

Developing a scalable framework for partnerships between health agencies and the Wikimedia ecosystem

Daniel Mietchen[‡], Lane Rasberry[‡], Thais Morata^{§,1}, John P Sadowski[¶], Jeanette Novakovich[§], James M Heilman[#]

[‡] University of Virginia, Charlottesville, United States of America

[§] National Institute for Occupational Safety and Health, Cincinnati, United States of America

[|] Cochrane Work, Amsterdam, Netherlands

[¶] National Institute for Occupational Safety and Health, Washington, United States of America

[#] University of British Columbia, Vancouver, Canada

Corresponding author: Daniel Mietchen (daniel.mietchen@virginia.edu), Thais Morata (tmorata@cdc.gov)

Academic editor: Victor Padilla-Sanchez

Abstract

In this era of information overload and misinformation, it is a challenge to rapidly translate evidence-based health information to the public. Wikipedia is a prominent global source of health information with high traffic, multilingual coverage, and acceptable quality control practices. Viewership data following the [Ebola crisis](#) and during the [COVID-19 pandemic](#) reveals that a significant number of web users located health guidance through Wikipedia and related projects, including its media repository Wikimedia Commons and structured data complement, Wikidata.

The basic idea discussed in this paper is to increase and expedite health institutions' global reach to the general public, by developing a specific strategy to maximize the availability of focused content into Wikimedia's public digital knowledge archives. It was conceptualized from the experiences of leading health organizations such as [Cochrane](#), the [World Health Organization \(WHO\)](#) and [other United Nations Organizations](#), [Cancer Research UK](#), [National Network of Libraries of Medicine](#), and Centers for Disease Control and Prevention (CDC)'s [National Institute for Occupational Safety and Health \(NIOSH\)](#). Each has customized strategies to integrate content in Wikipedia and evaluate responses.

We propose the development of an interactive guide on the Wikipedia and Wikidata platforms to support health agencies, health professionals and communicators in quickly distributing key messages during crisis situations. The guide aims to cover basic features of Wikipedia, including adding key health messages to Wikipedia articles, citing expert sources to facilitate fact-checking, staging text for translation into multiple languages; automating metrics reporting; sharing non-text media; anticipating offline reuse of

Wikipedia content in apps or virtual assistants; structuring data for querying and reuse through Wikidata, and profiling other flagship projects from major health organizations.

In the first phase, we propose the development of a curriculum for the guide using information from prior case studies. In the second phase, the guide would be tested on select health-related topics as new case studies. In its third phase, the guide would be finalized and disseminated.

Keywords

health communication; Wikipedia; consumer health information; health information systems; information networks; information science; information sharing; public health, health promotion; misinformation; infodemic.

Overview and background

In 2020, the WHO acknowledged it was not only fighting a pandemic caused by the SARS CoV-2 virus but was also engaged in an “infodemic”. An infodemic occurs when an abundance of information, both accurate and misleading, spreads rapidly alongside an epidemic (World Health Organization 2020a). This type of information ecosystem makes it difficult for people to locate trustworthy information and reliable guidance. Infodemics have occurred before the COVID-19 pandemic; however, its scale has required public health agents to develop new strategies for managing them (World Health Organization 2020b). Chou et al. (2020) argue that in this type of information ecosystem, a traditional approach to misinformation centered on debunking would likely prove insufficient. These authors called for interdisciplinary research to develop new health strategies that identify the most effective timing, manner, and forums for responding to misinformation. They suggest that one such strategy could rely on getting ahead of misinformation, by developing proactive responses or inoculations against misinformation (Chou et al. 2020). Making health information easier to locate and understand is an initial and necessary step in science translation, so that evidence-based public health practices can be put into use (Elliott and Resnik 2019).

A July 2020 roundtable on health literacy convened by the National Academies of Sciences, Engineering, and Medicine explored an additional challenge resulting from the proliferation of health and medical misinformation and disinformation during the COVID-19 pandemic. The roundtable identified “midinformation” as a category distinct from misinformation and disinformation, being an information crisis where the public experiences a state of “informational ambiguity based on scant knowledge or emerging scientific evidence” (National Academies of Sciences, Engineering, and Medicine 2020). Combatting “midinformation” requires making sure that the information users see first is the information you want them to see (National Academies of Sciences, Engineering, and Medicine 2020). A possible strategy to reach audiences who rely on internet searches is to

identify the information pathways used by them (National Academies of Sciences, Engineering, and Medicine 2020).

