Wikipedia-integrated publishing: a comparison of successful models

Thomas Shafee[1,2,3]

   [1] *La Trobe Institute for Molecular Science, La Trobe University, Melbourne*

   [2] *Editorial board, WikiJournal of Medicine*

   [3] *Editorial board, PLOS Genetics*

T.Shafee@latrobe.edu.au

Presented at the 14th Health Libraries Inc. Conference – 20th October 2017, State Library of Victoria

Abstract

Wikipedia is one of the most accessed sources of medical information by the general public and is also widely used by medical professionals and students. Its medical content consequently has a large impact on health literacy. However, relatively few experts contribute to creating and improving its content, and this is partly due to limited formal reward for contributing. One key method of encouraging greater involvement therefore involves rewarding medical and research professionals with citable academic publications that are dual-published in Wikipedia. Such publications provide an ideal outreach platform for the author, increasing exposure for the journal, and providing high-quality content to the encyclopedia.

Several successful dual-publication models exist. In the first model, peer-reviewed material is published in a journal and subsequently copied to Wikipedia under a compatible licence (typically Creative Commons). This produces new, high-quality articles and is easily consistent with current open access journal practices. A second, less common format is where material is first published in Wikipedia, then subjected to academic peer review before being published as a journal article. This model is also compatible with the recent practice of improving and peer-reviewing existing Wikipedia pages. A third model is where a journal requires authors to update Wikipedia as part of the journal's publication process. This allows content to be pitched at different levels for the journal and Wikipedia. Developing and extending these models is a vital component of ensuring that accurate biomedical information is available on a platform that is read by millions each day.

Introduction

Not long ago, an encyclopedia was a luxury item. Now, the largest encyclopedia the world has ever known is freely available. Wikipedia has grown rapidly since its beginnings in 2001, and now encompasses over 5.5 million articles in English, with a further 40 million across 299 languages (1). It is widely used as a source of medical information by the general public, as well as medical students, doctors, and scientists (2–5). Top medical articles get millions of views per year (Fig. 1). The accuracy of these articles consequently influences public health literacy and medical education (6,7).
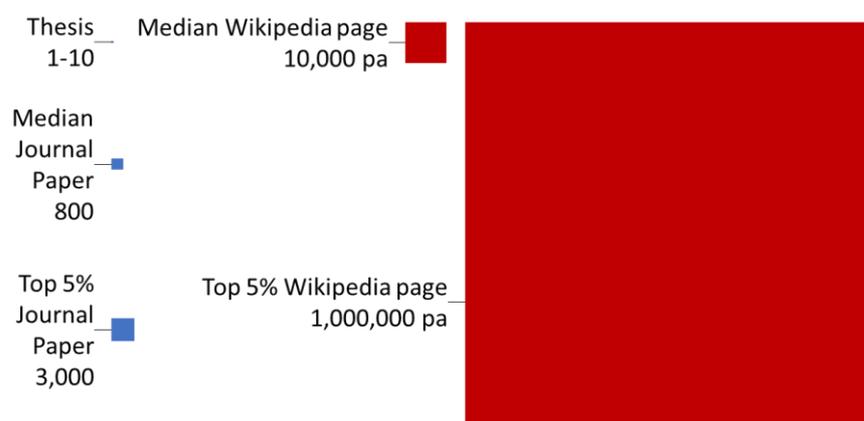
Figure 1 | Readership size comparison. Total journal article readership numbers in blue, annual Wikipedia article readership numbers in red.

Wikipedia is built entirely by collaborative volunteer labour, so there has long been caution over its accuracy and credibility. Indeed, in its early years it was not clear whether constructive edits would outweigh erroneous editing and deliberate vandalism. However, in 2005 the journal *Nature* famously found the error rate between Wikipedia and Encyclopedia Britannica to be comparable (8), and subsequent studies have explored in more detail accuracy within specific topic areas. Although high-importance articles are generally relatively well written, quality varies widely between articles more broadly (9). Therefore, there is a constant need to update and improve information, address omissions, add illustrations, and increase readability (9).

