OPEN COMMUNITY HEALTH: WORKSHOP REPORT

This report summarizes key outcomes from a workshop on open community health conducted at the University of Nebraska at Omaha in April 2018. Workshop members represented research and practice communities across Citizen Science, Open Source, and Wikipedia. The outcomes from the workshop include (1) comparisons among these communities, (2) how a shared understanding and assessment of open community health can be developed, and (3) a taxonomical comparison to begin a conversation between these communities that have developed disparate languages.

Download as PDF

AUTHORS

Georg Link <glink@unomaha.edu> University of Nebraska at Omaha, USA

Kevin Lumbard <klumbard@unomaha.edu> University of Nebraska at Omaha, USA

Nicole Damen <ndamen@unomaha.edu> University of Nebraska at Omaha, USA

Holly Rosser <hrosser@unomaha.edu> University of Nebraska at Omaha, USA

Matt Germonprez <germonprez@gmail.com> University of Nebraska at Omaha, USA

Sean Goggins <s@goggins.com> University of Missouri, USA

Andrea Wiggins <akwiggins@gmail.com> University of Nebraska at Omaha, USA

Vinod Ahuja <vahuja@unomaha.edu> University of Nebraska at Omaha, USA

Jonathan Brier <brierjon@umd.edu> University of Maryland, USA

Johanna Cohoon <jlcohoon@utexas.edu> The University of Texas at Austin, USA

Aaron Halfaker <ahalfaker@wikimedia.org> Wikimedia Foundation, USA

James Howison <jhowison@ischool.utexas.edu> The University of Texas at Austin, USA

Don Marti <dmarti@mozilla.com> Mozilla, USA

Greg Newman <gregory.newman@colostate.edu> Colorado State University, USA

Carsten Østerlund <costerlu@syr.edu> Syracuse University, USA

Ray Paik <rpaik@gitlab.com> GitLab, USA

Becky Rother <becky@zooniverse.org> Zooniverse, USA

Aaron Schecter <aschecter@uga.edu> University of Georgia, USA

INTRODUCTION

This report summarizes three significant findings from a diverse and interdisciplinary workshop on open community health in April 2018. The workshop was held at the University of Nebraska at Omaha and consisted of researchers and practitioners from the communities of Citizen Science, Free and Open Source Software (F/OSS), and Wikipedia. These communities have shared characteristics that allow them to be categorized as open communities.

Some of these characteristics are: Internet-facilitated collaboration, an open and welcoming attitude to new members, a shared mission among members, and membership defined through active engagement, since open communities cannot exist and flourish without active members. In Citizen Science, a shared mission may be to identify animals in pictures. In F/OSS a shared mission may be to develop a software library for securing Internet traffic. In Wikipedia, a shared mission may be to improve articles on female scientists.

Another important concern for any open community is the health of the community, which refers to its resilience and its ability to operate efficiently, sustainably throughout its life-cycle to fulfill its mission. Community health is not a generalizable concept because most communities have disparate missions, so it must be defined by individual communities. For example, in Citizen Science, community health may be informed by how community members collect or analyze data to aid a research team in producing valuable findings. In F/OSS, community health may mean that community members will continue developing quality software and adapt it for emerging technologies. In Wikipedia, community health may mean that community members will continue creating new content and maintain existing content. Participants in the workshop agreed that without viable community health, community members stop contributing and thus the mission of the open community can no longer be achieved.

Open communities have many stakeholders that care about and have a shared need for community health. Members want to have a community that they can identify with and that will continue contributing to a mission they believe in. Community managers want to foster community engagement and help their community be its best self. Community stewards, such as funding agencies or foundations, want communities to fulfill their missions in a self-sustaining manner or require justification for ongoing future investment.

