

Towards 'Health Information for All': Medical content on Wikipedia received 6.5 billion page views in 2013.

 blogs.lse.ac.uk/impactofsocialsciences/2015/05/19/towards-health-information-for-all-wikipedia/

5/19/2015

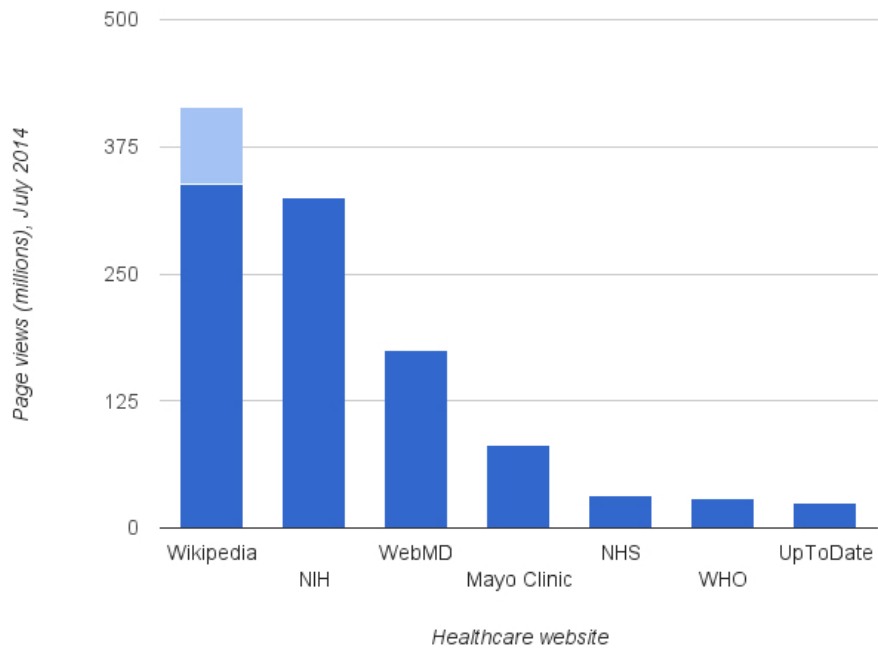
*The medical content in Wikipedia receives substantial online traffic, links to a great body of academic scholarship and presents a massive opportunity for health care information. **James Heilman** and **Andrew West** present their findings on the wider editorial landscape looking to improving the quality and impact of medical content on the web. Data points to the enormous potential of these efforts, and further analysis will help to focus this momentum.*



Recently we published "[Wikipedia and Medicine: Quantifying Readership, Editors, and the Significance of Natural Language](#)" in the Journal of Medical Internet Research (JMIR). Some of the key-findings include:

1. Medical content received about 6.5 billion page views in 2013. This was extrapolated from empirical per-article "desktop" traffic data plus project-wide mobile estimates.
2. Wikipedia appears to be the single most used website for health information globally, exceeding traffic observed at the NIH, WebMD, WHO et al..
3. Wikipedia's health content is more than 1 billion bytes of text across more than 255 natural languages (equal to 127 volumes the size of the Encyclopedia Britannica).
4. Nearly 1 million references support this content and the number of references has more than doubled since 2009, with citation density outpacing raw byte growth.
5. Article access patterns are heavily language-dependent, i.e., most topics show substantially differing popularity ranks across Wikipedia's different natural language versions.
6. The core community of editors numbers less than 300, having shrunk significantly over the last 5 years.
7. The core community is about 50% health care providers; 85% have completed a university degree; 50% have more than one university degree.

The paper contains many more data-driven insights we encourage readers to investigate. Equally important, this evidence provides us further motivation to expand translation of medical content across natural languages, continue to improve the quality of medical content, and identify partnerships and fundraising opportunities that will allow us to overcome current challenges.



Single most used website for health information – Figure 1:Health care site traffic comparison. Light blue portions represent official Wikimedia Foundation data.

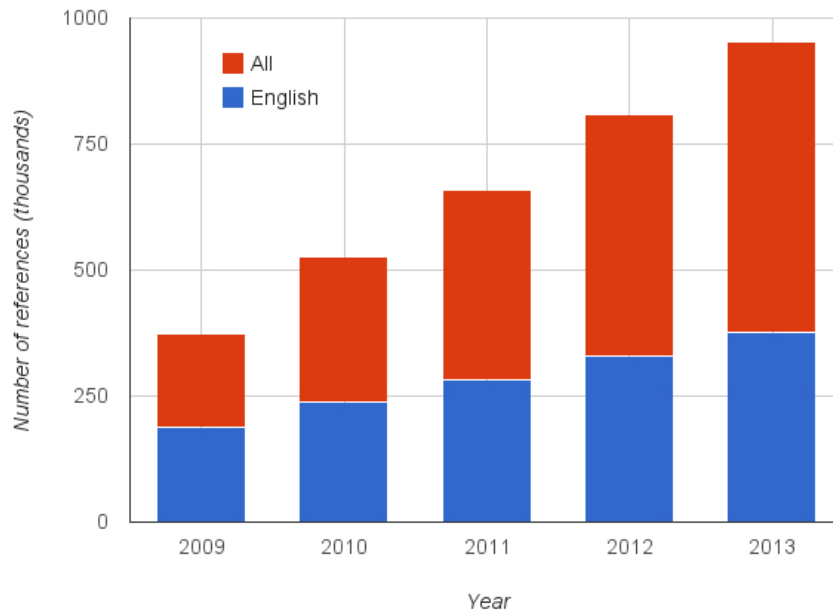
Translating medical content

Structured translation efforts for medical topics have been ongoing since 2012 as a collaboration between Wikiproject Medicine (WPMED) and Translators Without Borders (TWB). Naive approaches like using Google Translate fail because of their inaccuracy (especially with technical medical content) and limited language scope. In contrast, the current WPMED-TWB collaboration operates in 100+ languages and has translated more than 4 million words to date. This work may be useful for those developing translation engines.

