Undergraduate Thesis Prospectus

SitScape Internship Experience Reflection

(technical research project in Computer Science)

Establishing A Culture of Open Development: The Open Source Movement

(sociotechnical research project)

by

Ethan Buckner

October 27, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Ethan Buckner

Technical advisor: Rosanne Vrugtman, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

How do software organizations best promote the development of software as a whole?

Software is a unique industry due to its history and culture of shared innovation. Unlike other industries where intellectual property is closely guarded, using the work of others is frequent and expected with attribution made explicit in source code through library imports. This focus on building off the work of previous developers has led to a culture that balances an individual firm's need to turn a profit with the shared advancement of the software community. The open source movement has codified a system for open innovation and demonstrated its value through successful large scale projects like Linux and Firefox.

Machine learning has recently exemplified the power of open innovation to contribute to software advancement. The core libraries used for machine learning including Pandas, Numpy, Scikit-Learn, and TensorFlow are all open source projects (NumFOCUS, n.d.), providing the tools necessary for researchers, students, independent developers, and smaller firms to easily leverage cutting-edge technological improvements. TensorFlow, supported by Google, demonstrates the benefits of open sourcing research and development for large firms. Google benefited from of community driven enhancements, widespread adoption and standardization while contributing to the greater software community (Metz, 2015).

SitScape Internship Experience Reflection

What lesson of lasting professional value did I learn from my internship experience at SitScape?

Experiential option for CS 4991 with Rosanne Vrugtman advising. In my technical report, I will reflect on my experience as a developer in a small team working through agile cycles to produce a final tool. The report will contain commentary on development

methodologies, organizational communication, and the software tools used, such as Git, text embeddings, large language models, and OpenSearch.

Establishing A Culture of Open Development: The Open Source Movement

How have proponents of open source established it as a principle of software development?

The rapid adoption of personal computers combined with software's unique zero-cost reproducibility gave rise to norms emphasizing open innovation that cut against intellectual property practices in other industries. Early developers in the open source movement codified these norms into a series of software licenses. Today, many open source projects are widely used, including the Linux operating system and the Firefox web browser. Recently, license transitions by MongoDB and Elastic NV from open source licenses to the "fauxpen" sever side public license (SSPL) reflect disagreements within the software community about the role of large cloud providers like Amazon offering open source products as a service (Banon, 2021).

Much of the sociotechnical research examining open source software was published in the early 2000s, a few years after the widely publicized transition to open source by Netscape, establishing Mozilla (Netscape, 1998). This period also saw the founding of the Open Source Initiative and the definition and general adoption of the term "open source" (OSI, 2006). The highly cited book *The Success of Open Source* by Steven Weber, published in 2004, was pivotal in examining the open source movement through an economic lens. Weber argues that the open source movement is not an exception to economic principles, instead that the unique circumstances of software engineering make open innovation economically rational. Weber also describes a business strategy for a firm to earn a profit while publishing open source software, a model still used by businesses like Red Hat today. Open source development has also been

analyzed through alternative economic lenses including as a gift economy, which is particularly applicable to highly distributed projects worked on by many individual developers (Zeitlyn, 2003).

More recently, the move by some large software firms like Microsoft to open source some of their products prompted a case study by Kochhar et al. (2019) examining how these transitions worked to benefit the open sourcing company and the software community. This case study approach will be applicable when examining the reverse transition by MongoDB and Elastic NV.

Among participants in the open source community, the Free Software Foundation (FSF) is notable for its history and ideological bend. Founded in 1985 by Richard Stallman, the FSF was an early advocate for software sharing, inspired by Stallman's experiences at MIT and what he saw as the growing encroachment of proprietary software (Stallman, 1999). The FSF is responsible for maintaining the GNU project, a set of free programs usually bundled with the Linux kernel as the operating system GNU/Linux (FSF, n.d.-a). Critical to the FSF is Stallman's personal philosophy of computing freedom, summarized in his four essential freedoms: "The freedom to run the program as you wish, for any purpose", "The freedom to study how the program works, and change it so it does your computing as you wish", "The freedom to redistribute copies so you can help others", and "The freedom to distribute copies of your modified versions to others…" (FSF, n.d-b).

