

Open Source Software Ecosystems: A Systematic Mapping

Oscar Franco-Bedoya^{a,b}, David Ameller^a, Dolors Costal^a, Xavier Franch^a

^aUniversitat Politècnica de Catalunya, Software and Service Engineering Group, Barcelona, Spain.

^bUniversidad de Caldas, Manizales, Colombia

Abstract

Context: Open source software (OSS) and software ecosystems (SECOs) are two consolidated research areas in software engineering. OSS influences the way organizations develop, acquire, use and commercialize software. SECOs have emerged as a paradigm to understand dynamics and heterogeneity in collaborative software development. For this reason, SECOs appear as a valid instrument to analyze OSS systems. However, there are few studies that blend both topics together.

Objective: The purpose of this study is to evaluate the current state of the art in OSS ecosystems (OSSECOs) research, specifically: (a) what the most relevant definitions related to OSSECOs are; (b) what the particularities of this type of SECO are; and (c) how the knowledge about OSSECO is represented.

Method: We conducted a systematic mapping following recommended practices. We applied automatic and manual searches on different sources and used a rigorous method to elicit the keywords from the research questions and selection criteria to retrieve the final papers. As a result, 82 papers were selected and evaluated. Threats to validity were identified and mitigated whenever possible.

Results: The analysis allowed us to answer the research questions. Most notably, we did the following: (a) identified 64 terms related to the OSSECO and arranged them into a taxonomy; (b) built a genealogical tree to understand the genesis of the OSSECO term from related definitions; (c) analyzed the available definitions of SECO in the context of OSS; and (d) classified the existing modelling and analysis techniques of OSSECOs.

Conclusion: As a summary of the systematic mapping, we conclude that existing research on several topics related to OSSECOs is still scarce (e.g., modelling and analysis techniques, quality models, standard definitions, etc.). This situation calls for further investigation efforts on how organizations and OSS communities actually understand OSSECOs.

Keywords: Software Ecosystem, Open Source Software, Systematic Mapping, Literature Review, OSS, SECO, OSSECO.

1. Introduction

In the last decade, the study of open source software (OSS) has changed significantly. OSS employs new types of socialization processes, development practices, community networking, business models, organization structure, governance, and legality (Scacchi, 2007). On the other hand, software ecosystems (SECOs) are increasingly popular for their economic, strategic, and technical advantages (Berger et al., 2014). SECOs provide a new holistic point of view for understanding OSS.

In this sense, OSS initiatives typically create an adequate environment for making a SECO emerge from their projects, communities, and external actors (partners, public and private institutions, research groups). This means that SECOs provide a practical approach to understand all of the synaptic relationships between OSS heterogeneous elements. This paper aims at uncovering the existing research on OSS ecosystems (OSSECOs).

We conducted a systematic mapping study with the aim of identifying and analyzing the primary studies on OSSECOs published in academic venues. The mapping study took 2003 as the starting year because it was the year of publication of the seminal book on SECOs by Messerschmitt and Szyperski (2003). We retrieved and

Email addresses: ohernan@essi.upc.edu (Oscar Franco-Bedoya), dameller@essi.upc.edu (David Ameller), dolors@essi.upc.edu (Dolors Costal), franch@essi.upc.edu (Xavier Franch)

analyzed the literature on OSSECOs by defining and conducting a rigorous protocol following the guidelines described in Petersen et al. (2015, 2008) and Kitchenham and Charters (2007). We considered 652 papers published between 2003 and 2015, and, after a rigorous selection process (see Section 3.4), we obtained a total of 82 papers. We used these papers to answer three research questions divided into a total of 14 sub-research questions. As a result of the systematic mapping, this paper presents an overview of the research in the field by: (a) analyzing the information about demographic characteristics (i.e., type of publication sources, volume of research reported by year, predominant researchers, type of papers, and distribution of publications between industry and academy); (b) identifying and analyzing the different definitions, elements, measures, and instances of OSSECOs proposed by these papers; and (c) classifying the existing OSSECO representations.

The rest of the paper is organized as follows. Section 2 provides a brief background of the general characteristics of OSS, and we present an evolution of the SECO definition. Section 3 details the protocol and the research questions of this mapping study. Section 4 presents the synthesis results of the data extracted from the selected studies and answers the research questions. A discussion of the results of the systematic mapping is presented in Section 5. The article ends with a proposal of future work in OSSECO and a summary of the conclusions.

2. Background

The origins of SECO research have mainly been inspired by studies from business and natural ecosystems (Joshua et al. 2013). In this section, we briefly describe the definitions of OSS and SECO.

2.1. Open Source Software

Nowadays, the adoption of Open Source Software (OSS) by organizations has become a strategic need in a wide variety of application areas. It has changed (and still is changing) the way organizations develop, acquire, use, and commercialize software (van Angeren et al., 2011). Furthermore, OSS has become a strategic asset in software development with prospects of up to 95% of mainstream IT organizations including OSS in their mission-critical portfolios by the end of 2016 (Driver, 2013). Therefore, organizations are increasingly becoming OSS adopters, and several different OSS adoption strategies that are usually followed by the industry have already been identified and analyzed

(López et al., 2015). There are different terms for labeling the OSS phenomenon that reflect different views on what it is: open source software, free software, and free (libre) open source software (FOSS/FLOSS). We adopt an agnostic view and we will treat those terms as synonymous using OSS as the general term because those differences do not really affect the main goal of the paper. There is a need to understand the unique properties of OSS and then identify the concerns that might create barriers for organizations to adopt OSS (Li et al., 2013). The general concept behind OSS covers software artefacts including source code, licenses, development best practices, innovation, ethics, philosophy, social movement, community, culture, governance and organizational engagement. Typically, the developers are primary volunteers. In addition, the software emerges from a loosely coordinated, unsupervised community of developers and other contributors (van Angeren et al., 2011). Finally, the OSS development approach has assisted in the spread of emerging technologies, allowing users to utilize freely publicly available software and developers to incorporate third-party source code into their implementations. Individual and already tested libraries are often used as building blocks¹ for larger software systems, offering reusable functionality and providing the means for faster time-to-release (Kapitsaki et al., 2015).

2.1.1. OSS Projects and Communities

OSS projects are typically initiated by an individual or a small group with a specific need. This need is the motivation for the creation of the OSS project (Uden et al., 2007). Rather than a single corporate entity owning the software, a sometimes broad community of volunteers determines which contributions are accepted into the source code base and where the OSS project is heading (Riehle, 2007) (e.g., R SECO, Gentleman and Ihaka 2015).

OSS communities are keystone actors of OSS projects. They guarantee the development, support, and maintenance of OSS (Foulonneau et al., 2013). An OSS community involves organizations and individuals producing and consuming OSS components. There are many roles in an OSS community with different levels of participation e.g., users, reviewers, contributors, administrators, partners, and developers (Squire and Williams, 2012) (e.g., Eclipse Foundation 2015 SECO). As noted by several authors, the overall OSS commu-

¹Building blocks are software components that can be independently developed and deployed

nity forms a very peculiar complex system that is endowed by an inner short timescale dynamics and a long timescale evolutionary dynamics (Uden et al., 2007). OSS communities surrounding OSS projects provide access to all of the data related to their evolution. These can be used to evaluate the quality of OSSECOs.

2.1.2. OSS Resources

OSS projects accelerate and support the adoption of emerging component-based collaborative platforms. OSS projects typically provide access to several kinds of data sources to extract information about their evolution and the symbiotic relationships between OSS actors (Figay and Ghodous, 2009; Goeminne and Mens, 2013). According to Buford et al. (2011), collaboration tools help OSS communities to coordinate their activities and enable groups of adopters and providers to work as a team, sharing information and communicating as needed, without being co-located. Some of the traditional OSS data sources are: version control systems, mailing lists, bug trackers, web sites, wikis, discussion forums, etc. However, there are also non-traditional data sources such as adopter feedback, market share reports, sales reports, OSS actor surveys, decision-making notes, expert interviews, etc.

2.1.3. OSS Analysis

OSS projects typically provide public availability of historical data, which facilitates the analysis of OSS evolution (Goeminne, 2014). A number of studies have investigated large, well-known OSS projects through quantitative analysis (e.g., Linux kernel, Apache, Mozilla, Gnome, KDE). Several of these studies focus on social network analysis (Fitzgerald and Agerfalk, 2008). They rely on repository mining techniques to extract relevant data from OSS repositories or other data sources that are frequently used by OSS communities. These works use empirical software engineering methods by exploring and studying the OSS communities, including the way they work, cooperate, communicate, and share information (Mens and Goeminne, 2011).