One highlight is exemplified by the recent West African Ebola Outbreak, when the team was able to help quickly increase the number of languages with an “Ebola” article from 40 to more than 110. This work was additionally aided by a for profit translation company, Rubric, freely donating their services. Data from Microsoft shows that Wikipedia was the single most used source of Ebola information in the 3 most affected countries (Sierra Leone, Liberia, and Guinea), ahead of CNN, CDC, and the World Health Organization (WHO) during the worst part of the outbreak.

Partnerships to improve core/English content

A number of organizations have now partnered with Wikiproject Medicine to improve English medical content, including the University of California at San Francisco (UCSF), Mount Sinai, the Cochrane Collaboration, the Sustainable Sanitation Alliance, and the National Institutes of Health. Discussions are ongoing with WHO regarding a collaboration involving vaccine related content.



Nearly 1 million references support medical content – Figure 2: Citations/references appearing in the medical content of English Wikipedia and all Wikipedia languages based on year-end snapshots.

Notably, the UCSF collaboration began in 2012 after James Heilman presented lectures on “Wikipedia and medicine” at the institution. This was a catalyst for Amin Azzam, a professor at the school, to assemble an elective course that provided Wikipedia training, followed by a 4 week period where students actually edited and improved medical articles. Review and guidance was provided by librarians and faculty at UCSF, as well as core members of Wikiproject Medicine.

Another interesting experiment was performed where a high-quality Wikipedia article was submitted to the Journal of Open Medicine and underwent peer review. While there are expenses associated with this form of review and dissemination, the eventual publication of the article serves as proof that Wikipedia can support professional-quality medical content. Additionally, it is an opportunity for authors of Wikipedia to receive academic recognition for their work. Work is ongoing to develop an internal peer reviewed medical journal which is pubmed indexed.

Barriers to growth

Despite supporting 6.5 billion page views in 2013, the authors/translators/contributors to medical content are all volunteers. The project did receive a small amount of funding (\$12,000) via the Indigo Foundation to support the development of content in East African languages. This funding allowed content to be developed significantly more quickly than in other comparable languages.

It is also encouraging that Wikipedia as a whole has secured free and unlimited data access to its content via cellphone partnerships, enabling a knowledge resource for 400+ million people in the developing world. While this has enormous potential for the -consumption- of content, the mobile-devices which are pervasive in these regions are a difficult form factor from which to -contribute- content. Partly due to this, it remains difficult to find translators for languages of the developing world.

Expanding efforts

While the above points highlight some successes of Wikiproject Medicine and its partners, the challenges also indicate there is much work to be done. Data points to the enormous potential of these efforts, and further data

analysis can help us understand how to focus this work and make it more impactful. Human volunteers will undoubtedly continue their tireless contributions and further volunteers are key, but greater funding has the potential to greatly speed our progress. Paid staff could enable experts to focus on content instead of administrative work.

Additionally, more universities are interested in engaging with Wikipedia. However, programs that familiarize potential contributors with Wikipedia norms, like that at UCSF, involve a steep learning curve and the time-intensive involvement of existing experts. At current, these factors inhibit the scalability of the initiative. Expanding the program will require provisioning “online teaching assistants”, whose funding beyond a June 2015 pilot remains unclear.

Funding could also be used to expand translation centers around the world. If such assistance is obtained, it will also allow faster broaden of our scope of emphasis. Initial translation efforts focused on major medical disease (HIV/AIDS, tuberculosis, etc.) and we are expanding slowing into topics such as women’s health, sanitation matters, and items from the WHO List of Essential medicines; however, work moves slowly.

Featured image credit: [Wikipedia logo](#) [David Friedland’s “silver ball” logo](#) (CC BY-SA)

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Authors

James Heilman is one of Wikipedia’s most prolific medical editors. He is a Canadian emergency room physician and clinical faculty at the University of British Columbia. He is a founding member of Wiki Project Med Foundation, a NGO that works to promote the improvement of Wikipedia’s medical content. Email: [jmh649\[at\]gmail.com](mailto:jmh649[at]gmail.com)



Andrew West is a Senior Research Scientist at [Verisign Labs](#) of [Verisign, Inc.](#). He completed his Ph.D. at the [University of Pennsylvania](#) in the [Dept. of Computer and Information Science](#) in Philadelphia, PA. He received his M.S.E. from Penn in 2010 and did undergraduate study at [Washington & Lee University](#) in Lexington, VA where he received my B.Sc. in [Computer Science](#) in 2007.



- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.