

An analysis of Wikipedia references across PLOS publications

Jennifer Lin¹ & Martin Fenner²

¹jlين@plos.org; ²mfenner@plos.org

^{1,2} Public Library of Science (PLOS)

Introduction

The free online encyclopedia Wikipedia is currently the fifth largest website with 18 billion page views and nearly 500 million unique visitors a month (Cohen, 2014). Scholars are increasingly challenging the primary mechanism for scholarly communication, namely publication of scholarly journal articles. And some have opened their eyes to the potential of the Wikipedia platform to address the commonly identified drawbacks of the publishing system such as long publication delays, an inflexible & static format, peer review biases, etc. (Black, 2008). Subsequently, the connections between Wikipedia and scholarly research are growing stronger: journal editors are enriching research publications by cross-linking to Wiki pages (Penev et al., 2008), editors are actively soliciting and coordinating Wikipedia contributions alongside journal article publications (Poulter, 2014), and the first Wikipedian-in-residence at an academic institution is working on expanding engagement in “public scholarship” (Brown, 2014). Furthermore, Wikipedia offers researchers dynamic content creation and management tools that can enable closer collaboration during the research process. These factors have all led to an increasing interest from academics to contribute scholarly research on Wikipedia.

The growing significance of Wikipedia on scholarly research is found not only in deepening engagement by researchers but also elevating visibility and discoverability of research articles. Wikipedia provides a massive amount of traffic to formal scholarly research. CrossRef, the citation-linking network for scholarly publishers, calculated that Wikipedia is the 8th largest referrer to the CrossRef DOI resolver service of 65 million journal articles in their index (Bilder, 2014). This figure reveals not only how often research articles are referenced in Wikipedia pages, but more significantly, the extent to which Wikipedia readers access the journal article itself from a Wikipedia page. All of this has significant implications for our understanding of scholarly communication. At the same time, there is still insufficient data on how scholarly content is referenced in the Wikipedia altmetric. The Open Access publisher Public Library of Science (PLOS) is collecting this information for its entire corpus, which makes possible a detailed analysis of the reuse of PLOS content in Wikipedia.

The research questions are as follows:

1. To what extent are scholarly articles referenced in Wikipedia, and what content is particularly likely to be mentioned?
2. How do these Wikipedia references correlate with other article-level metrics such as downloads, social media mentions, and citations?

Data and Methodology

The data generated for this analysis comes from the PLOS instance PLOS ALM (<http://alm.plos.org>) of the open source ALM application (<https://github.com/articlemetrics/alm>). The application harvests data from a number of external sources to capture the engagement surrounding research articles after publication, including usage statistics, citations in scholarly literature, and a host of altmetrics including social bookmarking, sharing on social media outlets, and mentions in blogs and news media. For Wikipedia data, the ALM application collects the number of Wikipedia articles that reference PLOS articles in the 25 largest Wikipedia languages by number of articles (“List of Wikipedias,” 2014). This is done via a full-text search using the article DOI, which is part of the PLOS journal page URL. The Wikipedia user and file namespaces are not searched. The data for this analysis were obtained from the monthly ALM report in CSV format, generated March 10, 2014 (“Cumulative ALM Report,” 2014). The R statistical analysis software version 3.0.2 was used for analysis.

Results

Out of the 110,129 PLOS articles published before March 10, 2014, 4,553 articles (4.13%) were mentioned in Wikipedia at least once (“Wikipedia ALM Report,” 2014). All data were collected on March 10, 2014 and reflect the counts on the date of access. While the Wikipedia reference rate is similar to mentions in science blogs or the post-publication peer review service, F1000Prime, the nature of each activity is quite broad and the users behind it also vary. Fifty-one percent of articles mentioned in Wikipedia were also mentioned in Facebook.

