

Open source by the numbers at Google

At Google, open source is at the core of our infrastructure, processes, and culture. As such, participation in these communities is vital to our productivity. Within OSPO (Open Source Programs Office), our [mission](#) is to bring the value of open source to Google and the resources of Google to open source. To ensure our actions match our commitment, in this post we will explore a variety of metrics intended to increase context, transparency, and accountability across all of the communities we engage with.

Why we contribute: Open source has become a pervasive component in modern software development, and Google is no exception. We use thousands of open source projects across our internal infrastructure and products. As participants in the ecosystem, our intentions are twofold: give back to the communities we depend on as well as expand support for open source overall. We firmly believe in open source and its ability to bring together users, contributors, and companies alike to deliver better software.

The majority of Google's open source work is done within one of two hosting platforms: GitHub and [git-on-borg](#), Google's production Git service which integrates with [Gerrit](#) for code review and access control. While we also allow individual usage of Bitbucket, GitLab, Launchpad, and other platforms, this analysis will focus on GitHub and git-on-borg. We will continue to explore how best to incorporate activity across additional channels.

A little context about the numbers you'll read below:

- **Business and personal:** While git-on-borg hosts both internal and external Google created repos, GitHub is a mixture of Google projects, experimental efforts and personal projects created by Googlers.
- **Driven by humans:** We have created many automated bots and systems that can propose changes on both hosting platforms. We have intentionally filtered these data to ensure we are only showing human initiated activities.
- **GitHub data:** We are using [GH Archive](#) as the primary source for GitHub data, which is currently available as a public dataset on [BigQuery](#). Google activity within GitHub is identified by self registered accounts, which we anticipate under reports actual usage as employees acclimate to our [policies](#).
- **Active counts:** Where possible, we will show 'active users' and 'active repositories' defined by logged activity within each specified timeframe (for GH archive data, that's any event type logged in the [public GitHub event stream](#)).

As numbers mean nothing without scale, let's start by defining our applicable community: In 2019, more than 9% of Alphabet's full time employees actively contributed to public repositories on git-on-borg and GitHub. While single digit, this percentage represents a portion of all full time Alphabet employees—from engineers to marketers to admins, across every business unit in Alphabet—and does not include those who contribute to open source projects outside of code. As our population has grown, so has our registered contributor base:



This chart shows the aggregate per year counts of Googlers active on public repositories hosted on GitHub and git-on-borg

PRX124

1:23:cv-00108

What we create: As mentioned above, our contributing population works across a variety of Google, personal, and external repositories. Over the years, Google has released thousands of [open source projects](#) (many of which span multiple repositories) and ~2,600 are still active. Today, Google hosts over 8,000 public repositories on GitHub and more than 1,000 public repositories on git-on-borg. Over the last five years, we have doubled the number of public repos, growing our footprint by an average of 25% per year.

What we work on: In addition to our own repositories, we contribute to a wide pool of external projects. In 2019, Googlers were active in over 70,000 repositories on GitHub, pushing commits and/or opening pull requests on over 40,000 repositories. Note that more than 75% of the repos with Googler-opened pull requests were outside of Google-managed organizations (on GitHub).



This chart shows per year counts of activities initiated by Googlers on GitHub

What we contribute: For contribution volume on GitHub, we chose to focus on [push events](#), opened, and merged [pull requests](#) instead of commits as this metric on its own is difficult to contextualize. Note that push events and pull requests typically include one or more commits per event. In 2019, Googlers created over 570,000 issues comments, opened over 150,000 pull requests, and created more than 36,000 push events on GitHub. Since 2015, we have doubled our annual counts of issues created and push events, and more than tripled the number of opened pull requests. Over the last five years, more than 80% of pull requests opened by Googlers have been closed and merged into active repositories.

How we spend our time: Combining these two classes of metrics—contributions and repos—provides context on how our contributors focus their time. On GitHub: in 2015, about 40% of our opened pull requests were concentrated in just 25 repositories. However, over the next four years, our activity became more distributed across a larger set of projects, with the top 25 repos claiming about 20% of opened pull requests in 2019. For us, this indicates a healthy expansion and diversification of interests, especially given that this activity represents both Google, as well as a community of contributors that happen to work at Google.



This chart splits the total per year counts of Googler created pull requests on GitHub by Top 25 repos vs the remainder ranked by number of opened pull requests per repo per year.

Open source contribution is about more than code

Every day, Google relies on the health and continuing availability of open source, and as such we actively invest in the security and sustainability of open source and its supply chain in three key areas:

- **Security:** In addition to building security projects like [OpenTitan](#) and [gVisor](#), Google's [OSS-Fuzz](#) project aims to help other projects identify programming errors in software. As of the end of 2019, OSS-Fuzz had over [250 projects](#) using the project, filed over [16,000 bugs](#), including [3,500 security vulnerabilities](#).
- **Community:** Open source projects depend on communities of diverse individuals. We are committed to improving community sustainability and growth with programs like [Google Summer of Code](#) and [Season of Docs](#). Over the last 15 years, about 15,000 students from over 105 countries have participated in Google Summer of Code, along with 25,000 mentors in more than 115 countries working on more than 680 open source projects.

- **Research:** At the end of 2019, Google invested [\\$1 million in open source research](#), partnering with [researchers at UVM](#), with the goal to deepen understanding of how people, teams and organizations thrive in technology-rich settings, especially in open-source projects and communities.

Learn more about our open source initiatives at opensource.google.

By Sophia Vargas – Researcher, Google Open Source Programs Office