These common characteristics reflect a shared interest among Citizen Science, F/OSS, and Wikipedia researchers and participants in open community health. To date, however, discussions of open community health are fragmented among these different open communities. There has been a wealth of research into F/OSS and Wikipedia community health which has moved past the individual project level to also take into account ecosystem health. In contrast, there have not been major discussions about Citizen Science community health and, for the most part, localized efforts within projects have not yet been aggregated into a shared community health perspective. Nevertheless, workshop participants found that open community health has fundamental similarities across all open communities and that an exchange of knowledge and best practices can be beneficial to all. Furthermore, participants began to rationalize a taxonomy framework of open community health that may assist future conversations and help in developing a theoretical framework towards open community health.
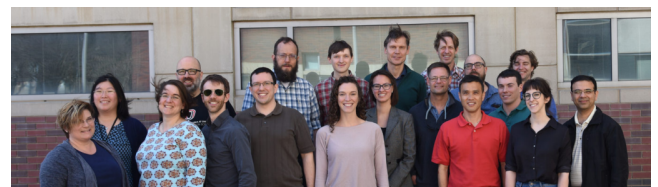
WORKSHOP DESIGN

Figure 1. Workshop participants from left to right. Front row: Holly Rosser, Christine Toh, Andrea Wiggins, Aaron Halfaker, Jonathan

Brier, Johanna Cohoon, Nicole Damen, Greg Newman, Ray Paik, Aaron Schecter, Becky Rother, Vinod Ahuja. Back row: Sean Goggins, Don Marti, Georg Link, Carsten Østerlund, Matt Germonprez, Kevin Lumbard, James Howison.

The workshop facilitated interaction, knowledge transfer, and idea generation between participants (see figure 1). The workshop started out with a primer about open community health and participants shared initiatives and projects in their own communities that address this issue. One such initiative is the CHAOSS project,[1] which focuses on metrics and tools that inform health in F/OSS communities. Other initiatives include Mozilla's diversity and inclusion efforts, led by Emma Irwin,[2] and the open source survey report. In Citizen Science, the SciStarter 2.0 platform[3] and Zooniverse Talk boards[4] were highlighted for their importance in building and analyzing open Citizen Science communities across projects. Wikipedia has run several open community health experiments on its platform to understand, for example, the effect of a new visual editor[5], a Q&A forum[6], and an orientation game[7] on new editors.

The workshop was structured in a series of breakout sessions followed each by a summary and discussion session. During each breakout session, two subgroups explored a given topic in detail. At the conclusion of the breakout session, the subgroups rejoined in a summary and discussion session to share and discuss their findings. In the first pair of sessions, participants self-selected into a group that focused on either F/OSS or Citizen Science. The goal was to discuss the current understanding of open community health within these communities. In the second pair of sessions, the participants broke into two new groups so that each group included both F/OSS and Citizen Science experts. The goal was to discuss similarities and differences between F/OSS and Citizen Science with regards to determining open community health. In the third pair of sessions, the first set of self-selected groups got back together again. The goal

was to discuss who else needed to be involved in future conversations and what the next steps were to advance the topic within each community.

Throughout the workshop, the Goal-Question-Metrics approach (Basili, 1992) fostered conversations. This approach structured and supported communication. Goals are objectives that people have when observing open community health. Questions are uncertainties and unknowns that people need answered to make decisions to reach their goals. Metrics are data points that can inform the questions and consequently allow for data-driven decisions. The data from these conversations was systematically collected through scribes (an online document was shared among scribes for note taking during each session to capture the conversation as it occurred), and participant notes (an online document was shared among participants using the Etherpad Software, to capture whatever they wanted). After each breakout session, session notes were analyzed and summarized before being presented to all participants in the next summary and discussion session. This served both to validate whether the notes correctly reflected the conversations, as well as to focus the subsequent conversations on central ideas and takeaways.

FINDING 1: CHALLENGES FOR A SHARED UNDERSTANDING OF OPEN COMMUNITY HEALTH

The contexts and idiosyncrasies of various communities are reflected in different languages, structures, and ideals. These shape how each community thinks and acts on open community health. Such discrepancies make it challenging to develop a shared understanding of open community health. Although all open communities are different, F/OSS and Wikipedia are more similar to each other than to Citizen Science and will be grouped as such for this finding. The three most discussed challenges to reaching a shared understanding are, in no particular order: 1) the platform an open community uses, 2) an open community's focus on artifacts, and 3) an open community's governance model.

