

# **UNICEF Innovation Fund**

## **Measuring the health of Open-Source Project – A Proposal**

Name: **Anusha Chandrakanth Lingadahalli**

Email: [lingadahalli.a@northeastern.edu](mailto:lingadahalli.a@northeastern.edu)

## ABSTRACT

This is a proposal to build some metrics to measure the health of an open-source project of a start-up in the Philippines as part of the UNICEF Innovation Fund. This start-up maps hard-to-reach areas and connects communities to resources using artificial geospatial analysis.

## DISCUSSION

A healthy open-source project is the one which uses open infrastructure, open practices, encourages community building and attains sustainability. But measurement of health, like in any other complex ecosystems is a difficult venture. There are multiple factors which contribute to the success (or the failure) of a project. Hence, it is important to breakdown the factors and address them as needed.

The simplest way to measure the health would be to see the latest commit date (as any developer would suggest) and to see the number of contributors. Although this is simple, it does accomplish the job when you want to compare two projects which do similar tasks. Well, this one does not help in case of our situation, since we have 1 full-time developer, 1 part-time developer and 1 full-time project manager.

Considering this is a new project and we would want to measure the health over a course of 3 months, here are some of the other ways which we can consider.

### ❖ **Metric: Social media**

- Social media shares and mentions can be a very good indicator of popularity which is always an important factor for an open-source project.
- Provided the project is useful, more popularity would attract more people and may also result growth in funding and the number of contributors.
- This one could be measured on a weekly basis. A graph of count against each of the weeks would provide a good indication to assess the growth of this project.
- There are Natural Language Processing tools available which can crawl the web to get this data.

### ❖ **Metric: Population structure**

- Any open-source project population will have coding (C)/non-coding (NC) contributors and experience(E)/new(NE) contributors. These are very useful factors to measure the growth.
- We can define a term to measure the ratio of these 2 indicators.
  - For coding/non-coding contributors, Coding Contributors Ratio  $CCR = |C - NC| / \text{Highest value between C and NC}$
  - For New Contributors Ratio  $NCR = |E - NE| / \text{Highest value between E and NE}$
- CCR and NCR values would range from -1 to 1. A value closer to 0 would indicate equal distribution population, respectively. These ratios seem very useful to measure the health in the long term.
- This approach and its applications have been discussed in detail [1] and it sounds very promising.
- This would be a monthly metric and more suited for long term analysis. The open-source project population is not expected to change drastically on a weekly basis; hence a monthly report would be a better approach.

#### ❖ **Metric: Pull requests**

- Number of pull requests merged in a week/month could be a good indicator of work done. Although it does not provide a full extent of work done, it certainly provides an overview of progress.
- This can be weekly/monthly report.

#### ❖ **Metric: Bug fixes and Issues reported**

- Number of issues reported could indicate the usage of the project. No systems are perfect to begin with, certainly not in this project with limited resources. As more people start to use the project, the issues and bugs reported could be good factor for growth.
- Higher count would be healthy for the system in the long term as it will only increase the robustness in the system.
- This report could be on a weekly/monthly basis depending on how popular the idea of this start-up has been.

## CONCLUSION

As discussed here, there is no one-size fits all approach when it comes to measurement of health in an open-source project. But there are various factors, which when studied in combination with each other, can provide valuable insights into the system.

## REFERENCES

- [1] Onoue, Saya & Hata, Hideaki & Monden, Akito & MATSUMOTO, Kenichi. (2016). Investigating and Projecting Population Structures in Open Source Software Projects: A Case Study of Projects in GitHub. IEICE Transactions on Information and Systems. E99.D. 1304-1315. 10.1587/transinf.2015EDP7363.
- [2] Kazuhiro Yamashita, Shane McIntosh, Yasutaka Kamei, and Naoyasu Ubayashi. 2014. Magnet or sticky? an OSS project-by-project typology. In Proceedings of the 11th Working Conference on Mining Software Repositories (MSR 2014). Association for Computing Machinery, New York, NY, USA, 344–347. DOI:<https://doi.org/10.1145/2597073.2597116>
- [3] <https://softwareengineeringdaily.com/2019/04/04/project-health-assessing-open-source-projects-with-more-context/>
- [4] <https://opensource.com/article/19/8/measure-project>
- [5] <https://www.linuxfoundation.org/en/resources/open-source-guides/starting-an-open-source-project/>