Over time, the encyclopedia is necessarily maturing from a site of article creation only towards one of article improvement. Wikipedia articles can be internally peer-reviewed by one editor to be awarded 'Good Article' status, or by 5-10 editors to achieve 'Featured Article' status (10). This peer review process is thorough, but is limited by the expertise of Wikipedia's editor community. Improving coverage of complex topics therefore benefits from more expert input, so there is now increasing interest in promoting academic engagement with Wikipedia (11,12).

Academics' opinions of the encyclopedia seem to be improving as its content matures (13–15). The goal of attracting busy academics and professionals to contribute to Wikipedia is supported by offering tangible career rewards (16,17). For example, there are moves to recognise editing Wikipedia and citation by Wikipedia through the research impact measure, Altmetrics (16,18).
One effective tool for bridging the Wikipedia-academia divide is dual-publication in academic journals. Dual-publication involves the publication of an article in both the scholarly literature as well as in Wikipedia. Here, I describe how this can be achieved, highlight some successful examples, and explain how they can benefit authors, journals, and the encyclopedia.

Dual-publishing models

There are several main models of dual-publishing. In all cases, the journal organises formal peer review and publishes a permanent version of record, whilst Wikipedia hosts the editable version. In "journal first" dual-publishing, material is first published in the journal with Wikipedia a secondary publishing outlet. "Wikipedia first" dual-publishing includes a prior step where Wikipedia is treated as a preprint server, where encyclopedia articles of sufficient quality can be subjected to academic peer review. In "parallel" dual-publishing, different content on the same topic is tailored to the journal and Wikipedia. The journal-published version of record stands as a stable, citable, indexed document. The Wikipedia-

published version acts as a living document that is constantly updated with new information and refined over time.

Generally, a good encyclopedia entry constitutes a broad academic review paper (diligent referencing, neutral point of view, avoiding copyright violation, etc). However, there are typically minor formatting and stylistic differences between the journal version and the Wikipedia version (Table 1). The journal version typically includes figure and table numbering as standard practice, whereas numbering is not included in Wikipedia, because they can be added, reordered, or replaced over time. The encyclopedia also preferences secondary and tertiary sources for medical statements (ideally clinical guidelines, meta-analyses, and reviews). Wikipedia also prefers dates over phrases that can become obsolete, such as "currently" or "the latest". Short section titles are preferred and "review-isms" are best avoided, such as "As Jones *et al* showed in their 2001 *Nature* paper…". Finally, it is also stricter in avoiding original research, and only summarising published information. Nevertheless, authors report finding it easy to adapt to the differences (17).

Table 1 | General features of the two different publication formats.

	Academic Journal	Wikipedia
Readership size	Small and brief (21) Median article - 800 total Top 5% article - 3000 total	Very large and extended Median article - 10,000 per year (22) Top 5% article - 1,000,000 per year (23)(24)
Readership composition	Other academics, often within narrow field	General public as well as experts and professionals
Peer review	Pre-publication, private review by 2-4 subject specialists	Post-publication, public review of a sort by subject generalists 'Good article' - 1 reviewer 'Featured Article' - 5-12 reviewers
Reputation	Varies by journal but generally extremely high	Public generally trust Academics have mixed opinions by improving
Authorship	Small number with relevant, accredited expertise. Organised group with lead and corresponding authors.	Large number with mixed expertise levels. Loose organisation. Many pseudonymous or anonymous.
Timeliness	Static Updated by new publications	Constantly updated Only one consensus version

