidata item referenced in the citation.

**Cascading**

A dependency tracking system must propagate notifications through any and all affected items across sites (Wikipedia article, citation item, Wikidata item), regardless of the origin of the change.

**Broader cross-wiki integration benefits**

Building a unified “Dependency engine” for Shared Citations would be directly applicable other cross-wiki integration like Wikidata Bridge, WikiLambda, Global templates, and longstanding requests in Commons.

More details available
2. Editing Principles

Progressive discovery

An editor must be able, but not required, to know where the citation is hosted, and easily access it.

Visual editing

An editor should be able to search, view the content of the citation, and when possible, edit it in place – in all editing environments.
Examples
<table>
<thead>
<tr>
<th>Citation</th>
<th>APA</th>
<th>MLA</th>
<th>MHRA</th>
<th>Chicago</th>
<th>CSE</th>
<th>Bluebook</th>
<th>AMA</th>
<th>BibTeX</th>
<th>wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>String as published: “Lyndall Gordon” First: Lyndall Last: Gordon Initial: L (Q6708612).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>String as published: “Harper Perennial” (Q5663419)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source type</td>
<td>Book – Secondary source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date/Edition</td>
<td>2005. 3rd, Hardcover. (4 other editions cited across Wikimedia - with 12 citations)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edition of</td>
<td>Wikidata item of overarching work: Q123456789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Citelinks**


By time
- First: 11 October 2007, User:CaptainSaru, on English Wikipedia - Feminist philosophy (104 other links expand)

By section
- Page 34: Mary Wollstonecraft on en.wikipedia.org (38 more links to page 34 expand)
- Pagerange 56-60: Mary Wollstonecraft on de.wikipedia.org (9 more links to page 56-60 expand)
- Chapter 3: Feminist philosophy on en.wikipedia.org (12 more links to Chapter 3 expand)

By project
- En.Wikipedia.org (9 links expand)
- Wikidata.org (54 links expand)
Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children (C44583)

Author: String as published: “Dr AJ Wakefield, FRCS” (Q508568) First: Andrew Last: Wakefield Initial: J

URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(97)11096-0/fulltext


Author: String as published: “Dr AJ Wakefield, FRCS” (Q508568) First: Andrew Last: Wakefield Initial: J

Source type: Scientific article – Primary source

Publication date: 28 February 1998

Publication status: Retracted. Date: 6 February 2010

Citelinks: - zero current Wikimedia project citations - [Former links: 150 across 27 wikis expand]
Hypothetical implementation of inspecting a shared citation in source editor:

```
<cite>{{CID|1234567#123|style=1|lorem ipsum}}</cite>
```

Optional citation style preference

Locally stored wikitext

Quick access to Cite DB item

Optional subsection (e.g. page)

Work's Cite DB identifier

New tag to differentiate functionality

Citation template

Inspecting a citation in VisualEditor, and in Wikidata: Would look and feel the same
How we might do it

1. Scope
2. Resources
## Scope

<table>
<thead>
<tr>
<th>Citation database</th>
<th>Wikidata</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service product</strong></td>
<td><strong>Sister project</strong></td>
</tr>
<tr>
<td>Records about specific publications: Scholarly publications, URLs, editions of books, news articles, archival records</td>
<td>Items about concepts: Authors, publishers, newspapers, websites, topics....</td>
</tr>
<tr>
<td>Only that which is cited as a Wikimedia reference. Only created upon their being used, individually.</td>
<td>Any works. Created in advance, en masse, for any use-case.</td>
</tr>
</tbody>
</table>

**Examples:**
- https://doi.org/10.1002/elps.200700396

**Examples:**
- *Democracy in America* (Q784882)
- *Alexis de Tocqueville* (Q140694)
- *Philosophy* (Q5891)

Estimated size: ~50m records

Current size: ~90m items
# Roadmap

## Product

<table>
<thead>
<tr>
<th>Q3 Current fiscal</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Architectural exploration</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st 6 months</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Design research</td>
<td>18 months</td>
</tr>
<tr>
<td>● Engineering design and prototype release</td>
<td></td>
</tr>
<tr>
<td>● Monitoring + Editing backend research</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd 6 months</th>
<th>Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Launch database; Community building + properties</td>
<td>Integrated Product-Platform team</td>
</tr>
<tr>
<td>● Iteration on Monitoring + Editing integration</td>
<td>1 PM, 2 front-end engineers, 2 platform engineers, .5 Designer, .5 QA, .5 Analyst, 1 PgM, .5 CL, .5 Eng Mgr.</td>
</tr>
<tr>
<td>● First user actions in 1st round target wikis</td>
<td>Overlapping work with existing teams (inc. WMDE)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3rd 6 months</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Completed workflow support for 1st round wikis</td>
<td>The community motivation is strong with citation-focused volunteer projects in all Wikis, external orgs, and grant-supported initiatives.</td>
</tr>
<tr>
<td>● Enable on 2nd round of Wikis</td>
<td></td>
</tr>
<tr>
<td>● Populating with content + support community tools</td>
<td></td>
</tr>
</tbody>
</table>

