Resource Discovery in the Age of Wikipedia

Jake Orlowitz

wikipedialibrary.org
@WikiLibrary
Imagine a world in which every person on the planet shares in the sum of all human knowledge.
Wikipedia’s volunteers

20 million registered

80,000 active users

1,400 admins
“It was like a giant community leaf-raking project in which everyone was called a groundskeeper. Some brought very fancy professional metal rakes, or even back-mounted leaf-blowing systems, and some were just kids thrashing away with the sides of their feet or stuffing handfuls in the pockets of their sweatshirts, but all the leaves they brought to the pile were appreciated. And the pile grew and everyone jumped up and down in it having a wonderful time. And it grew some more, and it became the biggest leaf pile anyone had ever seen anywhere, a world wonder.”

- NY Review of Books
Wikipedia’s scale

- 35 million articles
- 2 billion edits
- 288 languages
- 8000 views per second
- 500 million monthly visitors
- 7th most popular website
- 2000x larger than Britannica
- 8th highest doi referrer
Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

So far Altmetric has seen 8 citations on Wikipedia.

Nature (journal)
Cited by user DOI bot on en.wikipedia.org
Nature is a prominent interdisciplinary scientific journal. It was first published on 4 November 1869.

Evolution
Cited by user Thompsma on en.wikipedia.org
Evolution, also known as descent with modification, is the change in heritable phenotype traits of biological populations over successive generations.
Subdomain: en.m.wikipedia.org
Main domain: wikipedia.org.
Events from 1 October 2010 to 1 January 2015

<table>
<thead>
<tr>
<th>type</th>
<th>date / count</th>
<th>from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total referrals count from subdomain</td>
<td>190455</td>
<td>CrossRef Resolution Logs</td>
</tr>
</tbody>
</table>

Daily referral count from domain

![Graph showing daily referral count from domain from Oct 2010 to Jan 2015](image.png)
Wikipedia’s ubiquity

• 50% to 90% of physicians use Wikipedia

• 94% of medical students use Wikipedia
Wikipedia’s why

Fast, easy, and free to access
Understandable
Comprehensive
Generally accurate
Wikipedia’s pillars

Neutral point of view
Verifiability
Consensus
Civility
Open copyright
Virtual **filter**

1. **Edit Filter** automatically rejects known vandalism patterns
2. ClueBot reverts and flags suspicious edits with a **machine-learning bot**
3. Humans review malicious changes tagged with **language recognition** tools
4. **Vandalism patterns** are checked against metadata and historical trends
5. **Recent changes** patrollers scroll through new edits
6. Editors alerted to each change on all pages in their **article watchlist**
7. **Specialists and experts** report and fix mistakes when they see them
8. **Millions of readers** identify and correct errors when they come upon them
9. Link **blacklists** lock out known spam sites and unreliable sources
10. Detection mechanisms to determine **conflict of interest**
11. **Administrators** to block disruptive editors and protect pages
Wikipedia literacy

- Engage with global audience
- Identify bias and evaluate credibility
- Review literature with proper citation
- Avoid plagiarism and respect copyright
- Construct knowledge to fill content gaps
- Discourse in a community of practice
- Think critically and reflect on process
- Digital citizen in an online collaboration

All rated articles by quality and importance

<table>
<thead>
<tr>
<th>Quality</th>
<th>Top</th>
<th>High</th>
<th>Mid</th>
<th>Low</th>
<th>???</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>1,063</td>
<td>1,677</td>
<td>1,552</td>
<td>915</td>
<td>168</td>
<td>5,375</td>
</tr>
<tr>
<td>FL</td>
<td>137</td>
<td>520</td>
<td>618</td>
<td>564</td>
<td>124</td>
<td>1,963</td>
</tr>
<tr>
<td>A</td>
<td>189</td>
<td>370</td>
<td>538</td>
<td>303</td>
<td>68</td>
<td>1,468</td>
</tr>
<tr>
<td>GA</td>
<td>1,772</td>
<td>4,146</td>
<td>8,122</td>
<td>8,069</td>
<td>1,588</td>
<td>23,697</td>
</tr>
<tr>
<td>B</td>
<td>10,975</td>
<td>21,076</td>
<td>31,854</td>
<td>24,081</td>
<td>13,023</td>
<td>101,009</td>
</tr>
<tr>
<td>C</td>
<td>8,690</td>
<td>25,060</td>
<td>55,680</td>
<td>70,643</td>
<td>37,896</td>
<td>197,969</td>
</tr>
<tr>
<td>Start</td>
<td>15,605</td>
<td>67,680</td>
<td>275,152</td>
<td>652,713</td>
<td>261,189</td>
<td>1,272,339</td>
</tr>
<tr>
<td>Stub</td>
<td>3,907</td>
<td>27,584</td>
<td>202,675</td>
<td>1,592,673</td>
<td>840,977</td>
<td>2,667,816</td>
</tr>
<tr>
<td>List</td>
<td>2,575</td>
<td>9,532</td>
<td>27,810</td>
<td>75,026</td>
<td>58,104</td>
<td>173,047</td>
</tr>
<tr>
<td>Assessed</td>
<td>44,913</td>
<td>157,645</td>
<td>604,001</td>
<td>2,424,987</td>
<td>1,213,137</td>
<td>4,444,683</td>
</tr>
<tr>
<td>Unassessed</td>
<td>106</td>
<td>322</td>
<td>1,614</td>
<td>18,477</td>
<td>473,235</td>
<td>493,754</td>
</tr>
<tr>
<td>Total</td>
<td>45,019</td>
<td>157,967</td>
<td>605,615</td>
<td>2,443,464</td>
<td>1,688,372</td>
<td>4,938,437</td>
</tr>
</tbody>
</table>
The Research Connection

