Protocol for a SLR on software ecosystems
Technical report

Oscar Franco-Bedoya, David Ameller, Dolors Costal, Xavier Franch

Universitat Politècnica de Catalunya,
Department of Service and Information System Engineering,
Group of Software Engineering for Information Systems,
Barcelona, Spain.

Abstract

Open Source Software (OSS) and Ecosystems (SECO) are two emergent research areas in software engineering. We are interested on the published works that join these two topics, to do it we used a well-known technique called Systematic Literature Review (SLR).

Keywords: Ecosystem, Open Source Software, State of the Art, Systematic Literature Review

1. Background

Systematic literature reviews (SLR) are an accepted method to identify, evaluate, and interpret the available research relevant to a particular topic, research question, or phenomenon of interest conducting secondary studies in software engineering [1, 2]. This protocol follows the original guidelines as proposed by Kitchenham [3]. The review process was split into six stages (Fig. 1).

2. Context

Open source software (OSS) and software ecosystems (SECO) are two emergent research areas in software engineering [4]. The project proposal is
around these concepts. For this reason, we are conducting this systematic literature review on OSS-ecosystems.

2.1. Open Source Software

Open source software has witnessed an exponential increase in the last two decades and it is playing an increasingly important role in many companies and organizations. In recent years, the open source software (OSS) development model has gained significant momentum and is now generally considered a viable approach in commercial settings [1]. Eric Raymond describes the development of OSS as a bazaar like activity driven by volunteers, and claims that OSS is cheaper, has fewer defects, gets improvements faster, and is generally better than other kinds of software [5, 6, 7].

Open source software may be defined as software released under the terms of a license that allows the licensee to use, modify, and redistribute, either gratis or for a fee [1]. Open source software offers new capabilities for developing and reusing software that exceed the benefits offered by proprietary approaches to software development and reuse [15]. The benefits are often seen in modern efforts like Apps Challenges; in social networking and software development ecosystems like Github [9].

Software engineering researchers have used OSS products to study general software engineering problems like evolution [10], cloning [11], and the use of metrics to identify error prone classes [12].

2.2. Open source software ecosystem (OSS-ecosystem)

There are three generalized definitions of SECO. Jansen et al [13], define a software ecosystem as a set of actors functioning as a unit and interacting with a shared market for software and services. A SECO consists of actors such as Independent Software Vendors (ISV), outsourcers, and customers. A SECO typically is interconnected with institutions such standardization organizations, open source software communities, research communities, and the related ecosystems [14]. Bosch [15], define SECO as a set of software solutions that enable, support and automate the activities and transactions by the actors in the associated social or business ecosystem and the organizations that provide this solutions. The last is Lungu’s view, who defines a software ecosystem as a collection of software projects which are developed
and evolve together in the same environment [4].

Jansen definition includes both business and technology ecosystems components, and moreover the SECO elements can or cannot be software components, Bosh defined a SECO as technological infrastructure to support the business ecosystem and finally for Lungu the SECO is related with the software development platform.

Jansen gives the only OSS-ecosystem definition that we found in the SLR: “An open source ecosystem is one where it is possible to add contributions to a project create and publish components in the extension market, etc., without any barrier” [16].

3. SLR Protocol

Referring to Fig. 1, the review protocols consist of six stages enumerated as follows: research questions, search design process, selection process, quality assessment, data collection and data synthesis. In the first stage, a set of research questions were formulated based on the aim of this study. In the second stage, search strategies were designed in line with the formulated research questions, which consisted of the identification of search terms the choices of literature resources and the inclusion criteria. The third stage dealt with the selection process while the fourth stage concentrated and selected on applying the quality assessment criteria. In the fifth stage were extracted the data from the studies, and finally the sixth phase an analysis of all final studies were done.

3.1. Research questions

The overall research objective of the study is summarizing what we know about OSS-ecosystems. This has been broken down into four high-level research question addressed by this review:

• RQ1. What is OSS-ecosystem?
• RQ2. How to industry implements the OSS-ecosystem?
• RQ3. What phases, processes, or stages are described for OSS-ecosystems evolution?
• RQ4. How OSS-ecosystems are represented in the literature?