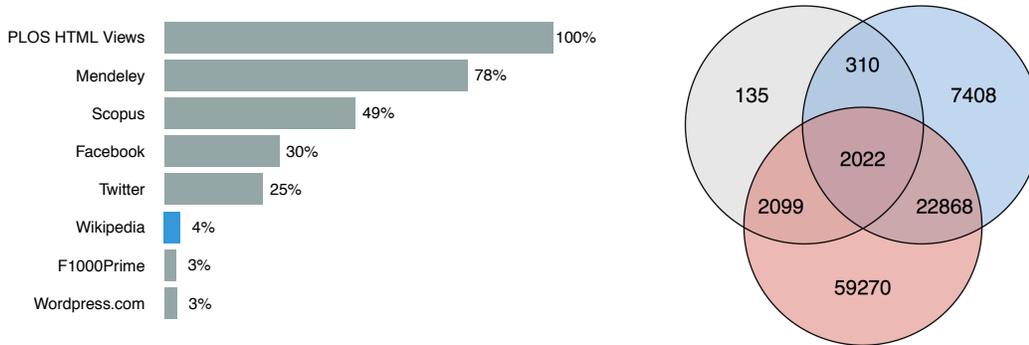


Figure 1. A) Percentage of all PLOS articles referenced in selected ALM data sources. B) Overlap between references in Wikipedia (gray), Facebook (blue), and Mendeley (red) for all PLOS articles.

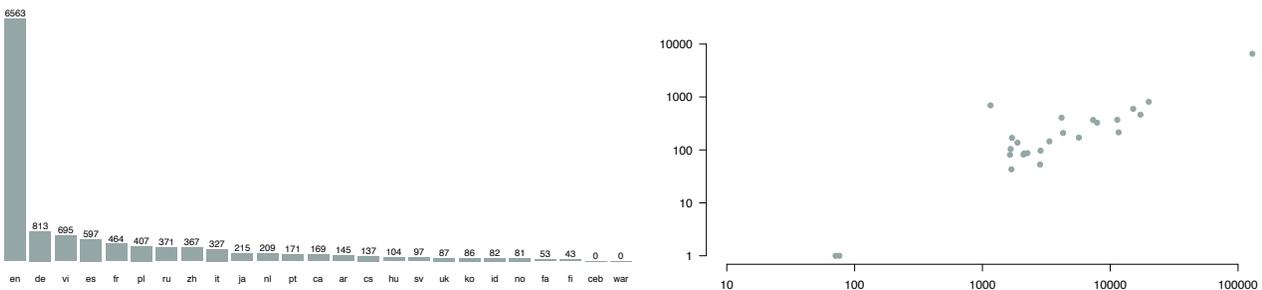


Figure 2. A) Number of Wikipedia pages referencing PLOS articles. Data collected from the 25 largest wikis. B) Correlation between Wikipedia active users (x) and number of Wikipedia pages referencing PLOS articles (y). Correlation = .098

By far the most referenced PLOS article is a study on the evolution of deep-sea gastropods (Welch, 2010) with 1249 references, including 541 in the Vietnamese Wikipedia. The 10 most referenced PLOS articles published in 2014 are listed in Table 1.

DOI	Title	References
10.1371/journal.pone.0008776	The “Island Rule” and Deep-Sea Gastropods: Re-Examining the Evidence	1248
10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288	The Tree of Life and a New Classification of Bony Fishes	145
10.1371/journal.pone.0012292	New Horned Dinosaurs from Utah Provide Evidence for Intracontinental Dinosaur Endemism	64
10.1371/journal.pone.0023852	Identification of Novel Functional Inhibitors of Acid Sphingomyelinase	60
10.1371/journal.pone.0014075	New Basal Iguanodonts from the Cedar Mountain Formation of Utah and the Evolution of Thumb-Spiked Dinosaurs	60
10.1371/journal.pone.0079420	Tyrant Dinosaur Evolution Tracks the Rise and Fall of Late Cretaceous Oceans	54
10.1371/journal.pone.0026964	A New Basal Sauropodomorph (Dinosauria: Saurischia) from Quebrada del Barro Formation (Marayes-El Carrizal Basin), Northwestern Argentina	52
10.1371/journal.pone.0029797	Ecological Guild Evolution and the Discovery of the World's Smallest Vertebrate	50
10.1371/journal.pone.0006190	New Mid-Cretaceous (Latest Albian) Dinosaurs from Winton, Queensland, Australia	50
10.1371/journal.pone.0002098	Multigene Phylogeny of Choanozoa and the Origin of Animals	48