There is increasing interest in research on software repositories, in particular in the software repository mining community. This community focuses on the analyses of the data available in OSS repositories. The researchers of this community explore a range of software engineering questions such as: software evolution, models of software development processes, characterization of developers and their activities, prediction of future software qualities, use of machine-learning techniques on software project data, software bug prediction, analysis of software change patterns, and analy-

sis of code clones (Xie et al. 2013). Obtaining data from OSS repositories is a tedious exercise, and the obtained datasets are often non-homogeneous, which makes further analysis difficult (Gousios and Spinellis 2012). However, there are collaborative development sites like like GitHub that provide access to their internal data stores through an extensive REST application programming interface, which enables researchers to identify a rich collection of OSSECO information (e.g., OSSECO cross-references, technical dependencies between projects).

2.2. Software Ecosystems

SECOs have emerged in the last few years as a novel way to understand the relationships between software projects, products, communities, and organizations (Franco-Bedoya et al., 2014). Furthermore, they are increasingly popular because of their economic, strategic, and technical advantages (Berger et al. 2014). Unfortunately, in contrast to natural ecosystems, there is no common definition of SECO. A SECO can be defined and interpreted in different ways, depending on the point of view (Mens et al., 2014). Two main viewpoints for SECO can be identified, namely business-centric and platform-centric.

The first view emphasizes a holistic, business-oriented perspective of a SECO as a network of actors, organizations, and companies. It is adopted by authors such as Messerschmitt and Szyperski (2003), Bosch (2009), and Jansen et al. (2009a). This view is similar to the commercial software ecosystem category in Bosch (2009), the external view level in Jansen et al. (2009b), the business dimension in dos Santos and Werner (2011) and the ecosystem-in-the-large in Goeminne and Mens (2013).

The second view highlights technical and social aspects of a set of software projects, technical platforms, and communities. This perspective is adopted by authors such as Lungu (2008) and Goeminne and Mens (2013). This view is similar to the social software ecosystems in Bosch (2009), the internal view level in Jansen et al. (2009b), the social and architectural dimensions in dos Santos and Werner (2011), and the ecosystem-in-the-small in Goeminne and Mens (2013). This Platform-centric viewpoint focuses on the platform environment.

The following subsection describes the evolution of the SECO definition that originated from the term ecosystem from the ecology domain.

²The whole complex of organisms present in an ecological unit may be called the biome.

Table 1: Ecosystem definitions

Definition (sorted by date)	Type
"The whole system (in the sense of physics) including not only the organism-complex, but also the whole complex of physical factors forming what we call the environment of the biome ² , the habitat factors in the widest sense." Tansley (1935)	ecosystem
"An economic community supported by a foundation of interacting organizations and individuals, the organisms of the business world." Moore (1993)	BECO
"It is a socio-economic development catalysed by information and communications technologies (ICTs)." Nachira et al. (2002, 2007)	DBECO
"Your own BECO include, for example, companies to which you outsource business functions, institutions that provide you with financing, firms that provide the technology needed to carry on your business, and makers of complementary products that are used in conjunction with your own. It even includes competitors and customers, when their actions and feedback affect the development of your own products or processes. The ecosystem also comprises entities like regulatory agencies and media outlets." Iansiti and Levien (2004)	BECO
"Traditionally, a software ecosystem refers to a collection of software products that have some given degree of symbiotic relationships." Yu et al. (2007)	SECO
"A software ecosystem is a collection of software projects which are developed and evolve together in the same environment." Lungu (2008)	SECO
"A software ecosystem consists of the 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 these solutions." Bosch (2009)	SECO
"We define a software ecosystem as a set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently under-pinned by a common technological platform or market and operate through the exchange of information, resources and artifacts." Jansen et al. (2009a)	SECO
"It is defined as a distributed adaptive open socio-technical system, with properties of self-organisation, scalability and sustainability." Stanley and Briscoe (2010)	DBECO
"We define the ecosystem as the source code together with the user and developer communities surrounding the software." Goeminne and Mens (2010)	SECO
"It is a network of organizations or actors, and a common interest in the development and use of a central software technology." Hanssen and Dybå (2012)	SECO
"A software ecosystem is a universe of shared assets centered around a common technical platform. In this universe, various roles, mainly suppliers and consumers, interact in order to develop, manage, and consume assets." Berger et al. (2014)	SECO
"The software and actor interaction in relation to a common technological infrastructure, that results in a set of contributions and influences directly or indirectly the ecosystem." Manikas (2016a)	SECO

2.2.1. Evolution of the SECO definition

Ecosystem (ECO) theorizing began in 1935 when Tansley (1935) matured the term coined by Clapham in 1930 to denote the physical and biological components of an environment when considered in relation to each other as a unit. Tansley realized the importance of the relationship between a community of units and the

environment in which they existed (van Angeren et al., 2011; Mens et al., 2014).

Fifty-eight years later, Moore (1993) introduced the concept of business ecosystem (BECO), which was later adopted by other works (e.g., Mizushima and Ikawa 2011; Stanley and Briscoe 2010; Dhungana et al. 2010). Moore's definition of BECO is *an economic community supported by a foundation of interacting organizations and individuals*. In 2004, Iansiti and Levien stated that a BECO evolves around a platform (e.g., software market, foundation technology, etc.). This definition is used in other works such as Viljainen and Kauppinen (2013), Jansen et al. (2013), and den Hartigh et al. (2013). For instance, Hartigh et al. operationalized the work of Iansiti and Levien to measure the health of the Dutch IT industry.

In 2002, the term digital business ecosystem (DBECO) was coined by adding *digital* to business ecosystem (Stanley and Briscoe, 2010). Nachira et al. refer to DBECO as *a socio-economic development catalysed by information and communications technologies (ICTs)* (Nachira et al., 2002, 2007). The matured definition of DBECO was proposed by Briscoe (2009) and Stanley and Briscoe (2010). They defined DBECO as *a distributed adaptive open socio-technical system, with properties of self-organization, scalability and sustainability*.

The term software ecosystem (SECO) was introduced by Messerschmitt and Szyperski (2003). This concept is cited in several works (e.g., Kabbedijk and Jansen (2011), Scacchi and Alspaugh (2012), Yu (2013)). However, the first definition of SECO was provided by Yu et al. (2007), this is based on the Messerschmitt and Szyperski (2003) ecosystem conceptualization. They defined a SECO as *a collection of software products that have some given degree of symbiotic relationships*. In 2008 Lungu, looked at SECO as *a federation of systems in a common platform* (Lungu, 2008). One year later, Jansen et al. (2009a) defined a SECO as *a set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them*. This definition was adopted by works such as Weiss (2011), Jansen et al. (2012), and Manikas and Hansen (2013a). This very year, Bosch (2009) discussed the implications of a software product line company transitioning to a software ecosystem approach. Finally, Hanssen and Dybå (2012) identify two fundamental factors of SECOs (i.e., network of organizations or actors and a common interest in a central software technology). This definition is used in Hoving et al. (2013), Berger et al. (2014), and Mens et al. (2014).

Because of its evolution, the SECO definition inherits several elements and features from the ECO, BECO, and DBECO definitions. Our study is mainly focused on SECOs in the context of OSS, namely OSSECOs. Nevertheless, the other types of ecosystems, specifically BECOs, are still worth considering. As result of this study and based on the definitions found in the selected papers, we constructed a general definition for the term OSSECO (see Section 6.1).

Table 1 summarizes the SECO-related definitions that are reported in this subsection.

3. Research Method

This research is based on both the guidelines proposed by Petersen et al. (2008, 2015) for the development of mapping studies and those described by Kitchenham and Charters (2007) for the development of systematic literature reviews. The mapping process is split into several phases (see Fig. 1), which are described in the following subsections.

3.1. Identification of the need for a review

As Kitchenham and Petersen state, prior to undertaking any systematic literature study, researchers should identify and review any existing systematic review of the phenomenon of interest against appropriate evaluation criteria. There is no procedure defined to implement this stage. However, similar to Oriol et al. (2014), we applied two strategies. First, to broaden the scope of the results, we searched other systematic literature studies in the area of SECOs (not only open-source). Second, we followed a procedure that is analogous to the main search of our systematic mapping. In other words, we defined a search protocol to identify other secondary studies. The protocol was based on the protocol defined in the main search, which will be explained in the following sections. In short, we used the same digital libraries (see Table 4), and we built the search string as a conjunction of population and intervention as recommended by Kitchenham and Charters (2007). From each term of the population and intervention, we identified a set of variants and acronyms:

(“*software ecosystem*” OR “*software ecosystems*”) AND (“*state of the art*” OR “*SLR*” OR “*review*” OR “*systematic mapping*”).