The research questions postulated in this review are exploratory, as we attempt to understand and identify useful quality data and clarify definitions about the OSS-ecosystem phenomena. The high-level research questions were decomposed into sub-questions. The following subsections describe the reasoning for each research question.

3.1.1. RQ1 What is OSS-ecosystem?

Open source software (OSS) and software ecosystems (SECO) are two emergent research areas in software engineering [17]. Therefore, by RQ1 we want to identify existing elements, definitions and general characteristics of OSS, SECO and OSS-ecosystems existing in the software engineering literature. To address RQ1, we decomposed the RQ into four specific research sub-questions:

• RQ1.1 What elements belong in an OSS-ecosystem?

• RQ1.2 What are the differences reported in the literature between the definitions of OSS-ecosystem, digital ecosystem and business digital ecosystem?

• RQ1.3 What metrics or attributes are defined to assess or evaluate OSS-ecosystems?
• RQ1.4 What instances of OSS-ecosystem are the most studied?

3.1.2. RQ2 How to industry implements the OSS-ecosystem?

The open source software strategy provides new opportunities for the companies [1]. Multi-project software ecosystems emerge, as software sharing or reuse within traditional software development enterprises [23]. With respect to RQ2, we considered study the implementation of OSS into organizations, but with a particular focus on industry. RQ2 was decompose into three research sub-questions:

• RQ2.1 What are the company roles in an OSS-ecosystem?
• RQ2.2 What are the participation strategies of the companies in the OSS-ecosystems?
• RQ2.3 What kind of infrastructure is used to enable the interaction of the companies in the OSS-ecosystem?

3.1.3. RQ3 What phases, processes, or stages are described for OSS-ecosystems evolution?

Since research in ecosystems from a software engineering perspective is quite new [20], little work to-date has been done to offer modeling analysis tools and techniques for software ecosystems analysis. By RQ3, we emphasized the need of identifying processes for modeling and analyzing OSS-ecosystems.

3.1.4. RQ4 How are representing in the literature the OSS-ecosystems?

Finally, by RQ4 we aim to identify the OSS-ecosystem representation in the literature, identifying its models related, particular notation and guidelines. RQ4 is divided in two sub-questions:

• RQ4.1 What models are related with OSS-ecosystems?
• RQ4.2 Which notation and guidelines have been used for modelling of the OSS based-ecosystems?.
3.1.5. RQ5 How publications in the topic are demographically distributed?

The review also highlights some information about demography attributes. RQ5 was decomposed into four research sub-questions:

- RQ5.1 In which type of venue are articles mostly published?
- RQ5.2 How the number of publications has evolved over the years?
- RQ5.3 Which countries publish more articles related to the topic?
- RQ5.4 How are publications distributed between research and industry?

3.2. Search design process

The aim of study selection is to find as many primary studies relating to the research questions as possible using an unbiased search strategy [1]. The search process combining searching digital libraries with manual search. In addition, we added specific papers from two secondary studies about software ecosystem [21, 22] and finally the articles in the Jansen book [23], were reviewed.

3.2.1. Literature sources

We selected a set of publication channels. Relevant journals and conferences were taken from previous literature reviews on software engineering and open source software [5]. Table 1 and Table 2 give an overview of the final sample of publication sources.

3.2.2. Search string

The aim for our search string is to capture all results that relate to the research questions. According to Kitchenham [1] the following steps were used to build the search terms:

- Derivation of major terms from the research questions.
- Identification of alternative spellings and synonyms for major terms.
- Usage of the Boolean OR to incorporate alternative spellings and synonyms.
- Usage of the Boolean AND to link the major terms.

The search string used on all databases is:
Table 1: Publication sources.

(OSS OR FOSS OR FLOSS OR Open Source OR Free Software OR Libre Software) AND ecosystem

Please note that:

- The potential synonymous to OSS have been identified, therefore we included the terms “Free Software”, “Libre Software” and the commonly used acronyms OSS, FOSS, FLOSS.

- To avoid overlooking relevant publications, we opted for a search strategy with high sensitivity (See Hauge). This means that instead of using keywords like open source software, we conducted all searches using the keywords open source (including quotation marks) and searched the digital libraries using all fields, including title, abstract and keywords where available or metadata.

- To improve the precision of the search, in some databases we limited the results to some research areas.