Once published, there are typically a higher number of page views of the Wikipedia version than the journal version. Wikipedia pageviews are also relatively stable, compared to the shorter half-life for page views of journal articles. Wikipedia articles that have been dual-published are indicated by an icon at the top of the page (Fig. 2A), as well as a detailed reference to the published version in the references section (Fig. 2B). Since standard Wikipedia articles should not be cited (19,20), this information provides readers of these articles with the citable journal version of record.
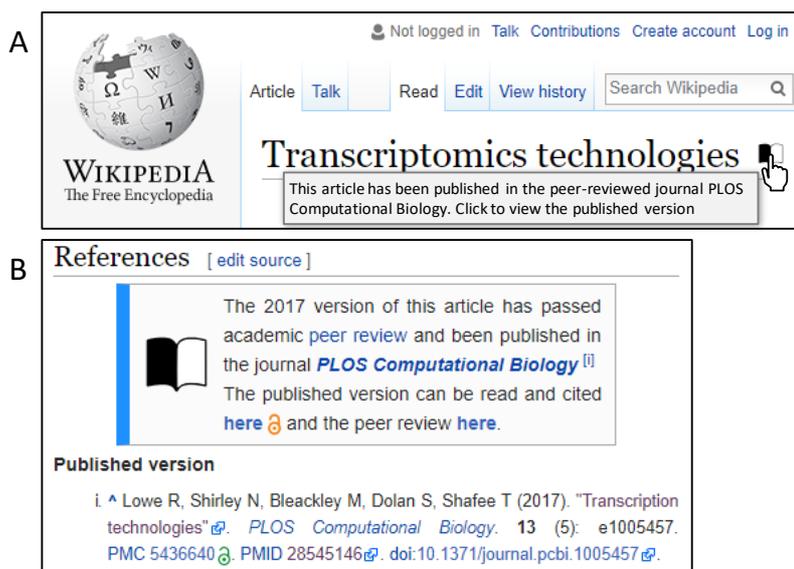
Figure 2 | Wikipedia article elements indicating. A) A journal icon is present in the top right of the article title. Upon mouse-over, a notification appears. Clicking the icon redirects to the journal article. B) A more in-depth description of the dual-publication is present at the top of the article's references section, along with the full citation and relevant links.

Journal first model

Journal first dual-publishing involves authors writing an article for a peer reviewed journal which is subsequently used for the creation of a new Wikipedia page (Fig. 3A). The format is relatively familiar, but with a slightly broader scope than usual. Authors are given the journal's guidelines, as well as a simplified set of Wikipedia guidelines for the article to adhere to. The article is then submitted for academic peer review by the journal (often public review). Articles approved by peer review are published in the journal with an open access licence that is compatible with copying the content over to Wikipedia, typically the Creative Commons licence CC BY (25). Once the content is published online in the journal, it is duplicated on Wikipedia as a secondary outlet.

There is no particularly direct comparison in contemporary publishing for using Wikipedia to reach a broader audience. The closest may be the increasing practice of researchers disseminating post-prints and open access publications via self-archiving and eLibrary databases (26).
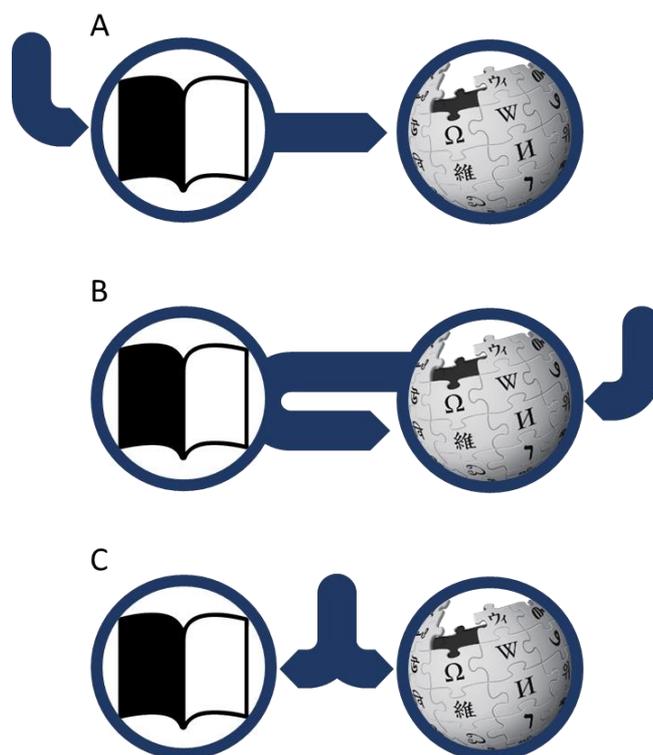
Figure 3 | Dual-publishing information flow. In each case, the journal (left) performs expert peer review and publishes a stable version of record, and Wikipedia (right) presents the information to a larger audience. A) *Journal first* model, where information content is first published in the journal article, and then published in Wikipedia as a secondary outlet. B) *Wikipedia first* model, where information content is first published in Wikipedia, then used as a pre-print for the journal article, then copied back to Wikipedia. C) *Parallel publication* where the same author creates different content on the same topic for both the journal and for Wikipedia.