As a result of this search³, we identified six sec-

³In www.essi.upc.edu/~gessi/PLATEOSS. There are several documents with detailed information that is not in this paper (e.g., the set of all papers containing search terms, author measures, list of terms, etc.)

ondary studies that presented a review on SECOs: Barbosa and Alves (2011); their updated work Barbosa et al. (2013); Manikas and Hansen (2013); their expanded work Manikas (2016a); Axelsson and Skoglund (2016) and Manikas (2016b). Afterwards, we analyzed all of the selected papers from these studies and found a new study by Hanssen and Dyba (2012), which is a kind of secondary study about theorizing in the SECO research literature.

Therefore, there were seven secondary studies finally selected: Barbosa and Alves (2011); Hanssen and Dybå (2012); Barbosa et al. (2013); Manikas and Hansen (2013); their expanded work Manikas (2016a); Axelsson and Skoglund (2016) and Manikas (2016b). It is worth noting that none of these works was conducted specifically on OSSECOs. Instead, all papers focus on SECOs in general except Manikas (2016b) that focuses on proprietary SECOs.

- Barbosa and Alves (2011). See Barbosa et al. (2013) which expands this work.
- Hanssen and Dybå (2012) described the theoretical foundations of SECOs. In their work, they identified openness and transparency as one of the fundamental concepts for further and deeper research in SECO theorizing. In addition, they presented five main theories related to SECOs: activity theory, transaction cost theory, systems theory, sociotechnical theory, and intermediary theory.
- Barbosa et al. (2013) conducted a systematic mapping study on SECOs. They defined four research questions about the characteristics, benefits, and challenges of SECOs. In their work, ten characteristics of SECOs were identified and eight main SECO research areas were found. According to Barbosa and Alves, the most relevant research areas in SECOs are open source software, ecosystem modelling, and business issues. Finally, they highlighted the relevance of OSS models in the context of SECOs.
- Manikas and Hansen (2013). See Manikas (2016a) which expands this work.
- Manikas (2016a) analyzed 231 papers collected from 2007 until 2014. He identified the term open as one of the keywords related to SECOs, which is specifically related to the domain of the studies (e.g., the OSS domain). Furthermore, he identified three signs of SECO maturity: (a) a rapid increase in the number of journal articles; (b) an increase in the empirical models within the last two

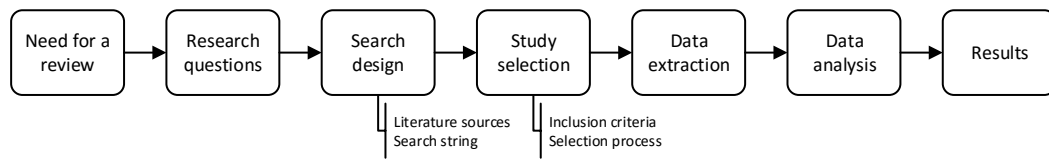


Figure 1: Stages for the systematic mapping

years; and (c) a large set of ecosystems studied. Finally, Manikas encouraged undertaking studies of specific SECO definitions rather than wide ecosystem studies in order to address SECO complexity. Our systematic mapping is an in-depth study of a specific type of SECO (i.e., OSSECO).

- Axelsson and Skoglund (2016) investigated the challenges related to quality assurance in software ecosystems and identified what approaches have been proposed in the literature. They selected six papers covering quality assurance in software ecosystems from different perspectives. The authors, also presented a list of research challenges that are specific to quality assurance in SECOS (e.g., stakeholder requirements definition and system architectural design). In their research agenda, Axelsson and Skoglund called for more research (primarily empirical) to better understand niche player needs (such as OSS communities).
- Manikas (2016b) investigates literature on non open-source ecosystem studies and identifies the aspects studied in this type of SECOS.

Given the lack of secondary studies specific for the OSSECO topic and the observation that OSSECOs have specific characteristics in the context of SECOS, such as the presence of an OSS community actor, we think that conducting a systematic mapping about OSSECOs is justified. In the next subsection we provide further details of the relationships of the research questions in these secondary studies compared to those in ours.

3.2. Research questions

The overall research objective of this study is to find and analyze the current state of the art in OSSECOs. This objective has been broken down into three high-level research questions (RQs) which, in turn, will drive the review method. The RQs postulated in this review are exploratory since we are attempting to understand and identify useful quality data and clarify definitions

about the OSSECO phenomenon. In addition, the high-level research questions are divided into research sub-questions. Table 2 shows the RQs and their motivation.

Once the RQs of this study have been formulated, we compare them with those of the secondary studies identified in Section 3.1 (see Table 3).

- RQ2 is partially addressed by the Barbosa et al.'s RQ1: *What are the main characteristics of a Software Ecosystem?* However, our goal in this RQ was to find a definition for OSSECOs, which is related but different.
- RQ1, RQ1.2 and RQ1.3 are addressed by Manikas (2016a). However, this type of research questions is a usual practice in systematic reviews, according to the common guidelines for this type of study. For instance, this information can be useful as input for further studies in the field in order to establish research trends.
- Manikas (2016a)' RQ: *How is the term software ecosystem defined?* is similar to our research question RQ2.1. However, we are searching definitions for OSSECO specifically.
- Manikas (2016a)' RQ: *Is software ecosystem research targeting real software ecosystems?* is related to our research question RQ2.4. However, we are searching for instances of OSSECO specifically.

3.3. Search design process

In every systematic mapping, the primary studies are identified by using automatic searches on scientific bibliographies or browsing manually by gathering the works from specific known journals and conferences of the target field. In our systematic mapping, we applied an automatic search that was complemented with manual searches in the specific venues listed in Section 3.3.1. The aim of this search process was to find as many primary studies related to the research questions as possible using an unbiased search strategy.

Table 2: Research questions

Research Question	Interest and motivation	Sub-questions
RQ1. What are the demographic characteristics of the studies about OSSECOs?	Identify the type of publication, in particular journals publications, and the type of papers, in particular empirical, is important because are indicators of the maturity in a new research field (Manikas, 2016a). The evolution in the number of publications is an indication of how the activity of a research field changes (Kitchenham et al., 2009). The information about geographical distribution of the publications and the classification between academy and industry is relevant because OSSECOs concept extend geographical and institutional boundaries. Finally, the OSSECO predominant researchers are important in order to identify the keystone authors in the growing network of OSSECO researchers.	RQ1.1 In which type of sources are articles mostly published? RQ1.2 How has the number of publications evolved over the years? RQ1.3 How are papers geographically distributed? RQ1.4 Who are the predominant researchers? RQ1.5 How are publications distributed between academy and industry? RQ1.6 What type of papers are published?
RQ2. What is an OSSECO?	OSS and SECOs are two emergent research areas in software engineering (Scacchi and Alspaugh, 2012). Consequently, by answering this RQ, we can get information about: existing elements, definitions and general characteristics of OSS, SECO and OSSECO existing in the software engineering literature.	RQ2.1 What definitions are related to the OSSECO definition? RQ2.2 Are there specific definitions of OSSECO? RQ2.3 What elements belong to an OSSECO? RQ2.4 What instances of OSSECOs have been reported in the literature?
RQ3. Which representations have been proposed for OSSECOs?	To identify which are the representations proposed in the literature for OSSECO, identifying modelling techniques, analysis, particular notations and guidelines.	RQ3.1 Which primary studies use models to represent OSSECOs? RQ3.2 Which of the proposed models, if any, are specific for OSSECOs? RQ3.3 Which notation and guidelines have been used for modelling OSSECOs? RQ3.4 What type of analysis was conducted using the models identified in RQ3.3?

Table 3: Relationships between research questions of our study and other secondary studies

RQ	Hansen and Dybå	Barbosa et al.	Manikas2016a	Axelsson and Skoglund
RQ1	N	N	A	N
RQ1.1	N	N	A	N
RQ1.2	N	N	A	N
RQ1.3	N	N	N	N
RQ1.4	N	N	N	N
RQ1.5	N	N	N	N
RQ1.6	N	N	N	N
RQ2	N	PA	N	N
RQ2.1	N	N	PA	N
RQ2.2	N	N	N	N
RQ2.3	N	N	N	N
RQ2.4	N	N	PA	N
RQ3	N	N	N	N
RQ3.1	N	N	N	N
RQ3.2	N	N	N	N
RQ3.3	N	N	N	N
RQ4	N	N	N	N

(N: Not addressed, PA: partially addressed, A: addressed).