Wikipedia is recognized for advancing public health (Heilman et al. 2011, Mathew et al. 2013, Cohen 2014, Masukume et al. 2016, Morata and Chadha 2019, Smith 2020, Wikimedia Foundation 2020, McNeil 2020) and is one of the most accessed health information sources (Heilman and West 2015, Smith 2020), with over 30,000 medical articles on the English Wikipedia alone receiving over 5 million pageviews a day in April-May 2019 on average (Maggio et al. 2020). Wikipedia ranks highly in Google search, and users often select it (Calabrese et al. 2019). It can be used to deliver information quickly and effectively to large, established, relevant audiences. While commercial media platforms such as Twitter or Facebook make it easy to post messages and measure impact, Wikipedia is a volunteer nonprofit with strict and unique content guidelines that can be more challenging for expert use. However, science-based information on health conditions, therapies, and events (including epidemics or disasters) can be missing or underdeveloped on Wikipedia. Some health agencies developed strategies to integrate content in Wikipedia and evaluate readership and community response (Lum et al. 2018, Morata and Lum 2018, Murphy et al. 2019). No one has yet produced generalized instructions specifically for health organizations on how to engage quickly, easily and effectively with Wikipedia's content and community.

Consequently, there is an as yet unmet need to develop a general-interest interactive guide on the Wikipedia and Wikidata platforms to support health professionals/communicators in quickly distributing key messages during crisis situations. Wikipedia has guides for publishing content such as biographies and specialized health information, but no guide for organizations to share updated population-level health messages. We propose creating this general-interest interactive guide using the Wikipedia platform's plentiful, developed native tools. The proposed infrastructure would be used to publish health information in Wikipedia to create a case study/demonstration and report pageviews. We aim to lower Wikipedia's barrier of accessibility to health communicators who wish to use Wikipedia for instant publication of institutionally backed messages. The proposed guide would cover basic features of Wikipedia: key messages, citations, and collecting readership metrics.

More advanced features might include

- using multiple language versions of Wikipedia via translation;
- setting up automated metrics reporting;
- sharing non-text media;
- anticipating off-Wikipedia or offline reuse of Wikipedia content in apps or virtual assistants such as Apple's Siri or Google Assistant;
- using Wikidata to collect, curate and share data; and
- describing other flagship projects which major health organizations designed and incorporated into Wikipedia engagement processes.

Our proposed guide should be tested on selected focused health-related topics as case studies of what Wikipedia communication and impact can offer.

Whatever the selected topic of interest, common elements are

- collecting key health messages from reputable agencies,
- empowering organizational representatives to get those messages into Wikipedia's ecosystem and
- accessing reports of the size of the audience which reached information through the relevant Wikipedia pages.

Objective

To increase and expedite health institutions' global reach to the general public, we propose the development of a general-interest interactive guide on Wikipedia to support health professionals/communication capacities to disseminate health-related information via the Wikipedia ecosystem.

Impact

The proposed plan would improve the timeliness and accuracy of communications and situational awareness regarding threats to the public's health and at-risk populations by expanding the delivery of health information to not only public health professionals, but also to large global audiences through Wikipedia. We expect it to achieve high communication impact, which we will evaluate through metrics reports. These reports can include a list of articles edited, photos and datasets shared, language communities served, publications cited, and pages viewed. Based on the precedent of classroom projects (e.g. Azzam et al. 2017, Weiner et al. 2019) and our review of traffic reports, we conservatively estimate the content would receive more than 10 million views in one year. Additionally, the guide and case studies developed under this project can make publishing to Wikipedia a more accessible option for health organizations globally. The guide would provide peer-to-peer advice for professionals to orient themselves and consider options and opportunities for reaching an expansive audience.

