

Colby Goettel

Reed

English 312

10 February 2014

Open Source More Software

GitHub is an on-line version control system for storing and maintaining code, and allowing others to build off your existing systems. It is not some small business idea that few people use, the big names are all present and accounted for including Google, Facebook, Yahoo, Microsoft, and all of their various projects. Because of this, it is not uncommon for developers, when applying for jobs, to put their GitHub account on their resume. Companies want to see what people do in their spare time, where their true passions lie, and GitHub is an excellent way for these companies to see how individuals spend their free time.

Because large and small companies, movers and shakers in IT, release their code open source, other developers are able to fork their projects and build off these existing systems. A common phrase in IT is that programming is 90% finding someone else's code that does close to what you need and 10% adapting it for your specific needs. Having code in public repositories allows others to gain insight into how systems are built and see a history of each project's commits (version control). It allows developers to fork projects and easily take off that first, difficult 90% of programming. This greatly increases the speed that products can be prototyped, tested, revised, delivered, and maintained. Encouraging developers to release more of their code open source hurries the production of robust code. It improves software standards, creates a more secure world, and decreases the number of people needed to properly build and maintain systems.

On-line repositories like GitHub allow projects to be built by many developers at once

(think Google Docs for programming). And when development of a certain project stops and others want to continue development, projects can be forked and built separately. GitHub also has a great feature called pull requests. In their own words: “Pull requests let you tell others about changes you’ve pushed to a GitHub repository. Once a pull request is sent, interested parties can review the set of changes, discuss potential modifications, and even push follow-up commits if necessary.” Code review is an essential part of building any system. A code review is done before a new version of a product is released and it involved checking for security holes, testing, and correcting any problems, all in conjunction with at least one other developer. Pull requests are a modern interpretation of code reviews that allow remote developers to co-develop.

When code is said to be robust, it is because it can handle major and minor functions of the systems to fail without bringing down the whole system. Robust code is security-minded and stable, it is not alpha or beta code that is not recommended for public consumption — it is the best of the best. It is what the general public expects their code to be. Having gone through rigorous testing, code becomes robust because bugs are fixed and users are able to test and give feedback.

But building robust code is increasingly difficult because of how quickly the IT landscape is changing. When developers can use others’ code and tests, it allows faster and better products to be delivered. There is no need for developers to re-invent the wheel. Why should software libraries be any different than open source repositories?

In October 2010, Instagram, a photo and video sharing service, launched. A year and a half later, Facebook offered \$1B in cash and stock options to purchase Instagram. At the point it was acquired, Instagram had thirteen employees and over thirty million customers. Contrast this example with the Affordable Care Act (ACA) web site that the federal government sunk billions of dollars into development, fifty-five government contractors (not fifty-five employees), and over

two years to build. And it didn't work on launch. These models show the ends of the spectrum. It shows what is actually required to build scalable products in today's IT world. When developers can use existing code instead of having to build every project from the ground up, then proven code is able to be used more often in production, in turn increasing the reliability and robustness of the software.

Open source code is code where the source code is open to the public. Simple as that. But just because code is open source doesn't mean that it has no protections against intellectual property infringements. There are numerous open source licenses designed to protect the rights of developers and their intellectual property. The biggest three of these licenses are the MIT license, the Apache license, and the GNU General Public License (GPL). The MIT license is short and sweet: it allows others to do whatever they want with your code, just as long as they attribute you for your work and don't hold you liable for any potential damages. This paper and its repository are open source on GitHub¹ and protected by the MIT license. The Apache license is quite similar, but also grants patent rights to the original developer. Apache, the most common web server in the world and the eponym of this license, is released under the Apache license. The GNU GPL requires the code and any derivative work to be released and available on the same terms. Large products like Linux and Git and Wordpress are released under the GNU GPL.

But this does not mean that entire projects should always be released as open source code. Certainly there is discretion in how much code should be released. Some projects, like Apache, Git, and Linux, are best released in their entirety. When corporations release products, it is justifiable to release portions of the project open source so that others can use their ideas and tested code to build better software.

¹<https://github.com/cgoettel/engl-312>

Some developers and small businesses fear that if their code is released open source that others will steal their ideas and code. Licenses legally protect open source code from maltreatment and organizations like the Electronic Frontier Federation (EFF) exist to protect the electronic rights of individuals from maltreatment by governments and corporations. They are an ally in fighting legal battles against oppressive regimes.

Another powerful ally in this fight are the use of private repositories. If developers are particularly concerned about the security of their ideas, private repositories can be created which allow for greater security with core elements of a product. But beyond core elements, the general recommendation is to release as much code as possible open source and licensed appropriately.²

With more products released open source and hosted in on-line repositories like GitHub or Bitbucket, developers have a central location to find code and build off existing systems. It is impossible to build off another's system if it is closed source. Open source is the only possible way for developers to accomplish this.

Having open source code and allowing others to co-develop enables safer, more robust code to be built. Production time is greatly decreased, staff requirements are lessened, and less budget and resources are needed for each project. Releasing code open source and allowing others to test and develop it is the same idea as using remote data centers controlled by other companies. It decreases the individual's need to maintain so many different parts and just concentrate on the task at hand. The entire business cycle is sped up because developers are able to work together to achieve a common goal. It is truly standing on the shoulders of giants.

²For more information, please see <http://choosealicense.com>