Examples

The most prominent examples of the journal first model are the *PLOS Computational Biology* and *PLOS Genetics* (27,28) journals. These journals established a 'Topic Page' format in 2012 for broad, encyclopedic articles that would be published first in the journal, and then copied into Wikipedia. Published topics encompass broad principles including circular permutation, as well as techniques like transcriptomic technologies (29,30). The *WikiJournal of Medicine* was specifically set up with Wikipedia-integration in mind (31,32) and publishes several related formats including full-length encyclopedic articles, image galleries, and brief articles on new images and diagrams (33,34).

Benefits and limitations

The journal first model best fits with current open access publishing practices. Its main requirement is some expertise in formatting text for Wikipedia, though this has been made simpler since the introduction of the VisualEditor interface, which makes editing Wikipedia similar to a Word document. *PLOS* achieves this by hosting a dedicated server for drafting wiki pages, topicpageswiki.plos.org. Of course, many topics have existing but under-developed Wikipedia pages. If a topic is covered by a 'Stub' or 'Start' class Wikipedia page, that page is completely overwritten by the new publication. It is always possible the page may revert to the earlier version, however this situation is unlikely and hasn't arisen so far. In the event of a dispute, the issue can be discussed on the article's 'talk page', and a

mediation committee can be asked to arbitrate if necessary. This leaves a grey area for topics where an in-depth Wikipedia page is present but out of date, since the journal version of the article would likely have to be integrated into the Wikipedia page in a more laborious fashion than most academic authors are likely to attempt. The vast majority of Wikipedia pages are under 500 words, so there is plenty of scope for this model (8).

Wikipedia first model

Wikipedia first dual-publishing involves taking an already existing Wikipedia page and subjecting it to external academic peer review (Fig. 3B). Such pages may be recent, or have existed since Wikipedia's inception, even having previously passed internal Featured Article review. External scholarly review allows its contents to be more stringently scrutinised by topic experts and are clearly marked as of higher quality. Wikipedia is therefore treated as a preprint server by the journal. Due to Wikipedia's Creative Commons by attribution share-alike (CC BY-SA) copyright licence, the peer reviewed journal version must also be published as CC BY-SA.

Examples

The first example of this model was the article "Dengue Fever" in 2014, published by *Open Medicine* (35,36). This was followed by the *WikiJournal of Medicine* with articles covering topics from anatomy to molecular biology (37,38). Overall, this model is the less common of the two.

Benefits and limitations

This format is far more flexible than the journal first model. It allows high quality Wikipedia articles ("Good" or "Featured") to be directly submitted for external peer review. It also allows medium-quality articles to be updated, expanded, and then submitted. This incentivises direct editing of Wikipedia even before the peer review process begins.

Although it has not yet been tested, it is worth noting that an article could go through the process of external peer review multiple times if the Wikipedia version later changes sufficiently to warrant another round of peer review and publication. It is conceivable that these updated journal articles would be re-published every few years, comparable to textbook editions.

In this format, the existing Wikipedia article is treated as a pre-print. Some academic journals are still wary of public pre-prints, since there is a strong drive for novelty and exclusivity. However, pre-print servers such as *ArXiv* and *PeerJ* pre-prints have advanced this considerably. A greater restriction is the requirement to publish under a CC BY-SA licence. This licence is uncommon in academic publishing, which typically uses either the more permissive CC BY, or the more restrictive, non-commercial CC BY-NC.

Although many Wikipedia articles have a small number of key contributors, often hundreds of individuals will make some contribution. These authors may also be pseudonymous or completely anonymous. This definitely conflicts with common practice in academia of clearly attributed and accountable authorship. Journals have adapted to this by listing the names of authors who submitted the article and responded to peer reviewer comments. The full list of all contributors is then included either in the acknowledgments, or as a hyperlinked '*et al*' in the author list. Author contributions are fully listed in the article's history (39) and graphically summarised as with the ArticleInfo tool (40). Exact text contributions can be tracked in more detail with the whoCOLOUR tool (41,42). Nevertheless, this may represent one of the greatest compatibility issues for journals looking to implement the model.

