

**PREPRINT**

*Author-formatted, not peer-reviewed document posted on 01/05/2021*

DOI: <https://doi.org/10.3897/arphapreprints.e68129>

---

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

 Daniel Mietchen,  Lane Rasberry,  Thais Morata, John Sadowski,  Jeanette Novakovich, James Heilman

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

Daniel Mietchen<sup>‡</sup>, Lane Rasberry<sup>‡</sup>, Thais Morata<sup>§,|</sup>, John P Sadowski<sup>¶</sup>, Jeanette Novakovich<sup>§</sup>, James M Heilman<sup>#</sup>

<sup>‡</sup> University of Virginia, Charlottesville, United States of America

<sup>§</sup> National Institute for Occupational Safety and Health, Cincinnati, United States of America

<sup>|</sup> Cochrane Work, Amsterdam, Netherlands

<sup>¶</sup> National Institute for Occupational Safety and Health, Washington, United States of America

<sup>#</sup> University of British Columbia, Vancouver, Canada

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

Reviewable

v 1

## Abstract

In this era of information overload and misinformation, it is a challenge to rapidly translate evidence-based health information to the public. Viewership data following the [Ebola crisis](#) and during the [COVID-19 pandemic](#) reveals that a significant number of readers located health guidance through Wikipedia and related projects, including its media repository Wikimedia Commons and structured data complement, Wikidata. In 2013, Wikipedia's medical content consisted of more than 155,000 articles and 1 billion bytes of text in over 255 languages, and the number of views during that year surpassed 4 billion, making it the most viewed medical resource worldwide.

The research idea discussed in this paper aims 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), [Cancer Research UK](#), [National Network of Libraries of Medicine](#), and CDC's [National Institute for Occupational Safety and Health](#) (NIOSH). Each has customized strategies to integrate content in Wikipedia and evaluate responses.

The research idea is to develop 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 translation into multiple languages; automated metrics reporting; sharing non-text media; anticipating 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 a discussion of 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, consumer health information; health information systems; information networks; information science; information sharing. public health, health promotion.

## 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 are not new to 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 and 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).

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). Wikipedia ranks highly in Google search

rankings and is selected about 93% of the time (Calabrese et al. 2019). It can be used to deliver information quickly, 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. Leading health organizations noted that 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 aggregated best practices or produced instructions for fast-and-easy Wikipedia engagement specifically for health organizations.

There is a 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 for specialized health information but has no guide for organizations to share updated top-level health messages. This research idea plans to use 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 could 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 Siri or Google Assistant; using Wikidata to collect, curate and share data; and describing other flagship projects major health organizations designed into Wikipedia engagement processes. The 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 1) collecting key health messages from reputable agencies, 2) empowering a representative of these agencies to get those messages into Wikipedia's ecosystem and 3) accessing reports of the size of the audience which reached information through Wikipedia.

## Objective

This research idea aims 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.

## 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 similar projects and our review of traffic reports, we conservatively estimate the content would receive more than 1 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 readers expect and need high quality information with bona-fide citations and links to reliable sources. Health organizations have the reverse challenge, as they have high quality content, but difficulty delivering 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](#), [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 goals are to:

- Develop mechanisms to alert experienced Wikipedia editors of reliable resources in a certain focused area as they are published.
- Develop a guide defining the scope and procedures for a range of activities that health agencies' staff can take to instantly publish institutionally backed key messages into the Wikimedia ecosystem.

**Specific activities include:**

1. Develop a Wikipedia case study for a specific situation or condition (cf. Fig. 1)
  - 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. In past epidemics, this included burial practices for Ebola; microcephaly and mosquito control for Zika; or social distancing, cloth face coverings, and other non-pharmaceutical interventions for COVID-19.
2. Collect communication impact metrics from Wikipedia analogous to social media metrics such as from Twitter, Facebook, and YouTube.
  - 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.
  - 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.
3. Reach new audiences.
  - Use Wikipedia as an additional media channel.
    - Transfer publications and metadata from 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 semiautomated translation.
  - 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 tool from Mt. Sinai and [Columbia University](#) public health programs and [others](#) will inform the effort.
4. Develop guide as basis for training materials.
  - Produce agency-specific guide for 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.