Due to this philosophical focus on personal freedom, Stallman and the FSF reject the term "open source" as overly pragmatic and business facing (Perens, 2007), preferring the term "free software" to emphasize the user's freedom rather than the open availability of source code (Stallman, n.d.). The "copyleft" provision of the GNU General Public License (GPL) issued by

the FSF exemplifies the distinction between free software and open source software. The GNU GPL is less permissive than other open source licenses, requiring all derivative works to also use the same or equivalent license, ensuring that free software will not be used to build proprietary software (FSF, 2022). This reflects the ideological focus on the expansion of free software rather than being as permissive as possible.

The Open Source Initiative (OSI) was founded in 1998 at a conference of developers in favor of software sharing after notable developments in the field (OSI, 2006). For profit companies like SUSE had shown that publishing free software with paid enterprise support was a viable business strategy (SUSE, n.d.), and Netscape had just made history by releasing the full source code of their Navigator browser to the public before the term "open source" had been coined (Netscape 1998). The founding members felt that a single label was needed to define the practice of software sharing, and that it must emphasize the business benefits of software sharing and community development. The decided on term of "open source" reflects this. According to OSI cofounder Bruce Perens, it is intentionally distinguished from the more political "free software" despite them being essentially the same, though this similarity is disputed by Stallman (Perens, 2007).

The role of the OSI is to maintain the official definition of "open source" and to determine which licenses fall under the open source umbrella (OSI, 2023). Having a central authority on this definition is intended to maintain its philosophy and permissively. The OSI notably rejected the new server side public license submitted by MongoDB as not open source and issued a press release decrying Elastic NV for deceptively abandoning open source when they switched to the SSPL (OSI Board of Directors, 2021). The OSI is widely supported by large tech firms including Amazon, Google, and Microsoft (OSI, n.d.).

Open source non-profits like the Linux and Mozilla foundations support many smaller open source projects (Perlow, 2020). In the case of Linux, Mozilla, and Apache, these foundations were established to support the development of an early open source project and grew to encompass many smaller projects. These early foundations were important for establishing open source norms by guiding these projects (Mozilla Foundation, n.d.). Other foundations like NumFOCUS are created to support many related projects to begin with. Open source foundations provide a single place for corporate users of open source software to contribute to many current projects and any future projects supported by the foundation (Linux Foundation, n.d.). Some foundations like Mozilla are also funded through taxable subsidiaries offering paid services (Mozilla Foundation, 2020).

Many for profit companies publish open source software and earn revenue without restricting source code. Often this is done using enterprise deal model described by Weber (2004). This usually involves customization, setup, and continuing support. Shifting revenue streams away from source code exclusivity allows the company to also benefit from community development. Linux distributions like SUSE and Red Hat are prominent examples because the benefits provided by these companies to enterprise are particularly applicable to operating systems (Red Hat, n.d.). Some companies will also earn revenue by offering their product as a service through hosting.

Cloud computing has grown massively recently, with most companies moving away from managing their own servers and instead relying on cloud providers like Microsoft Azure, Amazon Web Services, and Google Cloud. Cloud providers can take advantage of economies of scale to offer economical on-demand computing. Businesses using cloud providers need less IT staff, never have to be concerned with updates or issues and always have access to the state of

the art. In an extension of this offloading of concerns, cloud providers have shifted to a platform as a service (PaaS) model, which integrates many tools needed for development into a cloud suite (Microsoft, n.d.). This is convenient for businesses using the cloud provider, but it cuts out market opportunities for the companies that develop the included tools to offer them as a service.

MongoDB and Elastic NV caused controversy in the open source world by switching from open source licenses to the SSPL, which specifically disallows offering the product as a service. MongoDB was first to do this in October of 2018, stating in a press that the large cloud providers are threatening the open source community by offering open source products as a service and that shifting to the SSPL is a defensive move that protects open innovation (Horowitz, 2018). Elastic NV specifically called out Amazon for trademark infringement in addition to echoing these complaints (Banon 2021). The SSPL was rejected by the OSI as not open source, emblematic of the negative response from the greater open source community (OSI Board of Directors, 2021). Open source developers like Red Hat dropped support for MongoDB and any other software licensed under SSPL (Krazit, 2019). These moves were responded to by Amazon by forking both projects and maintaining parallel versions as open source projects (Meadows et al., 2021).