3.3.1. Literature sources

To ensure the consideration of appropriate venues, we selected a set of publication channels. The main purpose of this selection was to double-check that automatic searches covered all of these venues. In order to do this, relevant journals and conferences were selected from previous literature reviews on software engineering, OSS, and SECOs (Hauge et al., 2010). Furthermore, we added the four systematic literature reviews about SECOs mentioned in Section 3.1. Finally, we decided to add the book by Jansen et al. about SECOs

(Jansen et al., 2013) because, based on our knowledge and that of other authors (Mens et al., 2014), this is the only book that is completely devoted to the study of the concept of SECO. We finally selected the following list of journals, conferences, and workshops:

- Journals: (*Software engineering*) TOSEM, ASE, Communications, Computer IEEE, IEEE Software, DKE, EMSE, Engineering & Technology⁴, IEEE Review, TSE, IET⁵, ISJ, IST, JSS, REJ, SPE, SoSyM, SPIP⁶. (*OSSECOs and Information systems and management*) First Monday, Information Technology & People, IJOSSP, Journal of Industrial Economics, Knowledge Technology and Policy, Long Range Planning, Management Science, MIS Quarterly Executive, Research Policy.
- Conferences and workshops: (*Software engineering*) ASE, CAISE, COMPSAC, ESEC/FSE, ESEM, HICSS, ICSE, ISESE, METRICS, RE, SAC, SEKE. (*OSSECOs and Information systems and management*) FOSDEM, IWSECO, OS-

⁴Previously IEEE Software Proceedings⁵Previously IEE Review⁶From 2010 incorporated in Journal of Software Maintenance and Evolution (last issue December 2011)

CON, OSS, IFIP, WOSSE-ICSE, WFLOSS-ICSE, WSKS.

Note that these sources represent the main corpus whose exploration needs to be enforced by the systematic mapping. However, since we will use digital libraries (see Section 3.4), papers published in other venues will eventually be found.

3.3.2. Search string

The aim of our search string is to capture all of the results that relate to the research questions. According to Kitchenham and Charters (2007), a good way to create the search string is to structure it in terms of population, intervention, comparison, and outcome. However, similar to Ameller et al. (2015), we focused on the population dimension. Since, in fact, we are interested in two areas simultaneously, the search string is a conjunction of the two corresponding populations:

search string = OSS population AND Ecosystem population

There are several terms for OSS (see Section 2.1). The potential OSS synonyms have been identified from Hauge et al. (2010): "Free Software", "Libre Software", and the commonly used acronyms OSS, FOSS, and FLOSS. All of these are included in the search string to capture all relevant literature. In contrast, for the word ecosystem, we have identified "Software Supply Network" from Jansen (2007). He uses this term to define a network of linked, software products, hardware, and services to satisfy market demands. In addition, we have used the term "Software Supply Industry" from Messerschmitt and Szypersky's book (Messerschmitt and Szyperski 2003). The resulting query string was:

(OSS OR FOSS OR FLOSS OR "Open Source" OR Free Software OR "Libre Software") AND (ecosystem OR "Software Supply Network" OR "Software Supply Industry")

3.4. Study selection

The study selection strategy was designed to consist of a set of several steps, which is an adaptation of the steps proposed in Petersen et al. (2015) and Kitchenham and Charters (2007). Fig. 2 presents an overview of the study selection process and the number of publications included in each stage. The details of each stage are described in the following subsections. We excluded articles based on titles and abstracts as well as full-text reading.

3.4.1. Inclusion and exclusion criteria

The following criteria have been used to select the relevant publications:

- only publications in English.
- only papers published between 2003 (publication of the seminal book about SECOs) and 2015.
- only papers about OSSECO topics.

We excluded panels, prefaces of conferences and special issues, book reviews, news flashes, short papers (fewer than 4 pages), and PhD symposium papers, (i.e., publications without bibliographic information, papers that only report work in progress, and non-peer reviewed publications).

3.4.2. Stage 1 - Automatic search

In this stage, we identified a set of publications that serve as a basis for this study. For the selection of digital libraries, we determined a set of representative digital libraries (see Table 4) that cover the publication sources in Section 3.3.1. We executed the search on each digital library⁷ and saved the references in bibliography files. As a result, 1090 primary studies were identified.

Table 4: Digital libraries

Library	Link
ACM Digital library	dl.acm.org
Compendex/Inspec	www.engineeringvillage2.org
IEEE Xplore Digital Library	www.ieeexplore.ieee.org
Scencedirect	www.sciencedirect.com
SpringerLink	www.springerlink.com
Wiley Online Library	onlinelibrary.wiley.com

3.4.3. Stage 2 - Remove duplicates

Duplicate records were resolved in this stage by importing the references to a reference manager system and automatically removing duplicated papers. Finally, one of the authors manually reviewed the list of articles in order to identify duplicated records. A total of 407 papers were excluded in this stage (e.g., Figay and Ghodous (2009) is indexed in both IEEE and ACM digital libraries).

3.4.4. Stage 3 - Titles and abstracts

To identify publications that were indeed about OS-SECOs, all of the authors of this study reviewed all of the titles and abstracts and checked the inclusion and

⁷the last automatic search was made on February 12th, 2017

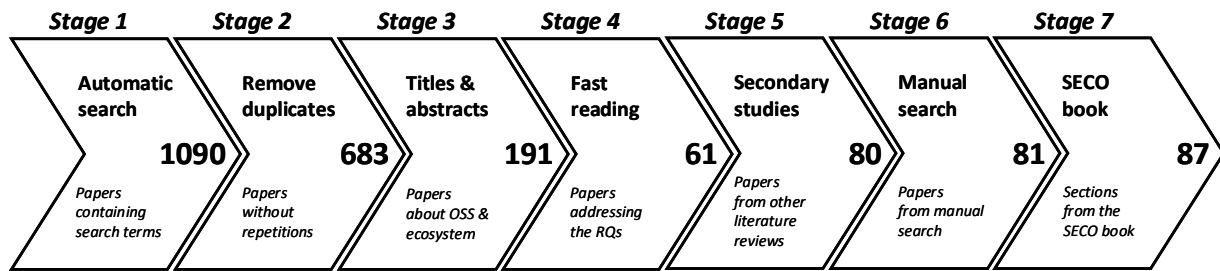


Figure 2: Stages of the selection process

exclusion criteria for each entry. When there was a disagreement, the authors discussed the issues until a consensus was reached. After this stage, 492 out of the 683 remaining papers were discarded, resulting in 191 publications (e.g., [260] was discarded because the title and abstract were not related to OSSECOs).

3.4.5. Stage 4 - Fast reading

Then, in order to filter out the papers from the third stage, the results and conclusions of each study were reviewed, and each researcher briefly studied their contents. Hence, all of the papers that did not reflect the study's topics, did not address any of the research questions, or were delta papers⁸ were excluded (e.g., [112] is delta paper of [R2]). At the end of this stage, 61 papers were selected.

3.4.6. Stage 5 - Secondary studies

Thereafter, in order to identify the maximum number of relevant papers that might have been missed, we reviewed the papers from the seven secondary studies (see Section 3.1). In this stage, we included 19 papers out of the 315 papers referenced by the secondary studies. These papers underwent the same process that we used for the rest of the papers from Stage 2 to Stage 4.

3.4.7. Stage 6 - Manual search

We complemented the search in the digital libraries with some manual searches in order to ensure that we had covered all of the editions of the literature sources listed in Section 3.3.1. One paper was identified using this manual search process (i.e., Morgan et al. 2013).

3.4.8. Stage 7 - The SECO book

Jansen et al. (2013) published their book: *Software Ecosystems Analyzing and Managing Business Networks in the Software Industry*. We applied Stage 2 to Stage 4 to the book chapters and selected six additional relevant studies.

3.4.9. Final result

Finally, after this last stage, the systematic mapping included 87 relevant papers (see Table B.16). These papers were coded with the prefix *R*.

3.5. Data extraction

We mainly used a qualitative data analysis approach based on the method of Miles et al. (2014) to extract the data from the selected studies. This process was conducted with the support of a qualitative data analysis tool called Atlas.ti⁹ to ensure consistent and accurate extraction of the key information related to the research questions. The extraction was performed by one of authors and reviewed and confirmed by the other three authors. We also frequently used consensus meetings to review the extracted data. Having other authors check the extraction is a common practice in systematic reviews for social sciences (Petersen et al., 2015; Hauge et al., 2010). The stages of the qualitative data analysis process were the following:

- *Data processing and preparation*: The 82 studies included in our systematic mapping were grouped into one Atlas.ti[®] hermeneutic unit¹⁰.
- *First cycle, (codes and coding)*: Codes are labels that assign symbolic meaning to the descriptive or

⁸Delta papers include a minimal new contribution with respect to some previous publication. The default selection criteria for delta papers was to select the oldest one, because it usually has the main contribution of the authors. However, we have overridden this rule when some newer paper is more relevant for the purposes of this study.