---

**Metrics**

Wikis requesting it enabled; citation records created;
Proportion of shared citations on a wiki; no of users creating/using a shared citation.
“Metadata is a love note to the future”

{{wikicite}}
Acknowledgements

The following people provided input as part of stakeholder research for this proposal

WMF
Amir Aharoni, Chris Albon, Asaf Bartov, Runa Bhattacharjee, Adam Baso, Amanda Bittaker, Carly Bogen, Kate Chapman, Bryan Davis, Carol Dunn, James Forrester, Satdeep Gill, Danny Horn, Ramsey Isler, Jon Katz, Daniel Kinzler, Samuel Klein, Carolyn Li-Madeo, Erica Litrenta, Josh Minor, Felix Nartey, Margeigh Novotny, Guillaume Paumier, Maryana Pinchuk, Evan Prodromou, Ed Sanders, Joseph Seddon, Moriel Schottlender, Ben Vershbow, Marielle Volz, Caitlin Virtue, Denny Vrandečić, Sam Walton, Leila Zia

Other organisations
WMDE – Lydia Pintscher, Amir Sarabadani, Adam Shorland;
Internet Archive – Mark Graham, Jake Orlowitz; OpenCitations – Silvio Peroni;

Community
Jean-Frédéric Berthelot, Anne Clin, Noé Gasparini, James Heilman, Toby Hudson, Andrew Lih, Andy Mabbett, Luca Martinelli, Mahir Morshed, user:MisterSynergy, user:NikkiMaria, Mike Peel, João Peschanski, Siobhan Leachman, Lane Rasberry, Thomas Shaffee, Doug Taylor, Nicolas Vigneron, Andra Waagmeester

And the WikiCite Steering Committee: Phoebe Ayers, Daniel Mietchen, Merrilee Proffitt, Alex Stinson, Dario Taraborelli
References in Wiktionary*

*The differences between the way in which different wiktionaries is possibly even bigger than the differences between wikipedias...
## References in Wikidata

<table>
<thead>
<tr>
<th>Patronage</th>
<th>3,738,189</th>
<th>January 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>point in time</td>
<td>1 reference</td>
<td></td>
</tr>
<tr>
<td>3,543,007</td>
<td>point in time</td>
<td>October 2018</td>
</tr>
<tr>
<td>3,580,940</td>
<td>point in time</td>
<td>November 2018</td>
</tr>
<tr>
<td>3,860,239</td>
<td>point in time</td>
<td>December 2018</td>
</tr>
<tr>
<td>3,047,596</td>
<td>point in time</td>
<td>February 2018</td>
</tr>
<tr>
<td>3,340,873</td>
<td>point in time</td>
<td>March 2018</td>
</tr>
<tr>
<td>3,289,772</td>
<td>point in time</td>
<td>April 2018</td>
</tr>
<tr>
<td>3,304,739</td>
<td>point in time</td>
<td>May 2018</td>
</tr>
</tbody>
</table>

Each field of each reference is stored independently, even if it’s the same.

Currently planning: Turn All References Blue [TARB]

- “a complete inventory of all cited references, to be continually updated.”
- “...A challenge to building this inventory is that cited references can be expressed in many different ways.”

A community-curated Wikimedia Citation DB would allow IA to:

- Focus on archiving, not on our template structures
- Leverage one database
- Identify gaps in their collection, or what needs re-archiving
- Freely access a de-duplicated, up-to-date, curated and filterable dataset of different types of publications, forming the backbone of the TARB project.

The IA Bot dashboard, tracking the progress of extracting URLs from Wikipedias, and re-inserting archive-urls to the reference – a task undertaken on each wiki separately.
Knowledge equity

Real time visibility into our citation graph allow us to:

- Help marginalized languages, communities and subject specialists to curate their reliable sources for easy reuse.
- Reduce the citation management workload, especially for smaller communities.

**Key Opportunity:** How might we identify citation gaps or imbalances, while not supercharging inequity?
Knowledge as a service

Structured citations contribute to the semanticization of our content and is core to Knowledge as a Service to the ecosystem around Wikimedia.