Parallel model

Parallel dual-publishing involves authors producing separate content for the journal and Wikipedia rather than requiring the content to be identical in both outlets (Fig. 3C). Typically, the content to be added to Wikipedia is peer reviewed at the same time as the journal article. The Wikipedia version can therefore act as a lay summary to accompany a more technical journal article.

Examples

The largest example of this format is by the journal *Gene*, which introduced the Gene Wiki Review format in 2013 (17). When authors write a review of a gene family for the Wiki Review, they are also asked to update the relevant Wikipedia page (17,43). Similarly, since 2008 the journal *RNA Biology* has required authors of articles about new RNA families to also create or update the corresponding Wikipedia page (44).

Benefits and limitations

This format has the advantage that the content length, depth and style can be tailored for the journal and Wikipedia versions. This may be particularly useful for highly technical topics, but involves additional work for the authors. It is also compatible with any Wikipedia article and allows the journal to use any license (including closed access). However, closed access publications have several problems. The use of any text or images from Wikipedia is disallowed, and paywalled journal articles are inaccessible by the public following the link from Wikipedia. It also doesn't provide the same level of validation for the Wikipedia content as other models. Overall, this model is easily implemented by any journal and expandable but comes with a few drawbacks from an open-access perspective.

Other models

Other models include the possibility for journals to be involved in organising expert peer review for articles even if they do not publish the outcomes. An example of this is *The BMJ* and its involvement in the peer review of the Wikipedia article on Parkinson's disease. This expands on other examples where journals have assessed accuracy in Wikipedia articles by inviting reviewers to identify errors, such as the famous *Nature* article in 2005 that listed all errors in 42 articles which were then corrected over the subsequent five weeks (7,45).

Who pays?

Wikipedia is free to read and edit, whereas scholarly publishing requires payment from either readers (subscription) or authors (open access). In the case of the *PLOS* journals, there is typically an article processing fee (APF) of $2250 (46). However, this is waived for Topic Page review articles and so publishing costs are borne by the journal. *WikiJournal of Medicine* takes greater advantage of volunteer labour and its web hosting cost is covered by the Wikimedia Foundation, so it is able to publish without fees of any kind (31,32). Conversely, the *Gene* and *RNA Biology* journals are subscription by default, but will publish via open access for an APF of approximately $3300 and $1500 respectively (47,48).

Conclusions

All of these models couple the extreme reach and accessibility of Wikipedia with the rigorous peer review and specialist readership of academic journals. They therefore impact both the general public and academic corpus. Each model has its relative advantages: the journal first model is easier to implement, but the Wikipedia first model offers more flexibility and scope.

Any open access journal can easily implement the journal first model. The only requirement is having an experienced Wikipedian involved in the project to help on the encyclopedia side. The technical requirements for editing Wikipedia are simple, and authors only need minor advice on style and scope of encyclopedia-like review articles.

These dual-publishing models are part of a larger push towards greater Wikipedia-academia integration by individuals, societies, and institutions. These include Wikipedia editing courses for students (49), article writing competitions (50), organising groups of experts to overhaul pages (51), donating multimedia content (33), or integrating scientific databases (52). Through all of this, the ultimate goal is that of accurate and comprehensive information, free to access for the world's population.