Challenge #1: Metrics from Open Community Platforms

Open communities require a collaboration platform to coordinate their work. These platforms are available through the Internet and are designed to support the specific mission of each community, allowing work to be done in the open. A single platform may be used by many communities, but a variety of different platforms may also be used for different tasks. This makes it difficult to reach a shared understanding of how to assess open community health. Trace data from these platforms can serve to inform open community health metrics, but due to the idiosyncratic use of platforms, the interpretation of metrics is varied at best and direct comparisons are even more problematic. Furthermore, standardized trace data is not available across all open community platforms.

F/OSS projects often use a software-sharing platform such as GitHub and GitLab as well as freely available tools like Git and Gerrit for versioning and editing. F/OSS communities often maintain communication channels through mailing lists or chat programs. Collecting trace data from F/OSS project platforms has been mostly solved and research on such data is widely accepted. Similar to F/OSS, Wikipedia has a shared platform for creating articles but differs in the availability of tools and platforms. Where F/OSS has many hosting and tooling options, Wikipedia provides both hosting for articles as well as tooling.

In contrast, Citizen Science lacks a unified shared platform or technology infrastructure for managing and performing project work. While platforms like Zooniverse and CitSci.org are designed to house a collection of projects, the citizen science work itself may be performed outside of the host site. Thus, many Citizen Science projects must develop or source their platform and tools independently. This may be explained by the variety of tasks and designs that Citizen Science communities follow, each customized to specific research objectives. SciStarter 2.0 is a newer platform that has the potential to serve as a social platform for members

of Citizen Science communities, but again, the work is organized and facilitated through other customized tools. Systematic tracing of individual project contribution across a multitude of Citizen Science platforms is yet to be developed. SciStarter is attempting to close this gap by offering participant data to scientists through free and fee-based data tracking services, but there is little else that can provide this cross-platform participant data at this level of granularity.

Challenge #2: Open Community Artifacts

Open communities differ in how they focus on artifacts, and consequently how contributors engage in communities. If the mission of an open community is to design and advance an artifact then there is a stronger focus on artifacts, which makes it easier to define a community as including anyone who contributes to the artifacts. The interaction of contributors with an artifact is measurable and the resultant metrics may inform community health. For example, community leadership may identify the number of individual contributions to the construction of an artifact as an indicator of community health. However, when the community focus is less concrete and independent of a specific artifact, then this approach is not possible.

F/OSS and Wikipedia communities have a strong focus on technical artifacts. The goals of an F/OSS community are to create and advance software artifacts. Similarly, Wikipedia communities have goals around creating and improving a set of article artifacts. These artifacts act as a technological centering point to which the community engages. A Wikipedia contributor may focus attention on articles about cars because this is their area of expertise or interest. An F/OSS contributor may focus attention on container software as this is an area of interest for the contributor or their employer.

In contrast, contributors to Citizen Science projects tend to 'dabble' in several projects while simultaneously contributing to one or two core projects based on their interests. Here, contributions

| 4

are not aimed at particular technical artifacts but instead are focused on the data artifacts that are most valuable for Citizen Science projects. Additionally, unlike F/OSS and Wikipedia, Citizen Science projects have long- to short-term ranges where the projects have determinate ends. However, the conclusion of a project does not always align with the conclusion of users' interest, who might want to participate beyond the project completion date. As some Citizen Science projects do not always have a clearly documented point of completion, participants may not necessarily know that the goals of a project have been reached. This phenomenon can lead to participation drop off and contributor fatigue – if they can't tell that their work means something for one project, they may not be willing to invest the time in contributing to another project. Therefore, participants should be provided with a means of discovering project outcomes, since moving users to a different project is an important means of engagement and a very real concern for research teams. In summary, Citizen Science communities are less bounded and more ephemeral due to their weaker focus on specific artifacts.

With a different artifact focus (i.e., technical and data artifacts), how open community health is determined will likely fall to different aspects of the project. In Citizen Science projects, health may be part and parcel of the quality of citizen data, while in F/OSS projects, health may be part and parcel of the resolution of software bugs. The variable focus leads one to understand open community health as related to the core practices associated with the particular artifact in question.

