# Fifteen Years of Open Source Software Evolution

Francis Bordeleau[1], Paulo Meirelles[2], and Alberto Sillitti[3]

[1] Ecole de Technologie Suprieure (ETS), Montreal, Canada
francis.bordeleau@cmind.io
[2] Federal University of Sao Paulo, Sao Paulo, Brazil
paulo.meirelles@unifesp.br
[3] Innopolis University, Innopolis, Russian Federation
a.sillitti@innopolis.ru

**Abstract.** The Open Source Software (OSS) ecosystem and community has evolved enormously from the first edition of the OSS conference that took place in Genoa (Italy) in 2005. Such evolution happened in every aspect of OSS including research, technology, and business pushing its adoption to an unpredictable scale. Nowadays, it is almost impossible for people not using OSS in every interaction they have with technology. This fact is a tremendous success for OSS but such evolution and adoption has not always followed the intended path and some relevant deviations have occurred during such long journey.
This paper provide an overview of the evolution of OSS in the three mentioned areas (research, technology, and business) highlighting the main aspects and identifying the current trends that will be the basis for its future evolution.

**Keywords:** OSS evolution · OSS research · OSS technology · OSS business.

## 1 Introduction

In an very famous article published in *The Wall Street Journal* in August 2011 titled "Why Software Is Eating The World" [1], Marc Andreessen provided a deep analysis of the IT industry and analyzed how value was moving out from hardware while it was increasing in software helping a new generation of companies and entrepreneurs to build new business at a fraction of the costs compared to the beginning of the century.

The identified trend is still in place and become even more complex with the development of synergies between hardware devices and software in many application areas (e.g., smart home assistants, autonomous vehicles, etc.) inspired by the Apple's iOS ecosystem.

However, what the article of the 2011 was not discussing is a similar evolution that was already happening inside the software domain. As the commoditization of hardware (also through virtualization and the introduction of cloud services such as Amazon AWS) reduced the costs for running the computing infrastructure, the extensive usage of quality OSS reduced the costs for running the basic

software infrastructure (e.g., operating system, database, web server, application server, etc.), the development environments (e.g., IDEs), and applications (e.g., office automation tools, browser, data analysis tools, etc.). This trend was and still is a powerful support to the creation of new startup companies that require a fraction of the budget compared to the early 2000s. However, such effect has not only positive aspects but also negative ones when we consider the entire software ecosystem enabling *free riders* taking advantage of the OSS community without giving back.

Moreover, with the development of IoT, Smart Cities, etc., the value is still moving upwards leaving even the software domain towards the data that actually enable a whole new category of applications that are software-based but not implementable with just software (e.g., applications based on machine learning algorithms).

The paper is organized as follows: Section 2 and subsections analyze the OSS evolution according to the different perspectives identified: research (Section 2.1), technology (Section 2.2), and business (Section 2.3); Section 3 discusses the current trends in the above mentioned areas; finally, Section 4 draws the conclusions and introduces possible future work.

## 2 Open Source Software Evolution

OSS has been evolved in the last 15 years from several perspectives that are partially related to each other. We have decided to focus on the following ones:

1. **Research:** how the research has evolved considering the problems investigated, the types of activities performed, and papers published.
2. **Technological:** how the technology has changed considering the popularity of the projects, how people contribute, and how the technology has been adopted by users.
3. **Business:** how business models have changed, which ones emerged and which ones has been dismissed.

Such perspectives are not the only ones that can be identified but they are just a starting point for a deeper investigation.

### 2.1 The Research Perspective

The research environment has changed dramatically in the last 15 years from several points of view:

– **Publication venues:** the number of papers published in top venues dealing with OSS was very limited. As an example, just look at the proceedings of the *International Conference of Software Engineering (ICSE)* in 2005 [11] and 2019 [12] for comparison. This could be related to the fact that results of research activities sponsored by companies were rarely released to the community and public sponsored research grants were not pushing researchers

to release their results other than through publications. Nowadays, most of the papers published in top venues deals with OSS since the research behind them is using OSS as a source of data (e.g., in terms of code, developers activity, etc.), use OSS as a tool for extracting and/or analyzing data (e.g., as in machine learning tools for building models), release tools as OSS, etc. Moreover, most companies have an OSS strategy that allows or even force researchers to release OSS and public sponsored research grants often require the release of results and tools as OSS.