**Key Problem:** How might reusers take advantage of a language-agnostic citation graph?
The product flywheel
Addressing community concerns

2014 English Wikipedia “delete” comments for Template:DOI

- “We, as a wiki, have chosen to sacrifice the benefits of consistency and elimination of redundancy in exchange for ease of editing and the idea that hitting "edit" on an article actually allows you to edit the article, not just the meta-structure of the article…”
- “Page watching is a major concern as well, as someone could change the cite doi child template, but that edit would go unnoticed and disconnected from the article itself…”
- “I hate the giant clots of wikisnot left in article markup by cite templates and migrating citation data to wikidata can’t come fast enough, but this was never a great solution…”

2017 English Wikipedia “delete” comments for Template: Cite Q

- “References are part of the page content, and editors should be able to view, edit and copy them by viewing the page source. Putting them on other pages made them harder to find and edit, harder to monitor for vandalism, harder to re-use by copy and paste….”
- “Maybe one day we will be able to edit Wikidata content within WP pages, and watch it for changes with other page changes. At that point it might make sense to host content such as references on Wikidata. But right now it makes the editing experience significantly worse and so should not be used…”
- {{cite doi}} was bad enough, but at least when an editor found the relevant page it was a recognisable citation, and could be examined, edited, copied as if it were in an article. Putting them on Wikidata is far worse, as it is a separate project and the way it works is completely unlike WP.
Malicious content

A service that centrally hosts content that is **visible to readers** of all sister-projects creates a new vector for malicious content contribution.

**Mitigation:**
- Consistent with [Risker's checklist for content-creation extensions](#), ensure the integration of:
  - checkuser & edit suppression/oversight [e.g. if name of a citation is changed to a libellous statement]
  - content logs, and user logs
  - #tags in edit history – indicating the username, the sisterproject originating the edit, and ‘via citations DB’
- Pan-Wikimedia project community-curation (via integration of monitoring tools) spreads the burden of patrolling

**Reduced risk:**
- Citation items are not ‘editorial’ content – there is minimised scope for edit warring
- Citation content and form are disaggregated – fields in the DB don’t need to be shown in the article
- Purely metadata, no mass-import, means very limited scope for copyright/database-rights violations
Spam

A service that centrally hosts content visible to reusers (e.g. search engines) creates a new vector for spam link promotion.

Mitigation:

- Centralising citations allows easier identification of poor quality links from common domains, especially useful to smaller language communities.
- Auto-delete Cite DB item when the reference was rapidly removed from its originating location
- Create Cite DB items only when used in mainspace (?)

Reduced risk:

- Being a service project, there is limited direct audience and therefore reduced spammer motivation
- Nofollow tags decreases SEO spam

English Wikipedia’s “Perennial Sources”. Only very few Wikipedias have an equivalent page. Pooling resources would help communities identify [un]reliable sources in other languages.
Why not within Wikidata?

**Wikidata is capable!** There is a vibrant community, extensive data modelling, and corpus of existing citation content. For example:

**Scale:** A separate database means that all Wikimedia citations can “Fit”. E.g. the 10s of millions of citations to specific URLs.

**Service:** By serving all Wikimedia projects, Wikidata itself can equally benefit from the shared citations, not merely host them.

**Scope #1:** Restricting to only Wikimedia citations ensures the ontology remains practical.

**Scope #2:** Restricting to only creation-upon-need ensures the community understands the difference from WD.

**Sovereignty**
WD and Cite DB would be editorially independent. **Some content overlap**, but would be overwhelmingly limited to individual scholarly journal articles, some book editions. WD does not normally have items about URLs, newspaper articles etc. Content policies would evolve.

“Let’s do it.” – Lydia Pintscher; “Sounds great.” - Adam Shorland; “It looks awesome to me” - Amir Sarabadani
Launched in 2004 – three years after Wikipedia – **Wikimedia Commons**’ role is to reduce duplication of effort across the Wikimedia projects.

Prior to this, a multimedia file had to be **locally hosted** in order to be able to be used.

- Many exactly matching files were uploaded to each Wikipedia language
- Metadata and captions were duplicated, but inconsistently and monolingually
- Curation of multimedia was handled by each community separately

The “service project” Commons.wikimedia.org [note: not WikiCommons.org, which implies a “sister project” status] centralised that work. The community later expanded its scope to include **any** freely-licensed multimedia.

**Citations DB** serves an equivalent role and, similar to continued “local upload” of multimedia, traditional citations would still operate in parallel. By contrast, the existence of **Wikidata**’s already fulfils the expanded scope use-case.

**Commons** still lacks the ability to notify Wikipedia when images-in-use are changed. That would become possible with the “monitoring” features in this proposal.
Rollout

1. WikiSources
   - All content is a citation destination
   - Most active community entirely among “emerging communities” ensuring diversity built-in to the culture from the beginning
   - Built on MediaWiki, the primary use-case

2. Wikidata
   - Large number of citations already in a consistent structured format
   - Multilingual and multicultural community, familiar with the format
   - Necessary community discussion on how properties and content should be mapped to each other
   - Built on WikiBase, requiring separate UI work

3. Wiktionaries
   - Community familiar with citation management
   - Diversity of citation types, template formats, and languages

4. Wikipedias and other sister projects
   - Largest block of content - most benefit from the ‘network effect’
   - Largest variety of citation formats and edge-cases