

Understanding What Constitutes a Vibrant Open-Source Community

Many academic and public libraries build or make use of open source projects to provide affordable and sustainable tools for the community built by the community. But how does an open source project become sustained? What does it take to build a growing and vibrant open source community? And which type of open source projects are safer to use or participate to?

To answer these questions, our team at Harvard's Institute for Quantitative Social Science (IQSS) proposes to run a two-year study that will analyze the characteristics and activities of 10 open source projects, which we believe to be successful, to determine a set of guidelines that help grow and maintain a healthy open source community. In this proposal, we request IMLS funding for a 12 month planning phase in the amount of \$50K. Collaborators from various open source communities will be part of the team to conduct the full study, but are not yet included in the planning phase.

Tasks in the 12 month planning phase

Aim 1. Define study design

During the planning phase, we will design the two-year study in detail. After conducting a literature review, we will define a set of attributes to measure the growth, success, and quality of an open source community. We will also identify a comprehensive set of activities that open source projects usually conduct. The study will not only include a quantitative approach (using metrics compared with the activities and governance of the community), but also a qualitative part (in the form of cases studies about each community)

Aim 2. Engage Partners

During the planning phase, we will define the full set of active open source projects that we plan to evaluate. We will start with the following projects:

Dataverse (<http://dataverse.org>)

Globus (<https://www.globus.org>)

Jupyter (<http://jupyter.org>)

Red Hat (<http://redhat.org>)

Open Nebula (<https://openebula.org/>)

We have selected these projects because they have matured, have been active and healthy at least for the last two years, and the IQSS is connected to them in some way, which makes it more feasible to obtain all the information that we will need to obtain for the study. The Dataverse project is being

The Institute for Quantitative Social Science, Harvard University developed, managed, maintained, and sustained by IQSS. We already have a relationship with the rest of the projects, and plan to establish a formal collaboration within the 12-month planning phase.

The projects above are not commonly used by libraries (although Dataverse and Jupyter have increasingly been used by academic libraries in the last 2-3 years). We plan to evaluate whether the attributes for success from these non-library open source projects are shared with open source projects more typically used or built by libraries. To do this, we will reach out to academic and public libraries (including our colleagues at the Harvard Library) to find 5 more open source projects that the libraries use, participate, or are familiar with, and they consider to be successful.

Aim 3. Host an experts workshop to review study design

We will convene a group of experts to review the study design, together with a selected members from the open source projects participating in the study. The workshop will take place in Month 6 of the planning phase, to ensure that we can incorporate the feedback into the final study design. The workshop will also serve as the place to review and agree among the partners (from collaborations in Aim 2) about the attributes used to define a successful open source community.

These 3 aims will inform us about how to best conduct the study in the next 2 years. In turn, the results of the study will inform libraries how to best engage an open source community to build and sustain digital tools for the *future library*.

Team

If this request is granted, the project will be led by PI Mercè Crosas with a research assistant (2.0 calendar months), and with the support of the Dataverse team at Harvard's Institute for Quantitative Social Science. The Dataverse team supports software and systems used to create repositories for research data sets, which are accessed by researchers and library personnel at academic, public, and professional institutions around the world. Dataverse is an open-source software project created by Harvard's Institute for Quantitative Social Science development team, who, in addition to Dataverse, has created numerous other open-source projects to enhance the experience of researchers both at Harvard and in collaboration with researchers around the world.

Budget

We request a total of \$50,000 for a 12-month planning grant. After the planning phase, to help us conduct the full study, we intend to request funding from IMLS and from the Sloan Foundation.

Planning Phase - Proposed Cost Breakdown:

PI, Merce Crosas: 0.5 calendar months, \$6,875 salary
Research Assistant- 2.0 calendar months, \$9,925 salary
Fringe Benefits (41.1%): \$6,905
Workshop: \$5,881
Indirect Costs (69% IDC on MTDC): \$20,414
Total Budget: \$50,000