- **Investigated topics:** the investigated topics were often related to the adoption of OSS in different contexts (e.g., in the public administrations), the migration from proprietary solutions to OSS [15] [13] [14], the quality of OSS from the point of view of the code produced and the development process [8] [9] [10], licensing aspects. Nowadays, the focus is more oriented to areas in which OSS enable research activities that are not possible in other environments such as the investigations related to the analysis of large code repositories and their evolution [2] [3] [4], the analysis of communities [5] [7], resource usage [6], security issues, etc. As a general trend, the focus now is more on the data and its analysis with machine learning approaches present almost everywhere.
- **Specialized venues:** the OSS conference series has evolved since OSS-related papers are now published everywhere and not in specialized venues anymore. OSS is now a first class citizen in any research-oriented conference or journal and not the exception. The current trend requires authors to release their code and datasets to enable other researchers to replicate their findings and justify if this is not possible for any reason. This is a fundamental paradigm shift that is not completed yet but it has started and it is spreading across a number of top quality conferences and journals.
- **Open access:** it is now a common practice for authors to pay for allowing people to access their papers for free to increase the dissemination of their work. This practice is often supported by public research grants to assure the maximum exposition of the research results they have sponsored. Even if the approach is controversial for many reasons (e.g., conflicts of interests of publishers that may lead to lower quality), it has the advantage of making research outcomes available to everybody without any paywall.

### 2.2 The Technological Perspective

The technology has evolved in many different areas but some have been affected more in depth:

- **Software technologies:** OSS is part of almost any software product due to the popularity of powerful and high quality libraries (e.g., the ones from the Apache Foundation, the ones released by major IT companies such as Google, Facebook, IBM, etc.) published according to licenses that allows their integration in both open and closed products (e.g., BSD, MIT, LGPL, etc.). Moreover, many basic components of the overall infrastructure are powered by OSS (e.g., databases, web servers, etc.).

- **Hardware technologies:** also any product that include a software component nearly always include OSS. TVs, cars, IoT devices, etc. often include stripped down versions of open operating systems, libraries, web servers to provide easy to use user interfaces, etc.
- **Software-as-a-Service (SaaS) technologies:** with the development of this approach to the delivery of software systems, a new set of problems have been identified for OSS. Open licenses were not ready to cope with this novel approach for releasing software and many companies took advantage of OSS breaking the basic ideas behind openness. As an example, the GPL license requires that the entire code of a product that use a GPL component needs to be released as GPL if it is released. With SaaS, the code is never released but only used through the Internet. This enabled many companies (including giants like Google, Facebook, Amazon, etc.) to build closed services taking advantage of the open source communities without giving anything back violating the spirit of the GPL. For this reason, a new version of the AGPL license have been developed. This license forces developers to release their code to the community even if the developed product is not released but just delivered as SaaS.
- **Mobile technologies:** with the introduction of the Android operating system, OSS has heavily penetrated the consumer environment becoming the most used operating system. Other open mobile operating systems have been proposed such as Tizen or Sailfish OS but they have a very small market share. Moreover, the popularity of OSS powered hardware platforms such as Arduino and Raspberry PI have made OSS the de facto standard for the development of prototypes, custom IoT projects, etc.
- **Cloud and big data technologies:** many technologies for running the cloud infrastructure has been released as OSS due to changes in the business models (see Section 2.3). A similar approach has been adopted in big data technologies where OSS is the major player with contributions provided by many kinds of companies that cooperate in the development of the technology (e.g., Apache Hadoop ecosystem) but compete in providing solutions to the customers.
- **Machine learning technologies:** all the major IT companies has released their tools as OSS (e.g., Google with Tensorflow, Facebook with PyTorch, Microsoft with CNTK, etc.) offering developers their cloud services for running them. Moreover, such tools are not very useful without a wide amount of data that is needed to build the models able to be used in actual applications.

### 2.3 The Business Perspective

Business models have evolved from focusing on selling software to focusing on providing on-demand services based on OSS. The business of software has changed in many aspects including the following:

- **Software commoditization:** the value in the software business is continuously moving upwards transforming the basic infrastructure and applications

into a commodity. Nowadays, leading applications include a relevant set of features powered by machine learning algorithms that are based on the analysis of a huge amount of data that is not available to the community while software is. For this reason, the business models are changing moving towards the data that will be able to provide competitive advantages to companies that own them. This is a relevant problem for the open source communities that currently do not have the ability to collect and exploit such amount of data preventing the creation of cutting edge applications.