**Phase 1 Evaluation:** Steps 1–3 would be developed during the first year, with limited implementation. Development and evaluation would be on-going throughout the project. The annual report for year 1 would include:

a) Metrics for Wikipedia content that was developed with the help of project activities. Evaluation will go beyond the number of views of articles that are created or expanded, to include the extent and the text of each edit, and whether others engaged to further develop

the page. The summary reports will inform the agency on what topics interest the general public.

b) Publication of the described alerts (step 2) on Wikipedia for a community of experienced editors, including the participant agency's researchers, data scientists, and health communication professionals.

**Phase 2 Evaluation:** Guide content will be developed based on findings from steps 1–3 and tested on selected health topics as case studies.

**Phase 3 Evaluation:** Guide will be finalized, disseminated, and evaluated by cohorts who participate in early training trials.

This initiative will help increase awareness of this science communication and diffusion strategy among stakeholder organizations and generate feedback that can be used to improve the project and increase the likelihood of a successful implementation and expansion to other areas of interest or centers. 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).

*Milestone 1: Structure a Wikipedia case study for selected topic and organize collection of impact metrics*

The goal of structuring a Wikipedia case study would be achieved by creating a conventional Wikipedia project page, which will present the following elements:

1. A focused case study topic we select.
2. A list of 20+ Wikipedia articles related to that topic, edited to remove errors, and reviewed on a regular basis.
3. Three of 20 edited articles will include key messages from specific agencies.
4. A list of 100 technical terms, curated as structured data in Wikidata.
5. A traffic report of all of this content measuring audience reach.
6. An engagement report describing community response to all of the elements above.

We would start by selecting a focused health-related topic to be used as case study. Wikipedia has mechanisms to respond to highly dynamic situations, to assess the reliability of resources, to collaborate and to leverage expertise. The proposed concerted effort between health agencies and the Wikipedia community would strengthen Wikipedia content when quick reactions are needed by systematically improving Wikimedia content one-step removed from target topics to help in situations such as preparedness for disasters, disaster literacy and disaster response. For example, in past epidemics, one step removed from target topics included burial practices for Ebola, microcephaly and mosquito control for Zika, or social distancing, face coverings and other non-pharmaceutical interventions for COVID-19.

### *Milestone 2: Structure mechanisms to identify coverage of publications of the participant agency within Wikipedia*

By structuring mechanisms to identify coverage of publications of the participant agency within Wikipedia, this project will generate the following products:

1. 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).
2. 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.

This milestone focuses on developing mechanisms to alert Wikipedians of new publications by the participant agency and alert its staff to which of their publications are mentioned in Wikipedia to increase and expedite their global reach to not only public health professionals but to the general public. Alerts will also note if mentions to publications are current or if they need to be updated. This milestone will be modeled after the approach developed by [WikiProject Medicine/Cochrane](#). 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.

### *Milestone 3: Expansion of health resources in Wikimedia Commons, Wikidata, and off-line Wikipedia*

While other milestones have goals to set up mechanisms for the tracking and reporting of media the participant agency publishes in Wikipedia, this milestone seeks to move beyond experiments and perform a sizable transfer of media into the Wikimedia platform. These goals include:

1. Publication of at least 300 media files from the participant agency in Wikimedia Commons.
2. Integration of these media files into Wikipedia, Wikidata, and other Wikimedia platforms.
3. Integration of these media files into the metrics tracking interfaces described above.
4. Audience reach of 10,000,000 pageviews/year.

To achieve and complement these goals and reach new audiences, several approaches will be used to increase the discoverability of published resources, including:

- Drawing on the experience in transferring metadata to the Wikimedia Commons media repository and the Wikidata information ecosystem on relevant topics. [Wikidata](#) is an open access, free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, the other wikis of the Wikimedia movement, to anyone in the world.
- Contributing to further develop offline Wikipedia to prepare for limited or compromised internet access caused by a disaster, so that information from the participant agency is included and updated as needed.
- Leveraging the experience from Wikipedia community projects.