Table 1. Most popular PLOS articles referenced in Wikipedia

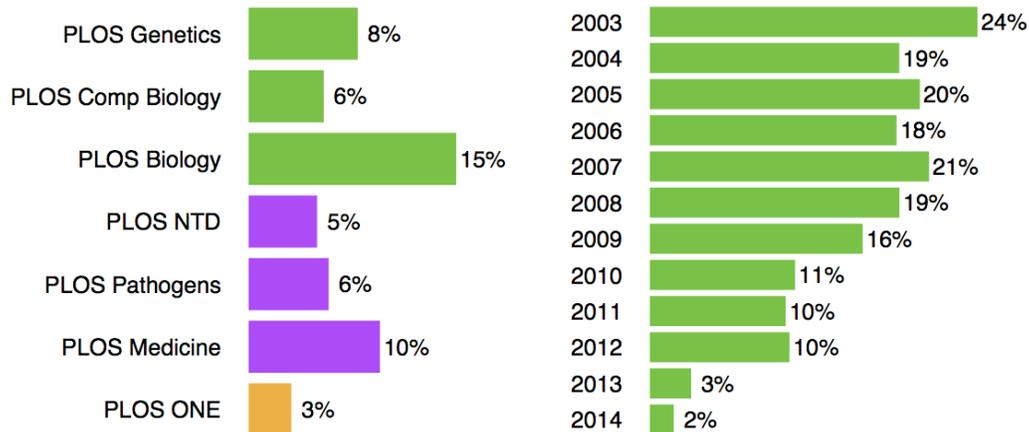


Figure 3. Proportion of papers with Wikipedia references by Journal. B) Proportion of papers with Wikipedia references by Year for PLOS Biology.

Conclusion and Discussions

The preliminary analysis uncovered evidence that suggests Wikipedia is an incredibly complex source, which needs more research attention, especially compared to other sources. The following dimensions impact Wikipedia behavior: dynamics of the editing process (i.e., adding and removing Wikipedia content by different users), politics of the proliferation of Wikipedia content across language pages, temporality of Wikipedia engagement, breadth of community engagement across public and private (scholarly) communities.

The Wikipedia references display a pattern distinct from popular social networks such as Facebook. While the references cover a broad set of topics, they particularly focus on articles from ecology, evolution and other subject areas that can enrich the encyclopedia with scholarly references. Forty-seven percent of references are found outside the English Wikipedia pages. The number of Wikipedia pages referencing a PLOS article highly correlates with the number of active users associated with that Wikipedia ($r^2=0.98$). For further analysis, we are interested in investigating the correlation between Wikipedia and citations as well as dig deeper into the subject areas covered (and hence, communities of practice represented). Finally, the research scope needs to be expanded across publishers so as to develop a more robust portrait of Wikipedia activity for scholarly literature.

References

- Black, E. (2008) "Wikipedia and academic peer review: Wikipedia as a recognised medium for scholarly publication?" *Online Information Review*, Vol. 32 Iss: 1, pp.73 - 88.
- Cohen, N. Wikipedia vs. the Small Screen. *The New York Times*.
<http://www.nytimes.com/2014/02/10/technology/wikipedia-vs-the-small-screen.html>
- Penev LI, Hagedorn G, Mietchen D, Georgiev T, Stoev P, Sautter G, Agosti D, Plank A, Balke M, Hendrich L, Erwin T. "Interlinking journal and wiki publications through joint citation: Working examples from ZooKeys and Plazi on Species-ID." *Zookeys*. 2011 Apr 14;(90):1-12. doi: 10.3897/zookeys.90.1369.
- Poulter, M. (2014, March 28). Publishing scholarly papers with, and on, Wikipedia. *Wikimedia UK Blog*.
<https://blog.wikimedia.org.uk/2014/03/publishing-scholarly-papers-with-and-on-wikipedia/>.
- Brown, K. (2014, March 14). Free plagiarism checker. *SF Gate*. <http://www.sfgate.com/technology/article/UC-Berkeley-grad-to-expand-Wikipedia-s-scholarly-5316009.php>.
- Bilder, G. (2014, February 24). Many Metrics. Such Data. Wow. *CrossTech Blog*.
<http://crosstech.crossref.org/2014/02/many-metrics-such-data-wow.html>.

List of Wikipedias. (2014). Retrieved April 18, 2014, from https://meta.wikimedia.org/wiki/List_of_Wikipedias#All_Wikipedias_ordered_by_number_of_articles

Cumulative ALM Report through 3/20/2014. http://article-level-metrics.plos.org/files/2012/10/alm_report_2014-03-10.csv

Wikipedia ALM Report through 3/20/2014. https://github.com/PLOS/altmetrics14-wikipedia/blob/master/data/alm_wikipedia_2014-03-10.csv

Welch, J.J., 2010. The “Island Rule” and Deep-Sea Gastropods: Re-Examining the Evidence S. Joly, ed. *PLoS ONE*, 5(1), p.e8776. doi: 10.1371/journal.pone.0008776