Finally, the impact of this project is likely to extend beyond science communication to the general public. Newer systems that evaluate scholarly impact today recognize the importance of a publication appearing in Wikipedia. For example, [Altmetric](#) measures the influence and online reach of scholarly output through readership counts, mentions in news sites, blog posts, citation manager bookmarks, social media, and Wikipedia citations. Contributing peer-reviewed scientific information into Wikipedia boosts not only the diffusion of an agency's knowledge, but also its scholarly recognition.

Implementation

Wikipedia serves a large number of readers who expect and need high quality information on specific topics, complete with bona-fide citations and links to reliable sources. Health organizations, on the other hand, have high quality content but difficulty delivering it to

readers at scale. Participant agencies and Wikipedia can provide each other mutual support by expediting and expanding access to selected content to Wikipedia's established large audience. In coordination with established Wikipedia community projects such as [WikiProject Medicine](#), [WikiProject Disaster Management](#), or [Wikimedians for Disaster Response](#), this proposed idea aims to develop processes to help ensure that health information in Wikipedia is current, high quality, accurate, and translated in multiple languages.

Our main goal is to develop a model which expert health organizations can use to implement a health communication campaign in the Wikipedia ecosystem. This model includes presenting a case study of Wikipedia publishing, demonstrating how to collect communication impact metrics, presenting examples and an explanation of how the publishing and audience metrics are evidence of reaching new audiences, and finally packaging all of these outputs into a guide which can support any health organization's engagement in Wikipedia. Specific activities include the following (cf. Fig. 1):

Develop a Wikipedia case study for a specific situation or condition

- As background, note the precedent of Wikipedia's prior crowd organized projects including for Ebola, Zika, COVID-19 as well as Medicine and disasters in general.
- Select a topic and analyze multiple communication models for Wikipedia.
- Systematically improve Wikimedia content that overlaps and is broader than the target disease or disaster.

A good way to start is by selecting a focused health-related target topic to be used as a case study. Wikipedia has mechanisms to respond to highly dynamic situations, to assess the reliability of resources, to collaborate and to leverage expertise (Keegan and Tan 2020, Benjakob et al. 2021). The proposed concerted effort between health agencies and the Wikipedia community would strengthen Wikipedia content when quick reactions are needed. This can be achieved by systematically improving - ahead of time - Wikimedia content not only about the target topic but also "one step removed" from it, so that when the need to contextualize new information arises, it meets a scaffold (or "scientific backbone", in the language of Benjakob et al. 2021) of already partly contextualized pre-existing knowledge ahead of public health emergency situations, thereby improving disaster preparedness, disaster literacy and disaster response. For example, in past epidemics, important complementary information included cultural context like burial practices for Ebola, microcephaly and mosquito control for Zika, or social distancing, face coverings and other non-pharmaceutical interventions for COVID-19.

Milestones in this phase of development include the following:

- Compilation of a list of 20+ Wikipedia articles related to that topic, which would be the focus of development and consideration in the case study.
- Identification of existing relevant publications on the chosen topic, from expert health organizations, from which we collect key messages to integrate into Wikipedia.

- A list of 100 technical terms, curated as structured data in Wikidata, which would be targets for translation as needed.
- A snapshot traffic report of all of this content measuring audience reach.
- A preliminary engagement report surveying existing Wikipedia community stakeholder engagement in all of the elements above.

Collect communication impact metrics

- Present the background context that in digital publishing, there is an established culture of monitoring social media metrics to measure audience engagement as a way of demonstrating meaningful content delivery. Communication professionals engaged with platforms such as Twitter, Facebook, and YouTube do this.
- Set up automated metrics reporting for selected Wikipedia articles and other Wikimedia resources, based on an inventory of key topics for each selected scenario. A survey of topics will make it possible to prioritize areas and materials of interest.
- Develop mechanisms to alert Wikipedians of new publications on the selected topic and alert the agency's staff which of their publications are mentioned in Wikipedia, noting if they are current or need to be updated. This process includes giving credit and doing disclosure of which organizations engaged in Wikipedia and how they interacted.