⁹<http://atlasti.com>

¹⁰A *hermeneutic unit* is an Atlas.ti container where all of the information, links, or paths to this information that are related to a specific project, are stored.

inferential information. They are primarily, but not exclusively, used to retrieve and categorize similar data chunks so that the researcher can quickly find, pull out, and cluster the segments relating to a specific research question (Miles et al., 2014). We defined a list of codes from the research questions (i.e., *deductive coding*). In Section 3.6, we detail the information that we used to define the initial codes.

- *Second cycle, (pattern codes)*: This is a way of grouping the list of codes into a smaller number of categories (i.e., *pattern codes*). These are explanatory or inferential codes that identify an emergent theme, configuration, or explanation (Miles et al., 2014). In Section 4, we describe these categories in the research questions where they were defined.
- *Displaying the data*: The goal of this stage is to condense the major data and findings from our study for further analysis and to represent and present the conclusions. In our study, we used different kind of methods to display the results (e.g., tables and charts).

To answer RQ1.4, we used social network analysis (SNA) because it is useful to assess authors' positions in the social networks (this is detailed in Section 4.1).

Finally, the process was developed based on several rounds of piloting and coding to ensure the validity and consistency of the results. To extract data from the identified primary studies, we developed the template shown in Table 5, which provides the initial codes for the data extraction process.

Table 5: Data extracted from each study

Data item
Source (Conference, Journal, Book chapter) and full reference
Year when the paper was published
Author(s) and their affiliation (organisation and country)
Type of publication
Definition(s), sources and authors of ecosystem term(s)
Elements related to OSSECO: name and type (defined, referenced, used)
Measures, if any, defined to evaluate OSSECOs
Instances, if any, of OSSECOs studied
Ecosystem model(s), if any, used
Scope of the ecosystem model(s) (SECO, BECO, DBECO)
Techniques and notations for modelling OSSECOs
Type of OSSECO analysis

3.6. Data Analysis

The information for each item extracted was tabulated and visually illustrated (see Section 4). Table 6 shows the data that was tabulated to answer the research questions.

Table 6: Data tabulated per research question

Data item	RQ
Number of papers per source	RQ1.1
Number of relevant publications per year	RQ1.2
Number of papers per country	RQ1.3
Social network measures (e.g., betweenness centrality)	RQ1.4
Number of papers of academy and industry	RQ1.5
Number of papers per type (e.g., experience report)	RQ1.6
Number of papers per type of ecosystem definition sources	RQ2
The sources of ecosystem definitions	RQ2.1
Number of papers per ecosystem definition	RQ2.1
The ecosystem concept definitions	RQ2.1
The definitions of OSSECOs	RQ2.2
Number of papers per ecosystem terms	RQ2.3
Number of papers per OSSECO instances	RQ2.4
Number of papers using models to represent OSSECOs	RQ3.1
The type of SECO modelled	RQ3.2
Number of papers per modelling technique	RQ3.3
Identify the type of OSSECO analysis	RQ3.4

4. Results

This section summarizes the results obtained from the data extraction process.

4.1. RQ1. What are the demographic characteristics of the studies about OSSECOs?

To answer this question, we applied the process defined in Section 3.5 without the pattern code cycle.

RQ1.1 In which type of sources are articles mostly published?

The distribution of the 87 primary studies is shown in Fig. 3. According to our data, conference proceedings (with 45 papers) are the most prevalent publication type. Table B.16 (in Appendix B) shows the publication type for each paper.

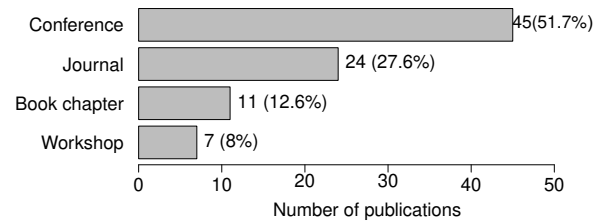


Figure 3: Publication type

RQ1.2 How has the number of publications evolved over the years?

We searched for primary studies between the years 2003-2015. We found the first papers published about OSSECOs in 2006 [R5, R6, R21]. Fig. 4 shows the number of papers per year.

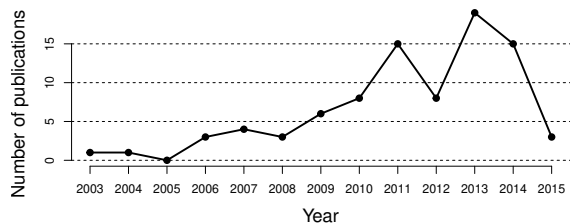


Figure 4: Publication year

RQ1.3 How are papers geographically distributed?

We determined the geographical distribution of the papers based on the country of affiliation of the first author. Europe (59 papers) is the most dominant continent with the Benelux countries (24 papers) and Scandinavia (9 papers) being the most active regions (see Section 5.1 for details). North America is next with 23 papers. There are few publications from Asia (3 papers, from Japan). Publications from other continents are scarce.

RQ1.4 Who are the predominant researchers?

We addressed this question by conducting a social network analysis (SNA), which allowed us to do the following: 1) to identify individual nodes that are of particular interest (i.e., relevant authors); and 2) analyze the whole graph and identify cohesive subgroups (i.e., authors' clusters)¹¹. This analysis was done only for the authors and coauthors of papers in our set of primary studies.

Identify the predominant researchers. We used an approach similar to Jabeur et al. (2010) to evaluate the authors' relevance SNA. In that work, they proved that centrality measures are the best ones to assess the social significance of a cluster of authors. According to their work, the social model is represented by a non-directed graph $G = (V, E)$, where V nodes correspond to authors. The set of edges $E \subseteq V \times V$ represents the social relationships connecting authors. First, we identified 151 researchers from the papers. Second, we identified the set E of edges as follows: 1) $(a_i, a_j) \in E$ if $a_i, a_j \in V$ and a_i, a_j have coauthored a paper; 2) $(a_i, a_j) \in E$ if $a_i, a_j \in V$ and author a_j is cited by a_i . Finally, we calculated the following measures to rank the authors:

- The *betweenness centrality* for a node N is the sum of the fractions of shortest paths that include N for every pair of nodes in the network. If a high betweenness node is removed, a number of links may

get disconnected (Toivanen et al., 2015). This measure quantifies the ability of a node to act as a mediator in the network (Kamei et al., 2008).

- Mathematically, *eigenvector centrality* is the first eigenvector of the adjacency matrix. The main principle is that links from important nodes are worth more than links from unimportant nodes (Golbeck, 2013). High eigenvector centrality nodes can be leaders of the networks (Toivanen et al., 2015). This measure scores nodes based on the principle that relationships with more important nodes confer more importance than relationships with less important nodes (Goeminne and Mens, 2010).
- *PageRank* measures the importance of a node within the network using a link analysis algorithm. It can be calculated using a simple iterative algorithm and corresponds to the principal eigenvector of the normalized link matrix of the network (Page et al., 1999). This measure score distinguishes the authority of each author in the social network (Jabeur et al., 2010).

Table 7 lists the top 10 authors ranked using these measures.

Table 7: Top authors ranked by social relevance

Betweenness centrality		Eigenvector centrality		Page rank	
Slinger Jansen	3519,859	Slinger Jansen	0.039	Slinger Jansen	6,087
James Herbsleb	1938,617	Sjaak Brinkkemper	0.030	Sjaak Brinkkemper	4,094
Tom Mens	1051,622	Tom Mens	0.026	James Herbsleb	3,905
Sjaak Brinkkemper	958,226	James Herbsleb	0.025	Tom Mens	3,119
Donald Wynn, Jr	892,983	Mathieu Goeminne	0.025	Daniel M. German	2,826
Daniel M. German	823,393	Daniel M. German	0.023	Mathieu Goeminne	2,682
Mathieu Goeminne	783,677	Walt Scacchi	0.022	Walt Scacchi	2,420
Lopez-Fernandez	584,000	K. Manikas	0.020	Brian Fitzgerald	2,157
Walt Scacchi	545,873	K. Hansen	0.020	Donald Wynn	2,150
Brian Fitzgerald	467,957	L. Luinenburg	0.018	Mircea Lungu	2,123

Cluster of authors. As in the previous case, we used the social model represented by a graph $G = (V, E)$, where nodes correspond to authors. However, in the current case, the set of edges E only connects two authors when they are paper coauthors in at least one publication. Then, we used the Newman (2001) algorithm implemented in NodeXL to identify authors clusters. (i.e., a set of at least two authors who collaborated on at least one publication). Table 8 summarizes the authors and coauthors network-wide measures.

