📶 (https://www.software.ac.uk/rss) 🐦 (https://twitter.com
/SoftwareSaved)

Blog (/blog)   News (/news)   Events (/events)   Contact (/contact)   Subscribe (/subscribe)

Software
Sustainability
Institute

(/)

About (/about)

Programmes and Events (/programmes-and-events)

Resources (/resources)

search

Open Community Metrics and Privacy: MozFest18 Recap

Posted by s.aragon on 12 November 2018 - 9:19am

By Raniere Silva, Software Sustainability Institute, and Georg Link, University of Nebraska at Omaha.

This post was simultaneously published here and in the CHAOSS Community blog (https://chaoss.community/blog-post/2018/11/08/open-community-metrics-and-privacy-mozfest18-recap/).

Image courtesy of Raniere Silva

TL;DR

Open communities lack a shared language to talk about metrics and share best practices. Metrics are aggregate information that summarise raw data into a single number, stripping away any context of data. Pedagogical metric displays are an idea for metrics that include an explanation and educates the user on how to interpret the metric. Metrics are inherently biased and can

lead to discrimination. Many problems brought up during the MozFest session are worked on in the CHAOSS project.

Introduction

Mozilla Festival (https://mozillafestival.org/), or MozFest, is an annual event organised by **Mozilla (https://www.mozilla.org/en-GB/)** volunteers to join thinkers and champions of the open internet. The event is located at the Ravensbourne University, Greenwich, London, United Kingdom, and hosts a safe space for attendees to talk about decentralisation, digital inclusion, openness, privacy, security, web literacy and much more.

At MozFest this year, we organised a 90-minute conversation on the topic of open community metrics and privacy. We seeded the conversation by sharing past examples (e.g., we are 1,000,000 mozillians) asked our nine participants to discuss what it means to aggregate metrics for open communities and what impact this can have for (non-)members of these communities. The following sections elaborate on thoughts, ideas, and discussions from the session.

What are Open Community Metrics?

The MozFest conversation revealed that we are lacking a shared language for talking about open community metrics. Our session is indicative that the problem exists across open communities. Maybe it is because open community metrics have not been standardised, have not been around for long yet, or that every community has different ways of dealing with them. To resolve this, we discussed the meaning of what is an open community and what are metrics.

We compared community to networks. A network consists of people who are connected to each other. In contrast, members of a community have a shared goal and purpose, often creating something together. Communities build a shared identity, and membership in the community is an alignment of one's identity. Adding in the 'open', an open community is one that uses communication tools that displays the conversations to anyone who wants to inspect them. An open community does not always welcome everyone as a member (inclusive governance) but it operates in a transparent way.

We contrasted metrics and data. Data is the raw information about who wrote a message to whom and what was the content of the message. The data needs to be enriched before it becomes information, such as how active a specific user is or how

popular a certain topic is within the open community. Metrics are aggregates of those information and summarise the information into a single number or graph. Metrics is the information you would expect to see on a dashboard that summarises aspects of your open community.

Open Community Data Example

The following example from a fictional open source project demonstrates how much information about participants in open communities is often available online. Such data can be aggregated by software, such as **Parceval (https://github.com/chaoss /grimoirelab-perceval)** that retrieves and gathers data from different places where open source contributors interact, for example GitHub, Mailinglists, Discourse, Slack, and Twitter. Due to the open nature of the community, much personal data is made freely available. For example, a Git commit retrieved by Parceval could be:

commit 944713e38290d8dd4a9ac7f02267ada72a0e5a10

Author: Jane Doe <jane.doe@mail.com>

Date: Mon Oct 29 14:56:59 2018 +0000

* Adicionado arquivo README.md*

The commit contains Jane's full name and email address. Even if Jane was a very private person and, except for their name and email, no other information about them was available online, we can narrow the search for the real person based on the Portuguese commit message, timestamp, and timezone. With access to other databases, we could enhance Jane's profile more and with a higher degree of accuracy and precision.

How can we protect Jane's privacy when open source communities are collecting, processing, aggregating and analysing data like the one that we describe? And how open source communities can stay true to their principles by making the data used on the analyse open (or FAIR – findability, accessibility, interoperability, and reusability) without violating Jane's privacy?

Sharing how Metrics Should be Understood

Now that we have a better shared understanding of what metrics are, we can discuss the intricacies of understanding the meaning behind metrics. A problem with metrics is that they only provide a single number or graph with no background information on how the data was collected, how it was summarised, and what were the intentions of the creators

of a metric. This can lead to misunderstandings. For example, when number-of-messages is used as a metric to gauge popularity in a topic, then we need to understand how conventions unfold in an open community and the tool used. A tool that has thumb-up reactions or votings for messages provides a way to show interest in a topic and interact with messages; whereas the count of messages in a tool that is lacking those features and only has text messages (e.g. mailing lists) will be inflated because it counts simple "+1" messages.

Another problem with metrics is that they are inherently biased. This runs counter to a popular belief that numbers are objective. At the foundation of metrics are data, which is biased because someone decided what data to collect and how to store it. The bias in data limits what information can be summarised by metrics. Furthermore, a metric reduces data to a single number or graph and as such removes detail and context. In this process, some data is weighted more than other data or not included at all. How data is summarised is determined by the metric creator, though what is done to the data is not obviously clear to a metric user.