This phase of development establishes channels for exchanging information between expert organizations and Wikipedia. Expert organizations require traffic reports, community feedback, and alerts of media uptake from Wikipedia. The Wikipedia community requires documentation of institutional engagement, Wikipedia-aligned media contributions, and a forum for stakeholder conversation from organizational partners. This information exchange would occur in a WikiProject or similar Wikipedia ecosystem documentation model, where both staff of the organization and community stakeholders can meet to view all communication metrics, shared resources, and documentation. The project will build from precedents including the Wiki Education Foundation's process for monitoring classroom student editing, Consumer Reports' practice of measuring Wikipedia article pageviews, and [Cochrane's interface](#) presenting source publications to cite in Wikipedia articles. The initiative will support the inclusion of relevant evidence within Wikipedia emergency-related articles, as well as processes to help ensure that emergency preparedness and response information included in Wikipedia is of the highest quality and as accurate as possible. Trusted, evidence-based research can ultimately help people make informed decisions about their own health.

Milestones in this phase of development include the following:

- A Wikipedia project page targeted to public health and expert communication managers tracking media use in Wikipedia:
 - A link to an online tool that counts and reports where Wikipedia links to the participant agency publications, as with a citation.

- A link to an online tool that counts and reports where the participant agency's images and non-text media appear in Wikipedia.
- A scholarly profile of the participant agency publications indexed in WikiCite (the Wikimedia project somewhat analogous to PubMed).
- A white paper explaining the significance of the Wikipedia project page, that:
 - Describes the communication impact of Wikipedia.
 - Describes how and why we chose these tools.
 - Provides a preliminary guide for replication.

Reach new audiences

- Use Wikipedia as an additional media channel to transfer publications and metadata from any target agency to Wikimedia Commons and Wikidata on relevant topics. Maintain a focused category in Commons for [media](#) and [publications](#).
- Examine the feasibility of contributing to multiple language versions of Wikipedia via assisted or semi-automated translation or dedicated community engagement mechanisms.
- Develop offline Wikipedia versions for limited or compromised internet access that include selected, updatable agency-specific resources. Offline Wikipedia will be customized and distributed by [Internet-in-a box](#) and by the WikiMed [app](#). Lessons learned from previous experiences with this tooling from the [Icahn School of Medicine at Mt. Sinai](#) and [Columbia University](#) public health programs and [others](#) will inform the effort.

The goal of this phase is demonstrating actual communication for impact by sharing the key health messages collected during the first phase while monitoring activity through the processes set up in the second phase. Wikipedia offers diverse options for communication, including decisions of emphasis on sharing either text, multimedia, or data; as well as achieving this with variations such as translation, optimization for off-wiki distribution, and deciding a balance of paid-staff engagement versus encouraging crowdsourced development.

Milestones in this phase of development include the following:

- Development of Wikipedia articles by adding key messages to them and citing the expert sources from which those messages originate.
- Publication of at least 100 media files from the participant agency in Wikimedia Commons.
- Integration of these media files into Wikipedia, Wikidata, and other Wikimedia platforms, including for metrics tracking.
- Audience reach of 10 million pageviews/year.

Create a guide which presents training materials

- Produce a guide to support expert health organizations in sharing health content in Wikipedia.

- Develop and deliver optional training for agency personnel on the scope and procedures for structuring communication efforts through Wikimedia in the context of emergencies.
- Develop a plan for data scientists from the agency to facilitate data integration with Wikidata and Wikipedia, with appropriate quality assurance.

This phase of development reviews all previous activity and documents it, targeting health communication professionals who would replicate the activity while representing an expert health organization which is publishing in Wikipedia. While the guide will be general, we will get feedback from specific health organizations which express interest in exploring its future use.

Milestones in this phase of development include the following:

- Finalize the development of the WikiProject or comparable on-Wikipedia documentation started in phase 1.
- Finalize drafting of the white paper started and circulated in phase 2.
- Produce a snapshot report of metrics recorded in phase 3, with interpretation including an estimate of the hours spent to achieve each sort of communication impact.
- Produce a standalone guide suitable for off-Wikipedia platform distribution, and which is in a format that health communication professionals would find familiar.

Resource considerations

Our proposal can only fully achieve success if carried out in coordination with support from Wikipedia community projects such as the ones mentioned here established Wikipedia community projects and staff from the participant organizations. A kick-off planning meeting with representatives of Wikimedia entities would be organized at the onset of the project to facilitate detailed conceptual and strategic planning. Recruiting public health informatics and data science expertise would be key in the first two phases of the project. Instructional design expertise should be considered for the third phase. Finally, minor purchases of supplies would be dedicated to customization of the offline Wikipedia through the Internet-in-a-box platform.

