

# Thesis Proposal

Institute for Software Research  
Societal Computing



## External Factors of Open Source Software Projects Sustainability

**Marat Valiev**

Friday, May 29th, 2020

1:00 PM - 3:30 PM EST

Virtual Presentation:

<https://cmu.zoom.us/j/99182957300>

Modern software development is heavily reliant on Open Source. It saves time and money, but, as any other non-commercial software, it comes on as-is basis. If not properly maintained or even abandoned by its community, Open Source Software (OSS) might impose extra costs or introduce bugs to the downstream projects. While software developers are well aware of these premises, existing techniques of mitigating such risks assume sustainability to be an intrinsic property of OSS, largely ignoring external factors. With plenty of examples of even high profile projects failing because of bugs in upstream dependencies, funding issues, lost competition or leaving key developers, it is important to understand the effect of these factors on OSS sustainability.

The results so far demonstrate that external factors play significant role in OSS projects sustainability. Shape of the dependency tree and access to resources from academic or commercial institution affect chances OSS chances of becoming dormant. While some of these issues, such as security vulnerabilities in upstream dependencies, recently became more visible on popular code hosting platforms, still there aren't enough signals for OSS project maintainers to get insights into the structure or size of the user community. As of today, platforms like GitHub provide only basic support for resource acquisition, usually limited to a donation page and lacking other resources a project might need, such as contributor recruitment, mentoring or infrastructure support.

In my thesis I propose to conduct several explorative studies to better understand role of external factors in OSS sustainability and develop techniques to include these factors into sustainability models. Using mixed method studies, we can shed light on the role of external dependencies, signals that can be used to inform maintainer decisions, ways to understand the structure of developer and user communities. Other factors to be explored in these studies include software adoption patterns, types of resources needed by projects and ways to obtain them, and how good practices of finding these resources. In addition to these mixed method studies, I also propose a methodological study on ways to produce embeddings for software libraries, projects and developer profiles that can be used to inform sustainability model design in many areas, including detection of software ecosystems, estimating supply and demand of developer skills etc. The resulting embeddings can then be used in mathematical models of OSS sustainability to identify external factors, such as presence of competing projects, size of the pool of potential contributors, and potential userbase.

**Committee: Dr. James Herbsleb (Co-Chair), Dr. Bogdan Vasilescu (Co-Chair),  
Dr. Audris Mockus (University of Tennessee), Dr. Vladimir Filkov (UC Davis)**