References

1. Article count [Internet]. *Wikipedia Statistics*. Cited 2017; Available from: stats.wikimedia.org/EN/TablesArticlesTotal.htm
2. Heilman JM, West AG. Wikipedia and Medicine: Quantifying Readership, Editors, and the Significance of Natural Language. *J Med Internet Res*. 2015;17(3):e62.
3. Allahwala UK, Nadkarni A, Sebaratnam DF. Wikipedia use amongst medical students – New insights into the digital revolution. *Med Teach*. 2012;35(4):337–337.
4. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. *Int J Med Inform*. 2009;78(10):645–55.
5. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: A Key Tool for Global Public Health Promotion. *J Med Internet Res*. 2011;13(1):e14.
6. Thompson N, Hanley D. Science Is Shaped by Wikipedia: Evidence from a Randomized Control Trial. *SSRN preprint.* 2017;3039505. Available from: ssrn.com/abstract=3039505
7. Masukume G, Kipersztok L, Das D, Shafee TMA, Laurent MR, Heilman JM. Medical journals and Wikipedia: a global health matter. *The Lancet Global Health*. 2016;4(11):e791.
8. Giles J. Internet encyclopaedias go head to head. *Nature News*. 2005;438(7070):900–1.
9. Shafee T, Masukume G, Kipersztok L, Das D, Häggström M, Heilman J. Evolution of Wikipedia's medical content: past, present and future. *JECH*. 2017;jech_2016–208601.
10. Viégas FB, Wattenberg M, McKeon MM. The Hidden Order of Wikipedia. *Online Communities and Social Computing*. 2007 (Lecture Notes in Computer Science; vol. 4564:445–54).
11. Hodson R. Wikipedians reach out to academics. *Nature News* [Internet]. 2015; Available from: dx.doi.org/10.1038/nature.2015.18313
12. Shafee T, Mietchen D, Su AI. Academics can help shape Wikipedia. *Science*. 2017;357(6351):557–8.
13. Soules A. Faculty perception of Wikipedia in the California State University System. *New Library World*. 2015;116(3/4):213–26.
14. Aibar E, Lladós-Masllorens J, Meseguer-Artola A, Minguillón J, Lerga M. Wikipedia at university: what faculty think and do about it. *The Electronic Library*. 2015;33(4):668–83.
15. Eijkman H. Academics and Wikipedia: Reframing Web 2.0+as a disruptor of traditional academic power-knowledge arrangements. *Campus-Wide Info Systems*. 2010;27(3):173–85.
16. Credit where credit is due. *Nature*. 2009;462(7275):825–825.
17. Tsueng G, Good BM, Ping P, Golemis E, Hanukoglu I, van Wijnen AJ, et al. Gene Wiki Reviews—Raising the quality and accessibility of information about the human genome. *Gene*. 2016;592(2):235–8.
18. Sugimoto CR, Work S, Larivière V, Haustein S. Scholarly use of social media and altmetrics: A review of the literature. *J. Assoc. Inf. Sci. Technol*. 2017;68(9):2037–62.
19. Bould MD, Hladkowicz ES, Pigford A-AE, Ufholz L-A, Postonogova T, Shin E, et al. References that anyone can edit: review of Wikipedia citations in peer reviewed health science literature. *BMJ*. 2014;348:g1585.
20. Rasberry L. Citing Wikipedia. *BMJ*. 2014 Mar 6;348:g1819.
21. Yan K-K, Gerstein M. The Spread of Scientific Information: Insights from the Web Usage Statistics in PLoS Article-Level Metrics. *PLoS One*. 2011;6(5):e19917.
22. Pageviews Analysis [Internet]. *WMFlabs tools*. Cited 2017. Available from: tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&range=latest-365&pages=Binswanger%27s_disease|Bipartite_patella|Recombinant_malaria_vaccine|Resistin|UK_Medical_Act_1876|Olney%27s_lesions|Pyknocytosis|Monoblast|Vasitis_nodosa|Kawashima_procedure
23. Pageviews Analysis [Internet]. *WMFlabs tools*. Cited 2017. Available from: tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&range=latest-

365&pages=Abortion|Epilepsy|Fever|Pregnancy|Pneumonia|Cancer|Influenza|Vaccine|Bone_density

24. Massviews Analysis [Internet]. *WMFlabs tools*. Cited 2017. Available from: tools.wmflabs.org/massviews/?platform=all-access&agent=user&source=category&target=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FCategory%3AHigh-importance_medicine_articles&range=latest-365&subjectpage=1

25. CC BY 4.0 [Internet]. *Creative Commons*. Cited 2017. Available from: creativecommons.org/licenses/by/4.0

26. Swan A. Open access self-archiving: An Introduction. *JISC/Key Perspectives*. 2005;4406.

27. Wodak SJ, Mietchen D, Collings AM, Russell RB, Bourne PE. Topic Pages: PLoS Computational Biology Meets Wikipedia. *PLoS Comput Biol*. 2012;8(3):e1002446.

28. Luk A. Continuing to Bridge the Journal-Wikipedia Gap: Introducing Topic Pages for PLOS Genetics [Internet]. *PLOS Biologue*. 2017. Available from: blogs.plos.org/biologue/2017/04/12/continuing-to-bridge-the-journal-wikipedia-gap-introducing-topic-pages-for-plos-genetics

29. Bliven S, Prlić A. Circular Permutation in Proteins. *PLoS Comput Biol*. 2012;8(3):e1002445.

30. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. 2017;13(5):e1005457.