Concluding remarks

At the moment of submission of this manuscript, our proposal is unfunded, but several of its elements are being executed by the agencies mentioned in the abstract and the WikiProjects aforementioned under Approach. The authors welcome the adoption of the proposed strategy or any of its elements by groups who share this project's goals.

Acknowledgements

The authors thank Tim Moody, Internet-in-a-box; Alexandre A.P. Montilha, Universidade de São Paulo; and CAPT David Byrne, National Institute for Occupational Safety and Health; who read and commented on earlier versions of this article. We also thank Amin Azzam at the University of California, San Francisco, for the constructive review.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention CDC. Mention of any company or product does not constitute endorsement by NIOSH or the CDC.

In addition, citations to websites external to NIOSH do not constitute NIOSH endorsement of the sponsoring organizations or their programs or products. The description of the project idea is not associated with any specific funding opportunity, nor will it have any bearing in funding decisions. Furthermore, NIOSH is not responsible for the content of cited websites. All web addresses referenced in this document were accessible as of the publication date.

References

- Azzam A, Bresler D, Leon A, Maggio L, Whitaker E, Heilman J, Orlowitz J, Swisher V, Rasberry L, Otoide K, Trotter F, Ross W, McCue JD (2017) Why Medical Schools Should Embrace Wikipedia: Final-Year Medical Student Contributions to Wikipedia Articles for Academic Credit at One School. *Academic medicine : journal of the Association of American Medical Colleges* 92 (2): 194-200. <https://doi.org/10.1097/ACM.0000000000001381>
- Benjakob O, Aviram R, Sobel J (2021) Citation needed? Wikipedia and the COVID-19 pandemic. *BioRxiv*[In English]. <https://doi.org/10.1101/2021.03.01.433379>
- Calabrese C, Anderson BN, Barnett GA (2019) Online Representations of “Genome Editing” Uncover Opportunities for Encouraging Engagement: A Semantic Network Analysis. *Science Communication* 41 (2): 222-242. [In English]. <https://doi.org/10.1177/1075547018824709>
- Chou WS, Gaysynnsky A, Cappella JN (2020) Where we go from here: Health misinformation on social media. *American Journal of Public Health* 110: S273-S275. [In English]. <https://doi.org/10.2105/AJPH.2020.305905>
- Cohen N (2014) Wikipedia emerges as trusted internet source for Ebola information. *New York Times*. URL: <https://www.nytimes.com/2014/10/27/business/media/wikipedia-is-emerging-as-trusted-internet-source-for-information-on-ebola.html>
- Elliott K, Resnik DB (2019) Making open science work for science and society. *Environmental Health Perspectives* 127 (7): 75002-75002. [In English]. <https://doi.org/10.1289/EHP4808>
- Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, Iberri DJ, Harvey M, Thomas B, Stomp W, Martone MF, Lodge DJ, Vondracek A, de Wolff JF,

