

Project Summary:

Coordination, Communication, and Collaboration in Open Source Software Development

Open source software runs the internet, dominates the server market, and is beginning to compete successfully in operating systems and even applications markets. Open source developments seem immune to the many problems that plague geographically distributed commercial development, as tens, hundreds, or thousands of developers collaborate successfully with minimal use of collaboration technology. The open source paradigm holds tremendous promise for supporting collaboration over distance in science and engineering, where software development, enhancement, and customization have become integral parts of the work. In fact, “open” style collaboration is being adopted in many and diverse settings other than software, from writing an encyclopedia to solving difficult engineering design problems.

Intellectual Merit: Theory-based empirical studies of coordination, social networks, and outcomes in open source software development.

Despite the interest in open source development, it is still not clear *how*, or even *if*, the open source paradigm has solved problems of coordinating technical work over distance. Open source projects have a distinctive set of practices, in which a core team or individual exercises control over what is incorporated in to the official code base; the vast majority of communication takes place in public, asynchronous forums; and individuals choose the work they will undertake. These practices are widely believed to help overcome coordination problems. Yet it may also be the case that open source succeeds by avoiding tasks and types of work where the need for coordination is high. So, for example, open source developments are generally focused on the relatively solitary tasks of maintaining and evolving the code, rather than the highly collaborative activities associated with determining requirements and generating a design.

We will investigate coordination in open source software development in the context of a theory of coordination. Building on our previous work in open source and in geographically distributed commercial development, we will first replicate and extend previous studies of commercial projects in order to determine if coordination in open source follows the same rules, or if it seems fundamentally different. We will perform comparisons of open source and commercial developments to see if theoretically predicted differences, driven by coordination needs, are observed. We will construct a computational model of open source projects to enable us to explore what-if analyses, and we’ll validate model predictions against actual observations. Finally, we’ll perform field experiments based on specific improvements we plan to introduce into selected open source developments as a further test of the theory

Research Plan

We will use archival data (change logs, version control logs) as well as surveys, interviews, and field experiments, to test hypotheses derived from this theory. We will initially use data from SourceForge, the largest collection of open source projects on the web, and the Apache server and associated projects. By virtue of our partnership with Avaya, we will also have access to similar data from a wide variety of commercial development projects in order to test our hypotheses about differences in processes and outcomes of open source and more traditional styles of development.

Broader Impacts

We seek a deep understanding of the principles that allow open source communities to coordinate technical work despite radical geographic separation. This holds substantial promise for improving how open source work is performed, especially in critical areas like science and engineering, to help make the vision of the NSF Cyberinfrastructure Report a reality. We anticipate that our results will also help to identify types of non-software tasks, settings, and organizational structures where open source style collaboration is likely to succeed. Especially as open source practices are applied broadly to all manner of distance collaboration projects, a deep, theory-based understanding will help guide effort into projects more likely pay off. We anticipate that the specific technologies we develop for our field experiments, which we will make available as open source, will be of benefit to many open source communities.