Wikipedia is as good as its sources
Libraries have the best sources
Wikipedia has the most eyeballs
Connect a circle of research, dissemination
Wikipedia as starting point for deeper learning
The Wikipedia Library

Gain access to **paywalled** sources

Facilitate **research** for editors

Connect to **libraries**

Lead to free and **local** sources

Promote **open** access
Access partnerships

Closed-access, paywalled sources
Donations for mutual benefit
Vetted, highly-active subject editors
Requirements and TOU respecting copyright
High impact metrics
OA signalling

Thinking big

What if every reference in a Wikipedia article had a robust citation including access and reuse status?
Wikipedia Specialists

Wikipedian-In-Residence  [[WP:WIR]]
Capacity focused, On-Staff, Wikipedian

Wikipedia Visiting Scholar  [[WP:WVS]]
Remote, Research Affiliate, Wikipedian

Wikipedia Library Interns  [[WP:TWL/I]]
Unfamiliar Staff, Self-Guided Curriculum
Thinking big

What if every library or research institution had a Wikipedia specialist to surface their collections and specialist knowledge on Wikipedia?
Wikipedia-in-the-Workflow

Use Wikipedia to solve public knowledge problems, integrated into your process

- National Archive
  Policy encourages Wikipedia editing

- State Library of New South Wales
  [[Wikipedia:GLAM/State Library of New South Wales]]
  [[List of Australian diarists of World War I]]
Linking archives

Wikipedia articles with citations to relevant archival items, specialist collections, and finding aids

==History==

==Further reading==

==External Links==
Case study

BALL STATE University

- 40 assets viewed 13,000 times, + 600%
- 10,000 pageviews referred via Wikipedia
- 5 times greater than the total views from any source
- Views for the whole collection tripled

“an overwhelming success... remarkable in its efficiency”

http://outreach.wikimedia.org/wiki/GLAM/Case_studies
REGISTER: Create an account for you individually – not your organization
DISCLOSE: Mention and explain your institutional affiliation
HIGHLIGHT: Expose your most relevant, substantial collections neutrally
INCLUDE: Add relevant citations to other archives’ materials and scholarship
ENGAGE: Respond thoughtfully and clearly to community concerns

“Curators, librarians, archivists, and similar are encouraged to help improve Wikipedia, or share their information in the form of links to their resources.”
What if every library or cultural institution exposed the best research in their field on Wikipedia?
Image uploads

GLAM-Wiki Toolset
Bulk upload open licensed media with robust metadata to Wikimedia Commons

Collection visibility
State Library of New South Wales:
➢ 93 images viewed by 480,869
New York Public Library images:
➢ 2114 images viewed by 6,881,713
NARA, British Lib, Smithsonian, Europeana
Thinking big

What if every reusable collection was shared on Wikimedia Commons?
Public reference desk
Thinking big

What if Wikipedia became the public’s digital reference desk staffed by professional reference librarians?
Building discovery

Expose collections

Remix metadata
George Washington

A good article from Wikipedia, the free encyclopedia. A former featured article candidate.


This article is about the first President of the United States. For other uses, see George Washington (disambiguation).
Named Entity Recognition

These names have been extracted from this entity's finding aid or biography. Select names that you would like to look up in VIAF.

In the next step, you will be able to make a final selection to create <cpfRelation> elements, with associated VIAF IDs, in the EAC-CPF record.

Each name can be edited to improve the search query, if appropriate. When editing, it is best to put names in inverted order (Last Name, First Name).

If names need to be split, or if you have additional names to add, you may click "Add New Row" to input appropriate data.

Note that geographical places are not included in VIAF and so should be ignored at present.

Please note that if you select several names to look up in VIAF, your query may take some time to run.

Please choose names to create <cpfRelation> elements:

- Abreu, Juan
- African Folktales
- African Tribal Groupings
- Afro Caribbean
- Afro Cuban
- Afro Cubans
- Afro-American
- Afro-American Literature, Homage
- Afro-Cuban
- Afrocuban Art
- Agbelemoge, Terry
- Alejandro Exter
- Alexandra Exter
What if Wikipedia helped readers research?
What if every Library Network shared best practices for using Wikipedia as a tool for Research, Literacy and Public Knowledge?
Wikipedia, Libraries = **natural allies**

Wikipedia is the **starting point** for research.

We lead readers **back to sources** at libraries so they can **critically engage** with knowledge.
Questions?
WikipediaLibrary.org

Jake Orlowitz
@JakeOrlowitz
jorlowitz@wikimedia.org

Tweet Us!
@WikiLibrary