- Liber C, Grover SC, Vickers TJ, Meskó B, Laurent MR (2011) Wikipedia: A Key Tool for Global Public Health Promotion. *Journal of Medical Internet Research* 13 (1). <https://doi.org/10.2196/jmir.1589>
- Heilman JM, West AG (2015) Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. *Journal of medical Internet research* 17 (3): e62. <https://doi.org/10.2196/jmir.4069>
 - Keegan B, Tan C (2020) A Quantitative Portrait of Wikipedia's High-Tempo Collaborations during the 2020 Coronavirus Pandemic. arXiv. URL: <https://arxiv.org/abs/2006.08899>
 - Lum M, Morata TC, Hare J, Sadowski JP (2018) Expanding and Improving Occupational Safety and Health Content in Wikipedia. *It Matters*. <https://blogs.cdc.gov/NIOSH-Science-blog/2018/07/23/Osh-wikipedia/>. Accessed on: 2021-1-21.
 - Maggio LA, Steinberg RM, Piccardi T, Willinsky JM (2020) Reader engagement with medical content on Wikipedia. *eLife* 9 <https://doi.org/10.7554/elife.52426>
 - Masukume G, Kipersztok L, Das D, Shafee TMA, Laurent MR, Heilman JM (2016) Medical journals and Wikipedia: a global health matter. *The Lancet Global Health* 4 (11). [https://doi.org/10.1016/s2214-109x\(16\)30254-6](https://doi.org/10.1016/s2214-109x(16)30254-6)
 - Mathew ME, Joseph A, Heilman JM, Tharyan P (2013) Cochrane and Wikipedia: The Collaborative Potential for a Quantum Leap in the Dissemination and Uptake of Trusted Evidence. *Cochrane Database of Systematic Reviews* <https://doi.org/10.1002/14651858.ed000069>
 - McNeil D (2020) Wikipedia and the W.H.O. join to combat COVID misinformation. *New York Times*. URL: <https://www.nytimes.com/2020/10/22/health/wikipedia-who-coronavirus-health.html>
 - Morata T, Chadha S (2019) Make Quality Hearing Health Information Available to All. *The Hearing Journal* 72 (5). <https://doi.org/10.1097/01.hj.0000559493.29061.35>
 - Morata TC, Lum M (2018) Partnerships to expand occupational safety and health content in Wikipedia. 75. *International Congress on Occupational Health, Dublin, May 2-6, 2018*. ICOH. Occupational and Environmental Medicine, 75, 2 pp. [In English]. <https://doi.org/10.1136/oemed-2018-ICOHabstracts.205>
 - Murphy W, Morata T, Montilha AP, Sadowski J (2019) Using Wikipedia to promote acoustics knowledge for the International Year of Sound 2020. 178th Meeting of the Acoustical Society of America <https://doi.org/10.1121/2.0001211>
 - National Academies of Sciences, Engineering, and Medicine (2020) Addressing Health Misinformation with Health Literacy Strategies. The National Academies Press, Washington. [In English]. <https://doi.org/10.17226/26021>
 - Smith D (2020) Situating Wikipedia as a health information resource in various contexts: A scoping review. *PLOS ONE* 15 (2). <https://doi.org/10.1371/journal.pone.0228786>
 - Weiner SS, Horbacewicz J, Raspberry L, Bensinger-Brody Y (2019) Improving the Quality of Consumer Health Information on Wikipedia: Case Series. *Journal of Medical Internet Research* 21 (3). <https://doi.org/10.2196/12450>
 - Wikimedia Foundation (2020) Wikipedia and COVID-19. Explore the data. <https://wikimediafoundation.org/covid19/data/>. Accessed on: 2021-1-20.
 - World Health Organization (2020a) Immunizing the public against misinformation. <https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation>. Accessed on: 2021-1-10.

- World Health Organization (2020b) The World Health Organization and Wikimedia Foundation expand access to trusted information about COVID-19 on Wikipedia. <https://www.who.int/news/item/22-10-2020-the-world-health-organization-and-wikimedia-foundation-expand-access-to-trusted-information-about-covid-19-on-wikipedia>. Accessed on: 2020-11-20.

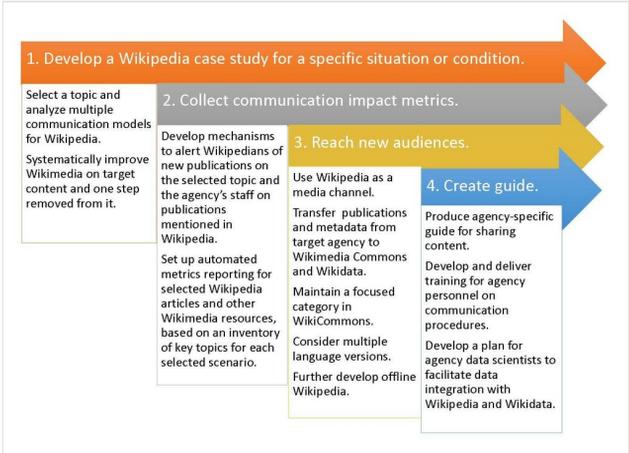


Figure 1. Schematic overview of the steps involved in developing a public health agency model for Wikipedia