31. Shafee T, Das D, Masukume G, Häggström M. WikiJournal of Medicine, the first Wikipedia-integrated academic journal. *Wiki J Med*. 2017;4(1):001.

32. About [Internet]. *WikiJournal of Medicine*. Cited 2017. Available from: en.wikiversity.org/wiki/WikiJournal_of_Medicine/About#Unique_publication_features

33. Staff of Blausen Medical. Medical gallery of Blausen Medical 2014. *Wiki J Med*. 2014;1(2):010.

34. Shafee T, Lowe R. Eukaryotic and prokaryotic gene structure. *Wiki J Med*. 2017;4(1):002.

35. Heilman JM, De Wolff J, Beards GM, Basden BJ. Dengue fever: a Wikipedia clinical review. *Open Med*. 2014;8(4):e105.

36. Maskalyk J. Modern medicine comes online: How putting Wikipedia articles through a medical journal's traditional process can put free, reliable information into as many hands as possible. *Open Med*. 2014;8(4):e116.

37. Wright M, Skaggs W, Årup Nielsen F. The Cerebellum. *Wiki J Med*. 2016;3(1):001.

38. Lalchhandama K. Plasmodium falciparum erythrocyte membrane protein 1. *Wiki J Med*. 2017;4(1):004.

39. Page history [Internet]. *Wikipedia*. Cited 2017. Available from: en.wikipedia.org/wiki/Help:Page_history

40. ArticleInfo [Internet]. *WMFlab tools*. Cited 2017. Available from: tools.wmflabs.org/xtools-articleinfo

41. Flock F, Laniado D, Stadthaus F, Acosta M. Towards Better Visual Tools for Exploring Wikipedia Article Development — The Use Case of "Gamergate Controversy." *Ninth International AAAI Conference on Web and Social Media*. 2015. p48–55.

42. Floeck F, Acosta M. whoCOLOR [Internet]. *f^2*. Cited 2017. Available from: f-squared.org/whovisual

43. Su AI, Good BM, van Wijnen AJ. Gene Wiki Reviews: marrying crowdsourcing with traditional peer review. Gene. 2013 Dec 1;531(2):125.

44. Butler D. Publish in Wikipedia or perish. *Nature news* [Internet]. 2008; Available from: dx.doi.org/10.1038/news.2008.1312

45. Nature December 2005 identified Errors [Internet]. *Wikipedia.* Cited 2017. Available from: en.wikipedia.org/wiki/Wikipedia:External_peer_review/Nature_December_2005/Errors

46. Publication fees [Internet]. *PLOS*. Cited 2017. Available from: journals.plos.org/ploscompbiol/s/publication-fees

47. Publication Costs [Internet]. *RNA Biology*. Cited 2017. Available from: tandfonline.com/action/authorSubmission?show=instructions&journalCode=krnb20

48. Elsevier. Guide for Authors [Internet]. *Gene*. Cited 2017. Available from: elsevier.com/journals/gene/0378-1119/guide-for-authors#13300

49. Azzam A, Bresler D, Leon A, Maggio L, Whitaker E, Heilman J, et al. Why Medical Schools Should Embrace Wikipedia: Final-Year Medical Student Contributions to Wikipedia Articles for Academic Credit at One School. *Acad Med*. 2017;92(2):194–200.
50. Kilpatrick AM. The 5th ISCB Wikipedia Competition: Coming to a Classroom Near You? *PLoS Comput Biol*. 2016;12(12):e1005235.
51. Childs O. We've joined forces with Wikipedia to improve cancer information [Internet]. *Cancer Research UK Science blog*. 2014. Available from: scienceblog.cancerresearchuk.org/2014/01/16/weve-joined-forces-with-wikipedia-to-improve-cancer-information/
52. Burgstaller-Muehlbacher S, Waagmeester A, Mitraka E, Turner J, Putman T, Leong J, et al. Wikidata as a semantic framework for the Gene Wiki initiative. *Database*. 2016;2016(1):baw015.