Challenge #3: Open Community Governance

Open communities often have governing mechanisms in place that may control or define how participants work together. An important aspect of these governance mechanisms is documentation of the function of community leadership and the path to leadership. Because all communities are different, measuring the health of a community may be a

unique exercise for each community. Understanding open community health requires understanding a community's governance, because it defines acceptable behavior within a community and what may be warning signs of health failures. One such warning sign could be that the path to leadership is blocked and no new leaders emerge from the community to carry forward the mission of the community.

Governance in F/OSS and Wikipedia communities is complex, with a mix of charismatic, participatory and bureaucratic forms. Many F/OSS projects have technical steering committees or may be hosted by a foundation. But these entities often have little authority and exist merely to provide general guidance or support. Alternately, some projects have very strong governing mechanisms in place. For example, Debian[8] has a constitution and elects a project leader every year. Although not always documented, work in F/OSS is usually self-assigned by contributors and there is often a path to leadership. As such, contributors can become leaders by earning a good reputation through contributing and participating. In Wikipedia, this path to leadership is semi-automated and partially consensus driven. Basic rights are earned through meeting participation thresholds — anyone who has contributed for four consecutive days, made 10 edits, and has not been banned, is granted the right to create articles. Advanced rights are earned via a community review of past activities and merit — Wikipedia's administrators who can delete pages and ban users are promoted through a nomination and consensus discussion process. The governing rules (policies and guidelines) are built through an emergent, conversational cycle of rule interpretation and reformation, ideally open to all contributors, and ratified by a consensus discussion. In theory, rules are enforced by any contributor through an open debate that includes direct citation to rules and consensus discussions about which rules are applicable in a specific situation.

Governance mechanisms in Citizen Science communities are more exclusive and restricted.

Authority structures are often easy to identify; and there is a strong differentiation between project setup and project execution. Citizen Science projects are structured and have a strong separation between science teams (designing, overseeing, and publishing research) and contributors (collecting and processing data). Tasks are serialized and split up to allow many different contributors to do work with less training than is needed for researchers, thus further reinforcing the separation. Science teams set the research agenda and much of the governance is conducted in back channels versus out in the open. Because there is a stronger distinction between leadership and contributors, paths to leadership are often non-existent (i.e., advancing from a contributor to a research team member is near impossible in most projects). Some Zooniverse volunteers have become Moderators on Talk boards, with research teams giving them a modicum of responsibility within their organization, but an elevation to becoming part of the research team is rare.

FINDING 2: STARTING POINT FOR A SHARED UNDERSTANDING OF OPEN COMMUNITY HEALTH

A goal of the workshop was to develop transferable insights into open community health that applied to F/OSS, Wikipedia, and Citizen Science. Despite the previously introduced challenges, participants in the workshop identified overlap in how they understand open community health. One characteristic all open communities found to have in common is that they are founded on people who collaborate to pursue the mission of the community. Although open community health may be measured in many ways, focus on people is a good starting point for developing a shared understanding of open community health across community types. We present three areas of overlap and offer three common Goal-Question-Metric compilations that were relevant to all three open community types.

Recruitment, Retention, and Engagement

The presence of active members is a major concern for open communities because projects only exist when they have members who collaborate in pursuit of a community mission. A focus of open communities is on developing and nurturing productive communities with diverse, inclusive, welcoming, and active membership. Three activities directly related to this are recruitment, retention, and engagement. The associated Goal-Question-Metric compilations are:

Goal: Recruit new contributors to an open community.

Question: How many new contributors participated in the community over a given period of time?

Metric: Number of new contributors over a given period of time.

Goal: Retain motivated new contributors.

Question: What type of experiences are important for new contributors?

Metric: Successful second contribution attempts and how long a contributor stays with project analyzed by what their first interaction was.

Goal: Engage online communities by having offline meetups.

Question: How does talking to someone "offline" affect willingness to participate "online"?

Metric: Attendance at activities, attendance over time, and variation in meet-ups in cities.

On-boarding: Mentoring and Training

Open communities have active members who want to contribute to the mission of the community. A shared concern is to enable these members to contribute in meaningful ways because the strength of open communities lies in the shared interest and knowledge of the contributors. Domain-specific training and communication of community

expectations ensure that the community has a shared roadmap for project outcomes. Activities directly relating to ensure that members are socialized, empowered, and skilled to contribute are integrating new contributors to the community through on-boarding processes to include mentoring and training. The associated Goal-Question-Metric compilations are:

Goal: Make it easier for new contributors to participate in the community.