These problems point to a need of bridging between metric creator and metric user. A solution we discussed at MozFest is to provide an explanation with every metric. The person or team that prepares a metric for an open community would be responsible for including a description to help with interpreting the metric. We might call this concept of a metric that includes an explanation and educates the user on its use a 'pedagogical metric display'.

Metrics by Community vs. Metrics about Community

Open community metrics can be created by open communities about themselves, or by externals. Metrics created by people external to an open community have issues that stem from their external perspective. For example, they may not know the meaning behind interaction patterns of a community and as such may choose to summarise data that is not aligned with how the data came to be. Externals also might not be familiar with the goals and identity of an open community and thus create metrics that are misaligned. To resolve these issues, metrics might be better created by open communities. Pedagogical metric displays could bridge between internal metric creators and external metric users.

An interesting question is why metrics are created for open communities. A small open community typically does not have metrics and we only observe metrics being introduced in large and growing communities. An intuitive answer could be that in a

small community it is easier to maintain a sense of how the community is doing but in a larger community it becomes infeasible to keep track of everything at the interaction or data level and thus people aggregate the data to metrics to maintain this sense of knowing how a community is doing. A follow-on question is whether communities use metrics to understand themselves better, or whether communities use metrics to drive their growth. For example, knowing and ensuring that new users who make their first post in a community forum receive friendly welcome replies can help a community thrive. Conversely, a daily notification to maintainers that sums up how many new posters did not receive a reply within 48 hours is in itself not helpful because it only provides a negative metric without actionable insights.

A question raised at MozFest is about creating and sharing metrics. By this, we mean the process of calculating a metric. For example, one open community might come up with a metric that serves the community well and other communities are interested in using the same metric for themselves. The immediate practical question is how metrics can be transferred between contexts and adapted for different tools that these communities might be using. An ideological question is whether the metric will actually be useful to the second community.

Who is (Dis)Advantaged by Open Community Metrics?

This section explores the implications of privacy in the context of open community metrics. First, anyone who participates in an open community can be tracked and profiled. A user may try to maintain a sense of privacy by using a pseudonym, a unique email address for each open community, and avoid revealing personal information. How successful these approaches are is subject to further study, but examples exist for failing to maintain privacy in online spaces.

Another problem with open community metrics is how they are being used. A funding agency or other benefactor might want to see open community metrics that demonstrate impact or outcomes. This seems like a simple case where an open community that is tracking metrics and can show how it is producing outcomes or making an impact has a leg up in the struggle for financial support. Producing metrics could become a matter of making an open community sustainable.

Other uses of metrics are also possible. An open community can use metrics to identify areas where community members are struggling. Identifying and fixing a bug in a community platform improves the experience. However, a problematic use of metrics could come from outside a community. What if a community of runners, who share their

tracked routes, heart rates, best times, and who they ran with, was analysed by an insurance company to identify new ways to structure insurance policies for people who run regularly (and consequently discriminates against parts of the community).

We have heard that membership in some open communities is a differentiating factor for job applicants. A developer with an active GitHub or StackOverflow profile could have a higher chance of landing a new job. Conversely, a user who posts pictures of drink parties in an open community could be disadvantaged in the hiring process. Granted, these are less issues of open community metrics and more about the data of open communities being available to external people. However, an open community can aid such discriminating use cases by providing metrics at the individual person level (ever wonder who looks at the green dots on GitHub and for what purposes?). The goal of metrics on the level of the individual might be to incentivise members of a community to participate more and behave in certain ways (e.g., avoid negative brownie points). But those same metrics could become signals and be used in unexpected ways. Naturally, people will begin to game a metric when it is displayed.

With the assumption that open community metrics are used in decision making (e.g., hiring, insurance premiums, credit reporting, national citizen score, …) the people who do not participate in a community might be naturally disadvantaged. Some people may choose to avoid open communities and stay anonymous. However, some people are lacking internet access, do not speak (fluently) the dominant language in their community, or do not have the time besides job and family to participate. These people are disadvantaged for reasons outside of the open community and the use of metrics is unfair from the onset.

Conclusion

The MozFest session was an exchange. We know from the participants that each of them took away something from the conversation. The central discussion points are summarised here. As facilitators, we see MozFest to be a mirror of the conversation that is happening in **CHAOSS (https://chaoss.community/)**. With this blog post, we hoped to document some of the discussion and provide a baseline from which we can move forward. Many of the problems brought up in the MozFest session and documented here are problems that the CHAOSS project are working on (I know, shameless plug for CHAOSS).

Thanks!

We are thankful to the attendees that help us to brainstorm open community metrics

and privacy.

If you wish to discuss this post with us, **send us an email (https://software.ac.uk /info%40software.ac.uk)***or contact us on* **Twitter (https://software.ac.uk /www.twitter.com/SoftwareSaved)** *@SoftwareSaved.*

Tags

- Raniere Silva (/author/raniere-silva)
- Georg Link (/author/georg-link)
- Open Community Metrics (/tags/open-community-metrics)
- MozFest (/tags/mozfest)
- Community (/taxonomy/term/3)

Subscribe

Every Friday we send out a digest of the week's news and blog posts.

Email Address   **Subscribe**

Terms (/terms)   Privacy Policy (/privacy)   Cookie Policy (/cookies)