- Finally, we proceeded in an incremental way and run several tests in order to verify the precision of our search string until it was satisfactory
3.2.3. Inclusion and exclusion criteria

The following constraints have been used in order to select the relevant publications:

- Only publications in English.
- Only papers published between 1998 and 2012.
- Have been excluded introductions of panels, conferences and special issues, book reviews, news flashes, short papers (less than 4 pages) and PhD symposium papers.

3.3. Selection process

The search process combined searching digital libraries with manual evaluation and classification of the results. Fig. 2 presents an overview of the review process and the number of publications included in each stage.

3.3.1. Stage 1 - Deleting duplicate articles

This stage involve searching the keywords, on the relevant literature sources. To have a clearly defined set of publications serving as a basis for this review, was result, 437 primary studies were identified.

3.3.2. Stage 2 - Deleting duplicate articles

Duplicate records are resolved in this stage by importing the references to a reference manager system (reference a RefWorks) and automatically remove duplicate papers. Finally, one researcher manually reviewed the list of articles in order to identify duplicated records. 139 papers were exludes in this stage.
3.3.3. Stage 3 - Reviewing titles and abstracts

To identify publications that in fact were about OSS ecosystems, one researcher individually reviewed all titles and abstracts and checks the inclusion and exclusion criteria for each entry. Only papers on OSS ecosystem topics were included. After this stage we discarded 188 of the 298 papers and ended up including 110 publications.

3.3.4. Stage 4 - Peer review titles and abstract

Next, to discard papers not related with domain-specific, OSS ecosystem, the other two researchers individually evaluated the output papers of the third stage by reviewing the titles and abstracts. At the end of these stage we discarded 36 papers and included 74 publications.

3.3.5. Stage 5 - Peer fast reading

Then, to select the papers from the fourth stage, to relevant ones, the title, abstract of each study was considered, and each researcher briefly studied their contents. Hence all the papers that noty reflect the study’s topics or not addressing any of the research questions were excluded. At the end of the exercise, 34 papers were selected.
3.3.6. Stage 6 - Identifying papers from secondary studies

Thereafter, in order to identify the maximum amount of important papers that might have been missed, we include the papers from two secondary studies [21, 22]. The Fig. 3 shows a Venn diagram describing the overlap between the papers from stage fifth and the papers from the secondary studies.

The Venn diagram shows that we considered 19 papers out of 90 papers from [21] and considered five papers out of a total of 44 papers from [22]. The classification process for the papers from secondary studies was the same process that we used in the stages (S1-S5). We included a set of 11 papers.

3.3.7. Stage 7 - Identifying papers from manual search

We performed a manual search in specific conference proceedings (see Table 1). One paper was identified using manual search process.

3.3.8. Stage 8 - Identifying papers from book: “Software Ecosystems”

While we were conducting the SLR, Jansen et al published your book: Software Ecosystems: Analysing and Managing Business Networks in the Software Industry [16]. We considered all papers from the book and selected 6 relevant studies. Finally, the last stage of the review included 53 relevant papers for this SLR. (see Appendix A)
3.4. Quality assessment

The quality assessment is performed only to evaluate the rigor of the presented research in each publication and not for want to exclude papers. We formulated a number of quality assessment questions using the nine quality metrics (QM) presented below. This schema is inspired by other SLRs [5]. Each question has only three optional answers: “Yes”, “partly” or “N”.

- QM1: Does the paper have a description of the research method?
- QM2: Does the paper describe an explicit research question/goal/purpose?
- QM3: Does the paper describe motivation for the research question(s)?
- QM4: Does the paper discuss limitations or validity?
- QM5: Does the paper describe the context of the research?
- QM6: Does the paper describe data collection?
- QM7: Does the paper describe data analysis?
- QM8: Does the paper describe sampling or selection of the study object(s)?
- QM9: Does the paper present any data?

3.5. Data collection

In the data collection stage we extracted the following from each of the publications: main topic, research question or research goal, year of publication, journal or conference, research method, a brief description of empirical evidence relevant to use of OSS ecosystem, and the affiliation and home country of the first authors.

3.6. Data analysis

For the coding and analysis of the data, we are using Atlas.ti an qualitative data analysis tool. Fig. 4 shows a network of codes for the research question 1.
Figure 4: Codes for research question 1

References


Appendix A. SLR Included papers