- **Value of data:** machine learning algorithms require a huge amount of data for the definition of reliable models. Such data is continuously collected and enhanced by large IT companies that base their products on them. Data are collected to improve the models that make the products more useful and reliable from which additional data are collected. For this reason data complement software enabling the creation of a new generation of products.
- **Introduction of new OSS licenses:** the introduction of the AGPL license was required to keep the spirit of the GPL license valid also in a world where software is not delivered anymore but offered as a service. This change in the software delivery paradigm requires an adaptation to the licensing approaches as it happens with the increasing value of data.

## 3 Current Trends

According to the three perspectives analyzed in Section 2, it is clear that OSS has penetrated all the business and research environments helping in shifting the focus from the basic software to higher levels that provide more value to the final customers. Moreover, it is clear that pure software is not enough anymore since data is the new king and applications cannot be tuned and work as customers expect without analyzing a massive amount of data extracting useful information that can be used to enhance the user experience.

Open source communities have to deal with these two new aspects and define a proper way to manage them. About SaaS, the community has introduced the AGPL that is able to keep the spirit of openness also in such environments. However, currently, there are no well established approaches to deal with data and assure that they stay open for the usage of the entire community. This is a new challenge that needs to be addressed as soon as possible to keep value in OSS.

## 4 Conclusions and Future Work

The paper presents an initial qualitative investigation of the evolution of the OSS in the last 15 years under three perspectives: research, technological, and business. The main challenges that OSS have to deal with are related to the delivery of SaaS and managing the value of data that is the key enabler for the development of a new generation of applications.

A more detailed analysis is needed through the collection and the analysis of quantitative data to provide a more complete overview of the identified trends and how they are connected to each other.

## References


1. Andreessen, M.: Why Software Is Eating The World. The Wall Street Journal, August 20, 2011. Available at: https://a16z.com/2011/08/20/why-software-is-eating-the-world/
2. Ciancarini P., Poggi F., Rossi D., Sillitti A.: Improving bug predictions in multi-core cyber-physical systems. In: 4th International Conference on Software Engineering for Defence Applications (SEDA 2015), Rome, Italy (2015)
3. Ciancarini P., Sillitti A.: A Model for Predicting Bug Fixes in Open Source Operating Systems: an Empirical Study. In: 28th International Conference on Software Engineering and Knowledge Engineering (SEKE 2016), Redwood City, San Francisco Bay, CA, USA (2016)
4. Ciancarini P., Poggi F., Rossi D., Sillitti A.: Analyzing and Predicting Concurrency Bugs in Open Source Systems. In: 30th International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, AK, USA (2017)
5. Di Bella E., Sillitti A., Succi G.: A multivariate classification of open source developers. Information Sciences, Elsevier **221**, 72–83 (2013)
6. Georgiev A., Sillitti A., Succi G.: Open Soruce Mobile Virtual Machines: An Energy Assessment of Dalvik vs. ART. In: 10th International Conference on Open Source Systems (OSS 2014), San Jose, Costa Rica (2014)
7. Jermakovics A., Sillitti A., Succi G.: Exploring collaboration networks in open-source projects. In: 9th International Conference on Open Source Systems (OSS 2013), Koper, Slovenia (2013)
8. Petrinja E., Sillitti A., Succi G.: Overview on Trust in large FLOSS Communities. In: 4th International Conference on Open Source Systems (OSS 2008), Milan, Italy (2008)
9. Petrinja E., Nambakam R., Sillitti A.: Introducing the Open Maturity Model. In: 2nd Emerging Trends in FLOSS Research and Development Workshop at ICSE 2009, Vancouver, BC, Canada (2009)
10. Petrinja E., Sillitti A., Succi G.: Comparing OpenBRR, QSOS, and OMM Assessment Models. In: 6th International Conference on Open Source Systems (OSS 2010), Notre Dame, IN, USA (2010)
11. Proceedings of the 27th International Conference on Software Engineering (ICSE 2005), ACM (2005)
12. Proceedings of the 41st International Conference on Software Engineering (ICSE 2019), ACM (2019)
13. Rossi B., Scotto M., Sillitti A., Succi G.: Criteria for the non invasive transition to OpenOffice. In: 1st International Conference on Open Source Systems (OSS 2005), Genoa, Italy (2005)
14. Rossi B., Scotto M., Sillitti A., Succi G.: An Empirical Study on the Migration to OpenOffice.org in a Public Administration. International Journal of Information Technology and Web Engineering **1**(3), 64–80 (2006)
15. Russo B., Braghin C., Gasperi P., Sillitti A., Succi G.: Defining the Total Cost of Ownership for the Transition to Open Source Systems. In: 1st International Conference on Open Source Systems (OSS 2005), Genoa, Italy (2005)