Fig. 5 shows the three most populated clusters identified. The graphs highlight the top authors in each cluster (according to the SNA measures) and the number of relationships between coauthors.

¹¹We used a tool named NodeXL to perform the network analysis (Smith, 2014).

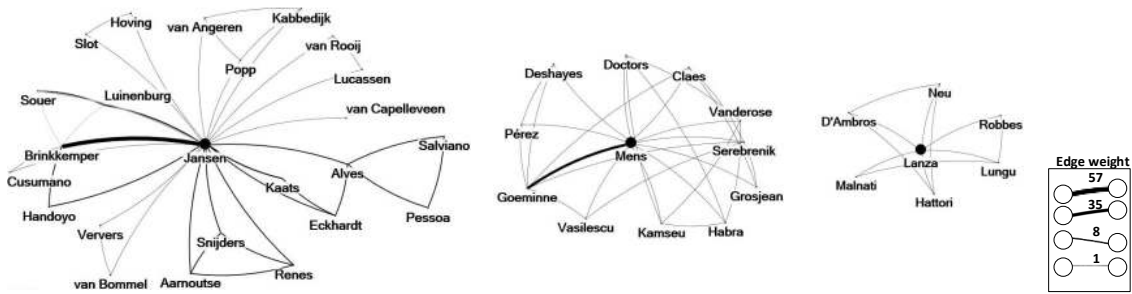


Figure 5: Research clusters (edge weight is the number of relationships)

Table 8: Authors and coauthors overall graph measures

Measure	Value
Number of Vertexes	180
Number of Edges	256
Number of Clusters	42
Maximum Edges in a Connected Component	40
Maximum Vertexes in a Connected Component	24
Graph Density	0.016

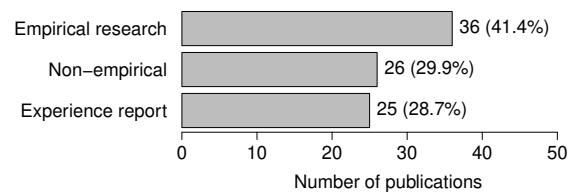


Figure 6: Type of research

4.1.1. RQ1.5 How are publications distributed between academy and industry?

In order to answer this question we analyzed whether at least one of the authors in each paper came from a non-academic institution (similarly to Ameller et al. (2015)). A total of 25 papers (28.7%) fall into this category, while 62 papers (71.3%) have authors solely from academy. We found that two papers [R18, R20] are exclusively from industry.

RQ1.6 What type of papers are published?

To answer this question, we classified the publications into three categories, similarly to Hauge et al. (2010) and Montesi and Lago (2008): (R) empirical research papers, where the authors present evidence from a research study having an explicit research question; (E) experience reports, where the authors report experiences without having defined an explicit research question; and (N) non-empirical papers, which include opinion papers and theoretical papers. Fig. 6 presents the number of papers of each type. The classification for each type of study is shown in Table B.16 in Appendix B.

4.2. RQ2 What is an OSSECO?

To answer this question, we applied the process defined in Section 3.5.

RQ2.1 What definitions are related to the OSSECO definition?

We found that 76 papers out of the total of 87 use ecosystem-related definitions based on the five different concepts introduced in Section 2.2.1 shown at the top of the Fig. 7.

Fig. 7-a shows the percentage of references of each ecosystem definition (calculated on the 76 papers that used ecosystem definitions). Fig. 7-b shows the list of the papers classified by ecosystem definition.

Table 9: OSSECO definitions

Definition
"An arrangement of individual and organizational units, involved in or affecting the circulation, transformation, and accumulation of capital (in various forms) in order to provide cooperative development, testing, marketing, distribution, implementation, and support of open source software." Wynn (2007) [R35]
"An OSS ecosystem is one where it is possible to add contributions to a project, create and publish components in the extension market, etc., without any barriers. Jansen et al. (2013) [R48]
"A set of developers functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. The result of the interaction is freely available for everyone." Hoving et al. (2013) [R65]

Fig. 7 shows that the SECO definition is the one that is most studied in the selected papers (36 out of the 87 studies), the BECO definition is the second most studied (22 studies), the OSSECO definition has eight pa-

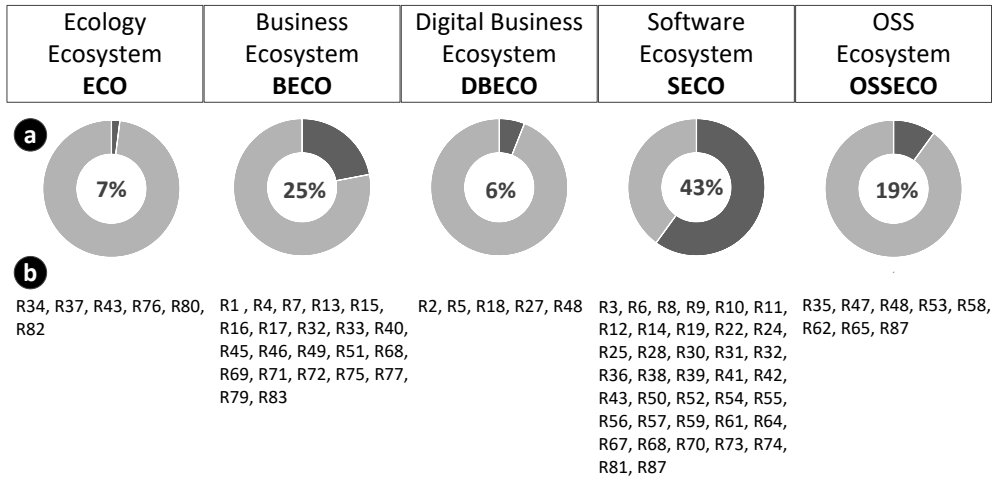


Figure 7: OSSECO classification

pers mentioning it, and ecosystem and DBECO¹² are the two least mentioned definitions (6 and 5 papers, respectively), Finally, 11 studies did not fit any of the classifications (i.e., [R20, R21, R23, R26, R29, R44, R63, R84, R85, R86, R88]).

RQ2.2 Are there specific definitions of OSSECO?

We identified that there have only been a few attempts in the literature to specifically define what OSSECO is. Specifically, in this mapping study, we obtained only eight papers out of the total of 87 papers that use the OSSECO definition (see Fig. 7). Only three of them give a definition of OSSECOs. On the other hand, 49 papers based their work on definitions related to the ecosystem-related definitions (i.e., BECO, DBECO, SECO); 17 of these papers also provide definitions related to ecosystems in their own words. The three definitions of OSSECO correspond to the papers [R35, R48, R65] (see Table 9).

RQ2.3 What elements belong to an OSSECO?

To answer this question, we applied the process defined in Section 3.5. First, we collected 64 related terms in the coding cycle belonging to OSSECOs¹³. among which *project*, *community*, and *developer* are the top three most used terms. Second, in the pattern codes

cycle, we classified the OSSECO terms into three categories, based on the type of term used in the study (i.e., the term was defined by the author), the article references to another author (in order to include the term definition), and the term is used in the article but is not defined. Table 10 shows the terms and the number of papers per category.

Table 10: OSSECO terms

Term	Own	Other	Use	Term	Own	Other	Use
Project	9	2	66	Survey	0	0	18
Community	27	5	42	Author	0	5	10
Developer	8	5	51	Keystone_player	7	6	2
Platform	8	2	49	Node	10	1	5
Source.Code	0	0	59	Integrator	4	1	9
Contributor	18	1	34	Adopter	5	0	7
Product	6	1	48	Artefact	1	1	10
Service	6	0	47	Niche_Player	4	4	3
Repository	9	0	37	Practitioner	3	1	7
Feature	4	3	39	Behavior	0	0	9
Market	3	4	37	Resseller	1	0	8
Bug	0	0	43	Email	0	0	8
Reviewer	3	5	32	Platform_Provider	3	1	4
Roadmap	2	0	37	Active_User	3	0	4
License	10	1	25	Transactions	1	0	6
Partner	5	0	31	Coordinator	2	0	5
Mailing_List	4	0	33	IRC	0	0	6
Foundation	12	6	16	Bug_Fixer	0	4	1
Measure	1	5	28	Passive_User	1	0	4
Dependency	7	0	27	Dominator	3	0	1
Member	6	2	26	Sub_Community	1	0	4
Actor	10	3	20	Vocabulary	1	0	3
Stakeholder	4	1	25	Community_Manager	0	0	3
Bug_Tracking	1	0	29	Forge	2	0	1
Commit	3	1	24	Bug_Reporter	0	2	1
Event	4	1	23	Entropy	1	0	2
Goal	3	0	21	Translator	1	1	1
Niche	15	2	6	Committer	0	0	1
Boundary	2	2	17	Configurator	0	0	1
Social_Network	7	2	13	Supplier	0	0	1
Edge	16	0	5	Super_Repository	1	0	0
Data_Source	4	0	16	Wishlist	1	0	0

RQ2.4 What instances of OSSECOs have been reported in the literature

To answer this question, we identified the specific OSS communities studied in each paper. We found that 49 papers out of 87 studied specific OSS communities.

