https://docs.google.com/forms/d/e/1FAIpQLSdGHbhTp1K9SBVKZnf1FGKQPV3uM1kf-fR7TPjZK72HnBIemw/viewform?vc=0&c=0&w=1&flr=0

Open Source Software in the Federal Government: An Analysis of Code.Gov

Gizem Korkmaz[1], Rahul Shrivastava[1], Anil Battalahalli[1], Ekaterina Levitskaya[2], José Bayoán Santiago Calderón[3], Ledia Guci[3], Carol Robbins[4]

[1] Westat
[2] Coleridge Initiative
[3] Bureau of Economic Analysis
[4] National Science Foundation

Open source software (OSS) is ubiquitous, serving as specialized applications nurtured by devoted user communities, and as digital infrastructure underlying platforms used by millions of people. OSS is developed, maintained, and extended through the contribution of independent developers as well as people from businesses, universities, government research institutions, and nonprofits. Despite its prevalence, the scope and impact of OSS are not currently well-measured. Recent policies of the U.S. Federal Government promote sharing of software code developed by or for the Federal Government. While the policy to promote reusing and sharing of software created with public funding is relatively new, public funding plays an important and not fully accounted role in the creation of OSS.

This paper aims to measure the scope and value of OSS development in the U.S. Federal Government. We collect data from Code.gov, the government's platform for sharing OSS projects, and study contributions of agencies. The dataset contains 17K repositories from 21 agencies, with the majority of contributions originating from the DOE, NASA and GSA. In addition, we collect data on development activity (e.g., lines of code, contributors) of the repositories on GitHub, the largest hosting facility worldwide. Adopting a cost estimation model from software engineering, we generate estimates of investment in OSS that are consistent with the U.S. national accounting methods used for measuring software investment. Finally, we generate and analyze collaboration network resulting from cross-agency contributions to repositories and explore the centrality of agencies in the network.