References

Banon, S. (2021). Amazon: NOT OK - why we had to change Elastic licensing (press release). Elastic NV. elastic.co/blog/why-license-change-aws

FSF (n.d.-a). Free Software Foundation. The Free Software Foundation (FSF) is a nonprofit with a worldwide mission to promote computer user freedom. We defend the rights of all software users. fsf.org/about

FSF (n.d.-b). Free Software Foundation. What is Free Software? gnu.org/philosophy/free-sw.en.html

FSF (2022). Free Software Foundation. What is Copyleft? gnu.org/licenses/copyleft.html

Horowitz, E. (2018). MongoDB now released under the Server Side Public License (press release). MongoDB. mongodb.com/blog/post/mongodb-now-released-under-the-server-side-public-license#:~:text=MongoDB%20has%20created%20a%20new,litigation%20over%20enforcing%20the%20AGPL.

Kochhar, Kalliamvakou, Nagappan, Zimmermann, & Bird. (2019). Moving from Closed to Open Source: Observations from Six Transitioned Projects to GitHub. *IEEE Transactions on Software Engineering*, 1–1. Available from Web of Science. doi.org/10.1109/TSE.2019.2937025

Krazit, T. (2019, Jan 16). MongoDB's licensing changes led Red Hat to drop the database from the latest version of its server OS. *GeekWire*. geekwire.com/2019/mongodbs-licensing-changes-led-red-hat-drop-database-latest-version-server-os

Linux Foundation. (n.d.). Members. linuxfoundation.org/about/members

Meadows, Graybill, Davis, & Shah (2021). *Introducing OpenSearch* (AWS Open Source Blog) (press release). Amazon. aws.amazon.com/blogs/opensource/introducing-opensearch

Metz, C. (2015, Nov 9). Google Just Open Sourced TensorFlow, Its Artificial Intelligence Engine. *Wired*. wired.com/2015/11/google-open-sources-its-artificial-intelligence-engine

Microsoft. (n.d.). What is PaaS? azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-paas#:~:text=Platform%20as%20a%20service%20(PaaS,%2C%20cloud%2Denabled%20enterprise%20applications.

Mozilla Foundation. (n.d.). The Mozilla Manifesto. mozilla.org/en-US/about/manifesto/details

Mozilla Foundation. (2020). Introducing Mozilla VPN. blog.mozilla.org/futurereleases/2020/06/18/introducing-firefox-private-network-vpns-official-product-the-mozilla-vpn

Netscape. (1998). Netscape Announces Mozilla.org, a Dedicated Team and Web Site Supporting Development of Free Client Source Code (press release). web.archive.org/web/20021004080737/wp.netscape.com/newsref/pr/newsrelease577.html

NumFOCUS. (n.d.). Sponsored Projects. numfocus.org/sponsored-projects

OSI (n.d.). Open Source Initiative. Our corporate supporters. opensource.org/sponsors

OSI (2006). Open Source Initiative. History of the OSI. opensource.org/history

OSI (2023). Open Source Initiative. Open Source Definition. opensource.org/osd

OSI Board of Directors. (2021). The SSPL is Not an Open Source License (press release). Open Source Initiative. blog.opensource.org/the-sspl-is-not-an-open-source-license

Perens, B. (2007, Aug 17). Review: Perens Speaks about Free Software in Copenhagen (forum post). Technocrat. web.archive.org/web/20080212193546/http:/technocrat.net/d/2007/8/17/25149

Perlow, J. (2020). The Linux Foundation: It's not just the Linux operating system (press release). Linux Foundation. linuxfoundation.org/blog/blog/the-linux-foundation-its-not-just-the-linux-operating-system

Red Hat. (n.d.). The Value of Red Hat. redhat.com/en/about/value-of-Red-Hat

Stallman, R. (n.d.). Why Open Source Misses the Point of Free Software. Free Software Foundation. gnu.org/philosophy/open-source-misses-the-point.html

Stallman, R. (1999). The GNU Project. Free Software Foundation. gnu.org/gnu/thegnuproject.html

SUSE. (n.d.). About SUSE. suse.com/company/history

Weber, S. (2004). *The Success of Open Source*. Harvard University Press. Available from JSTOR. doi.org/10.4159/9780674044999

Zeitlyn, D. (2003). Gift economies in the development of open source software: Anthropological reflections. *Research Policy*, *32*(7), 1287–1291. Available from Web of Science. doi.org/10.1016/S0048-7333(03)00053-2