¹²The papers [R2, R48] are classified in this definition; however they use the concept of digital ecosystem instead of DBECO.

¹³The list of all the terms and definitions that we found in the primary studies are in <http://www.essi.upc.edu/~gessi/PLATEOSS/>.

Most of them studied the Eclipse ecosystem (16 papers) and the GNOME ecosystem (10 papers). The rest of the OSS-communities were studied by only one or two papers (except Ubuntu and Ruby, with 3 papers each one). Table B.17 (in Appendix B) shows the name and the sources of all the OSSECO instances studied.

4.3. RQ3. Which representations have been proposed for OSSECOs?

To answer this question, we applied only the first cycle of the process explained in Section 3.5. We used the codes defined for RQ3 in the data extraction section (see Table 5).

RQ3.1 Which primary studies use models to represent OSSECOs?

Fig. 8 shows that 56 papers (64.4%) of the 87 studies in the collected literature use models to represent the actors, resources, and their relationships in the specific OSSECO under study.

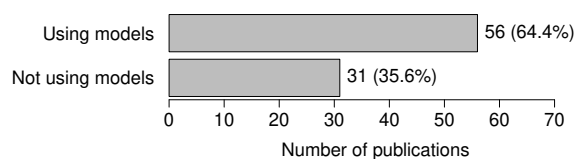


Figure 8: Use of models

RQ3.2 Which of the proposed models, if any, are specific for OSSECOs?

According to our review, none of the 56 studies using models develops a specific technique for modelling OSSECOs although most of them studied OSS communities. Fig. 9 shows the different definitions of these studies. The SECO and BECO definitions are the most frequently used ones. In contrast, the DBECO definition is used in two papers, and the ecosystem definition is used in only one paper. Table B.18 in Appendix B shows the type of ecosystem definition of each paper.

RQ3.3 Which notation and guidelines have been used for modelling OSSECOs?

To answer this question, we applied the second cycle of the process explained in Section 3.5 to the results of the RQ3.1. In the 56 papers that use models, we found several modelling techniques to describe or visualize software ecosystems: conceptual maps (e.g., R2, R15, R47, R69), tabular representations (e.g., R35, R52, R62, R79), mathematical notations (e.g., R6, R31,

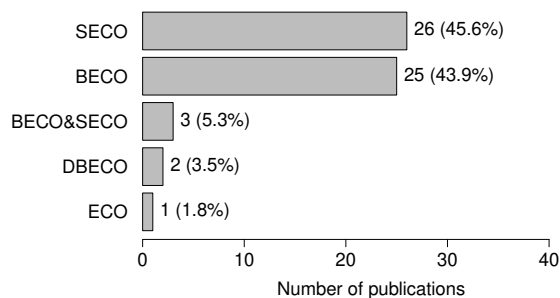


Figure 9: Use of models by ecosystem definition

R34), metamodels (e.g., R12, R16, R39), social networks (e.g., R8, R25, R59, R77, R86, R87), class diagrams (e.g., R36, R46, R65), iStar (e.g., R72, R83), and also ad hoc notations (e.g., R22, R27, R57, R75, R88). When a paper used more than one type of modelling technique, we selected the dominant one. Fig. 10 depicts examples of OSSECO models according to each type of modelling technique. Fig. 11 shows the distribution of papers by modelling technique. It shows that ad hoc notations (31 papers) are predominantly used to model OSSECOs. Table B.18 (in Appendix B) lists the modelling technique used for each paper and the goal pursued by the model.

RQ3.4 What type of analysis was conducted using the models identified in RQ3.3?

We classified the papers that use OSSECO models into four categories (i.e., social network analysis, statistical analysis, visual analysis, and mathematical analysis). This classification is based on the approach proposed by Jansen et al. (2015).

Fig. 12 shows the distribution of selected studies that use models in the type of OSSECO analysis. twenty-three of these studies (41.1%) do not show any evidence of analysis.

5. Discussion

In this section, we discuss each of the answers to our research questions. For the analysis of some of the results, we performed a correlation analysis between all the codes used to answer the research questions. However, in this study we only considered the statistically significant correlations¹⁴. The independence test used in this paper is Fisher's Exact Test for Count Data. In our study, all the contingency tables were small enough to run Fisher's test in a reasonable time.

¹⁴They are considered statistically significant when their *p-value* is less than 0.05.

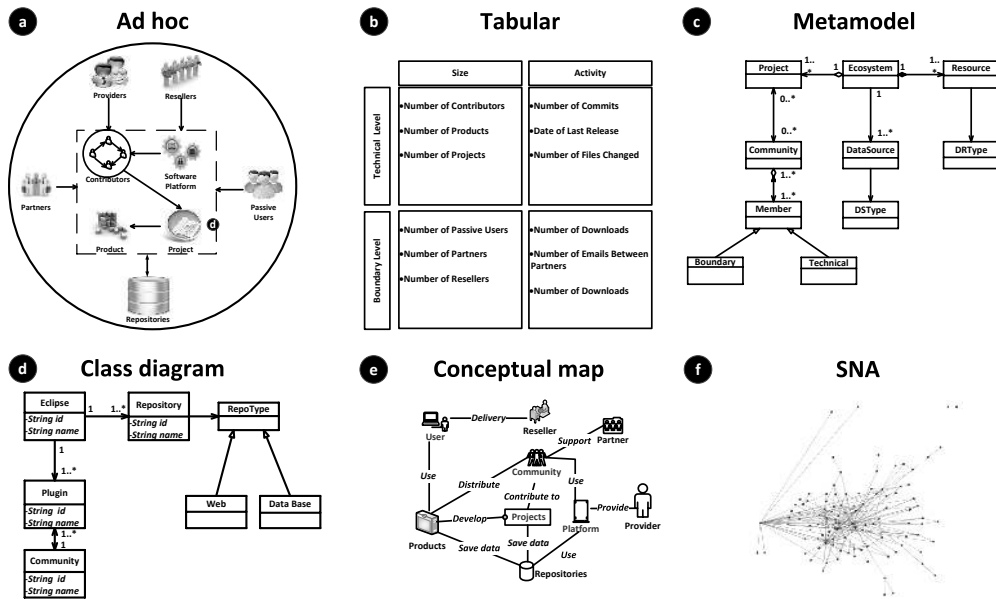


Figure 10: Examples of modelling techniques: OSSECO models a, c, d, and b are static; model f is an OSSECO dynamic model; and figure b shows a level-oriented framework model.

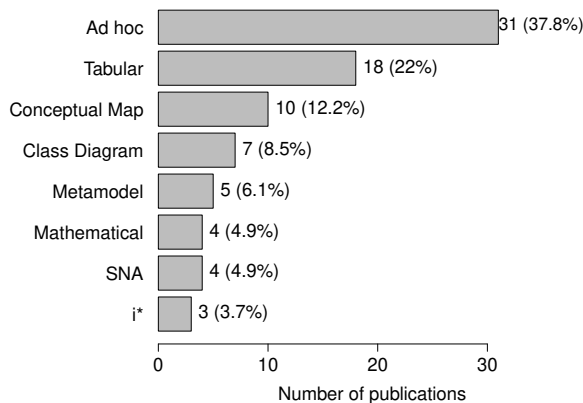


Figure 11: Modelling techniques and notations

5.1. RQ1. What are the demographic characteristics of the studies about OSSECOs?

RQ1.1 In which type of sources are articles mostly published?