Question: Are newcomers able to get engaged and contribute to the community in a reasonable timeframe?

Metrics: Availability of on-boarding materials (e.g., wiki, documentation, FAQ), mentorship programs, and places where people can post questions (e.g. Askbot, mailing list, live chat).

Goal: Train people.

Question: How do you know that your training works?

Metric: Skill level of community members.

Goals: Mentoring across communities by sharing best practices and avoiding duplication of effort.

Questions: How much redundancy do we see within communities? Which best practices have been adopted and by whom?

Metrics: Reuse of practices, research protocols, or source code. Number of similar communities.

Diversity and Inclusion

Creating a diverse and inclusive environment that is conducive for a wide variety of potential members to feel welcome and empowered to contribute is vital to open community health. For example, diverse communities produce innovative project outcomes and often foster more volunteer-driven collaborations. Further, to maintain activity,

contributors need to have a desire to be involved, which is negatively impacted by toxicity or exclusive behavior. Thus, diversity and inclusion are key factors in recruiting and retaining contributors to open communities. The associated Goal-Question-Metric compilations are:

Goal: Project meets the needs of the communities and technologies that depend on it.

Question: In F/OSS, are downstream projects that depend on this project able to use it productively? In Citizen Science, how do science teams integrate Citizen Science work into their scientific practices?

Metric: Changes in dependency patterns. In F/OSS, the number of forks maintained by other projects, i.e. number of people using the same software. In Citizen Science, the number of articles being published based on citizen science data/analysis.

Goal: Engage new and diverse audiences.

Question: How many new and diverse participants are attracted?

Metrics: Number of new diverse segments engaged over time, annually, and since community inception.

Goal: Encourage and measure pathways between disparate projects.

Question: How does a Zooniverse volunteer, citizen scientist, or an F/OSS developer who participates in multiple projects differ from one who specializes in just one?

Metric: Construct profiles based on projects, actions, and timestamps at the person-level.

FINDING 3: TAXONOMY FOR A SHARED UNDERSTANDING OF OPEN COMMUNITY HEALTH

Citizen Science, F/OSS, and Wikipedia communities show great potential for cross-domain knowledge transfer, but major efforts in this direction have not

yet been undertaken with regard to open community health. One particular exercise that elicited much enthusiasm and interest from workshop participants was the development of a taxonomy, or term-mapping, between the communities to facilitate understanding and learning between communities. The results from this early effort can be found in the Appendix. Such a taxonomy is a necessary first step in clarifying what each community means by the terms they employ (Eitzel, et al., 2017), thus helping to bridge the differences between open communities and help to deepen shared understanding of open community health.

Continuing the conversation

The workshop was successful in bringing together representatives from Citizen Science, F/OSS, and Wikipedia communities. It was also successful in fostering a dialogue among the communities, with the key takeaways presented here. Additionally, continuing the dialogue and working across communities would entail benefits for all involved.

The Goal-Question-Metrics approach was effective in facilitating the workshop and could be a good structure for advancing our understanding of open community health. Challenges that need to be overcome include disparate languages, varying artifacts of focus, and a shared understanding of metrics and strategies. Future work can complete the taxonomy for a shared understanding which will help bridge language differences between open communities. Developing tools for determining open community health might not work across the different platforms and varying artifacts. However, each community can implement their own tools but a shared understanding of metrics and strategies can be rooted in implementation-agnostic metrics definitions and a collection of best practices. Sharing experience reports of overcoming open community health issues and fostering healthy communities with lessons learned are beneficial to all.

ACKNOWLEDGEMENTS

REFERENCES

Basili, V. R. (1992). *Software modeling and measurement: the Goal/Question/Metric paradigm*. Retrieved from https://drum.lib.umd.edu/bitstream/handle/1903/7538/?sequence=1

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., ... Jiang, Q. (2017). Citizen science terminology matters: exploring key terms. *Citizen Science: Theory and Practice*, *2*(1). https://doi.org/10.5334/cstp.96

APPENDIX: TAXONOMY MAPPING FOR A SHARED UNDERSTANDING OF OPEN COMMUNITY HEALTH

This document contains an early effort into taxonomy mapping, based on an Open Community Health workshop held in April 2018. It would greatly benefit from further discussion and should therefore be viewed as a starting point to initiate further conversations. Empty cells may indicate a missing equivalent concept or that we did not know which word to put there. As such, the taxonomy of the terms, their meanings, and their cross-domain counterparts presented are a work-in-progress, and can be freely added onto and modified by anyone interested in picking it up.