## Resource considerations

This research idea can only fully achieve success if carried out in coordination with several of the established Wikipedia community projects mentioned here and the participant agency staff from different centers. A kick-off concept and strategic planning meeting with representatives of Wikimedia entities would be organized at the onset of the project (cf. roadmap in Fig. 2). Public health informatics and data scientist 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.

### Phase One

1: Review health communication strategy in place on the selected topic, and processes that exist within Wikimedia platforms. Organize kick-off concepts and strategic planning meetings.

2: Purchase supplies to customize Internet-in-a-Box for the development of an off-line Wikipedia.

3: Recruit public health informatics/data science expertise.

### Phase Two

1: Attend conferences and other events on the selected topic to network with stakeholders and exchange information on needs and project progress.

2: Recruit public health informatics/data scientist expertise.

### Phase Three

1: Purchase supplies to test strategy and maintain equipment.

2: Recruit instructional design expertise.

3: Prepare reports and journal articles and travel to meet with stakeholders.

## Concluding remarks

At the moment of submission of this manuscript, this research idea remained unfunded, but several of its elements are being executed by the agencies mentioned in the abstract and the WikiProjects aforementioned under Approach. Authors welcome the adoption of the proposed strategy or its elements by groups who share this project's goals.

## Acknowledgements

The authors thank Tim Moody, Internet-in-a-box, Canada, Alexandre A.P. Montilha, Universidade de São Paulo (Brazil) and CAPT David Byrne, National Institute for Occupational Safety and Health who read and commented on earlier versions of this article.

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 research idea is not associated with any specific funding opportunity, nor will 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

- 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>
- 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.
- 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>
- 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.

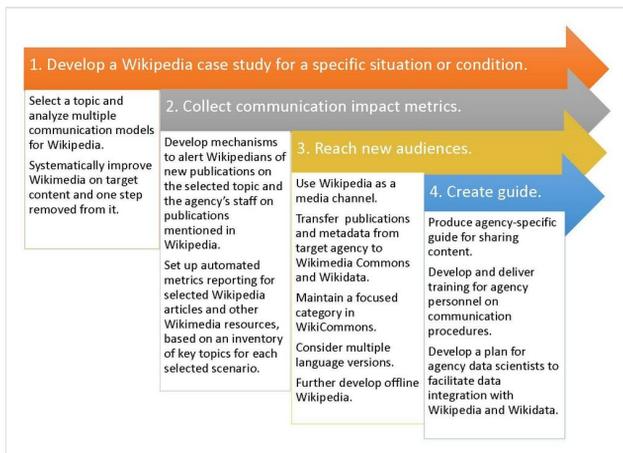


Figure 1. Schematic overview of the steps involved in developing a public health agency model for Wikipedia

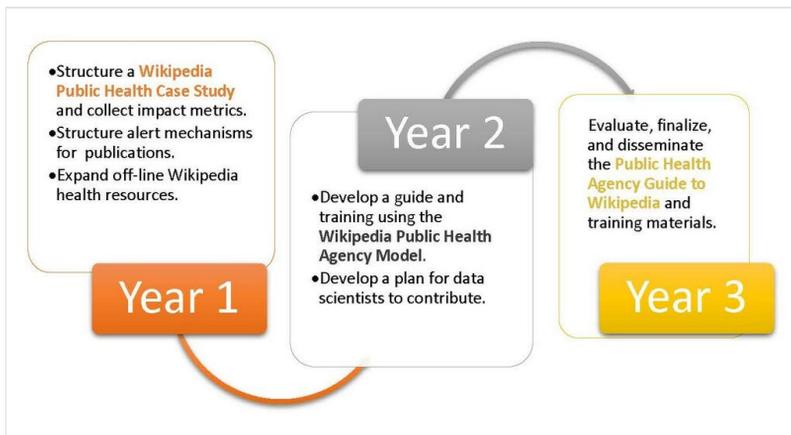


Figure 2.

Roadmap for developing a public health agency model for Wikipedia