Terms can be grouped into two major categories: *People* and *Infrastructure*. However, more categories can be added. Due to the centrality of people and the many different functions, capabilities, motivations and contributions that exist among those involved, the *People* grouping involves anyone who is actively involved with the community. Open communities are largely built on and supported by various forms of online and offline systems and

JOURNAL OF
PEER PRODUCTION

The Journal of Peer Production
New perspectives on the implications of peer production for social change
Journal of Peer Production Issue 13: OPEN
http://peerproduction.net — ISSN 2213-5316

tools. Terms to talk about these platforms have been grouped under *Infrastructure*.

INFRASTRUCTURE TAXONOMY

Generic Description	Community Terminology		
	Citizen Science	Open Source	Wikipedia
Platform infrastructure	Zooniverse	GitHub, Gerrit, TravisCI, Jenkins, wiki's, JIRA, etc.	MediaWiki, Huggle, ORES
Project infrastructure	Protocol, Datasheet	Repository	Bots & test script
Collection/artifact /codebase	Dataset	Repository, Codebase	Content namespace, Articles, Items/ Properties/ Entities
Platform infrastructure	Dataset	Collection	Wikipedia page
n/a	Datasheet	Artifact	n/a
n/a	Data entry	Codebase	n/a

PEOPLE TAXONOMY

Generic Description	Community Terminology		
	Citizen Science	Open Source	Wikipedia
Someone who takes the initiative or the lead on a project	Lead Scientist, Project Owner, Project Champion	Founder, Benevolent Dictator for Life, Sponsor	Project Organizer (not a title that can be owned), Bot Maintainer
Person initiating or starting a project and maintains governance of project	Project Leader	Founder, Benevolent Dictator, Project Lead	N/A Note: Wikipedia doesn't have this because you can't "own" a page
Someone who works on the organizing side of the project. Organizational contributor	Science Team, Project Staff. Note: Within Citizen Science there is some pushback against "owner"	Community Manager, Evangelist	WikiProject organizer
Someone with rights or a role in which they can make decisions	Moderator, Reviewer, Core Team, Science, Project Staff, Platform Team	Committers, Maintainer, Project leads, Core Contributor, Technical Steering Committee, Advisory Board, Governing Board	Administrator, Steward, Mediator, Closer, Consensus-processes, Architecture Committee, Staff Engineers, Product Managers
Editorial Contributor	Volunteer, Participant	Developer, Contributor	Editor
Community organizer / Cheerleader. Animateur/Animatrice	Ambassador, Super Volunteer	Community Organizers, Cheerleaders, Ambassador	Wikipedian-in-residence
Someone with a more evaluative role	Reviewer	Project Lead, Reviewer	Stewards
Someone who is new to the community	Inexperienced	Inexperienced	Non-autoconfirmed
Someone with scientific interests and contributions	Science Team	Core Team	Mediators
Someone who oversees things in an organizational manner	Maintainer	Technical Steering Committee	Discussion Closer
Someone who works on the technical side of a project	Core Team	Developer	Architecture Committee
Someone who supports the technical infrastructure of a project	Unicorns? (no term given because they are very rare but much wanted)	Release Engineers, Dev-ops	Tool Developer

ENDNOTES

[1] https://chaoss.community/

[2] https://wiki.mozilla.org/Diversity_and_Inclusion_for_Communities_and_Contributors/DI_Call_05_02_2018

[3] https://scistarter.com/

[4] https://www.zooniverse.org/talk

[5] https://meta.wikimedia.org/wiki/Research:VisualEditor%27s_effect_on_newly_registered_editors/May_2015_study

[6] https://doi.org/10.1145/3233391.3233544

[7] https://doi.org/10.1145/2685553.2699022

[8] https://www.debian.org/devel/constitution