Fig. 3 in Section 4.1 shows the distribution of the primary studies per publication type. In this section, to analyze this distribution, we compare it with that of the general context of publications in software ecosystems. In order to do this, we used the results reported in the secondary studies: Hanssen and Dybå (2012); Barbosa et al. (2013); Manikas (2016a); Axelsson and Skoglund (2016). Fig. 13 shows that the percentage of publica-

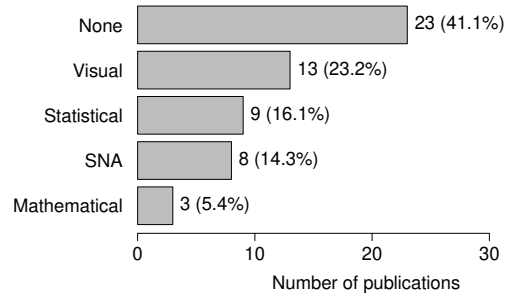


Figure 12: Analysis techniques

tions in journals is quite significant, and most of them are from journals with high impact factors such as IST and JSS (i.e., 17 studies from Manikas and 6 studies from these journals). Similar to Manikas (2016a), we think that this is beneficial for the maturity of the SECO and OSSECO fields.

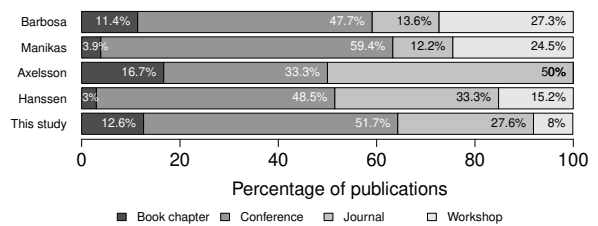


Figure 13: Comparison of types of publications

RQ1.2 How has the number of publications evolved over the years?

OSSECO is a growing research area in software engineering. Fig. 14 shows a significant increasing trend in the number of publications related to OSSECOs with 56.3% out of the 87 papers studying and analyzing OSSECOs. Furthermore, since 2006 there has been a regular increase in the number of publications each year, with the exception of years 2012 and 2015, which does not significantly affect the overall trend. In addition, we have witnessed the emergence of a research community that shares interest in OSSECOs: IWSECO is an international workshop on SECOs with several publications on OSSECOs (e.g., Syed and Jansen (2013); Spauwen and Jansen (2013); Van Lingen et al. (2013)), tutorials in relevant conferences like ICSE (Bosch, 2012), specialized workshops such as WEA (workshop on software ecosystem architectures), and special issues about SECOs¹⁵ in journals (e.g., IST¹⁶, JSS¹⁷).

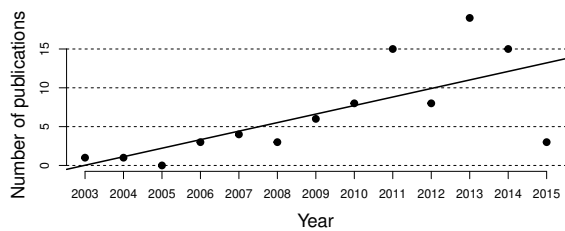


Figure 14: Publication trend

RQ1.3 How are papers geographically distributed?

We have put the results of our study in a general context of publications. In this case we use the context of computing science in the period 2006-2015 as presented in the bibliometric indicator database of the SCImago journal & country rank (SCImago, 2015) (see Fig 15). It is no surprise that European and North American authors are the dominant researchers. However, in our study, the percentage of publications from Europe is significantly higher than in the SCImago database (67.8% and 35.7%, respectively). This could be due the increasing research on the OSSECO topic in some countries (e.g., The Netherlands and Belgium). On the other hand, the number of publications from Asia is surprisingly and significantly lower (38.7% in the SCImago database and 5.8% in our study).

¹⁵Several papers of these special issues are about OSS

¹⁶<http://www.sciencedirect.com/science/journal/09505849/56/11>

¹⁷<http://www.sciencedirect.com/science/journal/01641212/85/7>

In the distribution of papers in Europe, The Netherlands and Belgium are the countries with the most publications (25.4% and 13.6%, respectively). These values are corroborated by the countries of affiliation of the dominant researchers. This highlights the fact that in a relative, new discipline such as OSSECOs, leading research groups can create predominant niches in a specific research area, as it happened with Jansen and Brinkkemper’s research group from the The Netherlands and Mens and Goemine’s research group from Belgium.

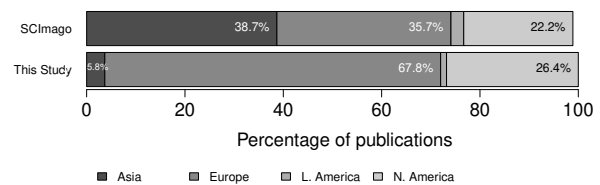


Figure 15: Comparison of publications per continent

RQ1.4 Who are the predominant researchers?

Table 7 shows that five of the predominant researchers are included in all of the top 10 lists. Brinkkemper, Herbsleb, Jansen, Mens and Goemine are key entities in the social networks (i.e., keystone actors, network brokers, etc). This is due to their strategic position in the social networks of authors. The measures used in this work highlight the authors connecting dispersed partitions of the OSSECOs researchers. Thus, we can identify that there are clusters (i.e., sets of authors collaborating together) around the main researchers. This would mean that amount of the research on the OSSECO topic is growing around these authors and their approaches. Also, this cluster enables independent authors to come together as a larger social network of shared knowledge about OSSECOs.

Some authors like Jansen, Mens and Goemine have several publications about OSSECOs (i.e., 17, 9, and 7, respectively). This may explain the clusters around them. The number to citations to these publications explains the high values of their measures in Table 7. In contrast, other authors like Herbsleb and German, with high values in Table 7, are not in main clusters because they have only two papers in our set of primary studies. On the other hand, Lungu is on one of the predominant authors lists because of the number of references to his publications (23). Also, he is in one of the main research clusters because he is a coauthor of Lanza, who is one of the main nodes in his cluster.

Graph density has a value between 0 and 1 and

describes how interconnected a network is (Golbeck, 2013). Table 8 shows that the author-coauthor network density is very low (only 1.6%). This suggests that most of the authors only have a high density relationship only with a small number of other authors. It is also an indicator of low network cohesion and membership. Other studies about the measures of author-coauthor networks have similar density measure result (e.g., Li et al. 2010; Abbasi and Altmann 2011; Cheong and Corbitt 2009). This could be due to the youth of the field, and it could mean that it is a challenge to grow partnerships inside the OSSECO research community. In addition, it is necessary to find brokers that connect dispersed clusters.

RQ1.5 How are publications distributed between academy and industry?

Fig. 16 shows that in the period 2006-2016 for both the context of publications in computing science (Ruíz, 2012)¹⁸ and for this mapping study, the great majority of the papers are from academy. It is no surprise that academics are clearly more motivated to submit papers to journals and conferences. This is particularly true in the OSSECO domain where researchers are more interested in abstract concepts and definitions than practitioners, who are more attracted by practical questions. However, the number of papers from industry indicates that OSSECO is a topic of interest from the industrial perspective. We found a correlation between *Ecosystem definitions* and *Type of papers*: 53 (66%) out of the total of 80 papers that use ecosystem definitions are from academy ($p=0.031$).

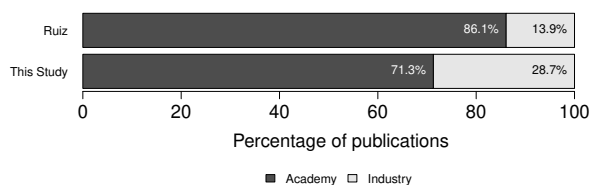


Figure 16: Comparison of affiliation

RQ1.6 What type of papers are published?

To analyze the distribution of the type of papers found in our study, we contextualized our results with Hauge et al. (2010), which is a SLR about adoption of OSS in software-intensive organizations. Fig. 17 shows that there are no remarkable differences in the distribution of papers between our study and Hauge et al.'s study.

¹⁸To obtain these values we contacted the author of the paper and asked him to provide the updated information.

This is an interesting fact because each type of paper contributes differently to the research community (Chen and Hirschheim, 2004). While non-empirical studies help to develop concepts and build theory, empirical studies provide concrete evidence for testing theories. For instance, on the non-empirical study side, we have: [R29], which describes the use of active theory in OSSECOs; and [R35], which proposes a conceptual framework to evaluate OSSECO's health. On the other hand, on the empirical study side we have: [R48], which makes a survey on SECO governance; and [R36], which makes a survey on SECO associated models. Fig. 18 shows that neither of the top continents, Europe and North America, have more industrial papers than academic ones. Finally, experience reports provide examples of the use of theories in this side, we have: [R39, R42], which visualize the GNOME dynamism; and [R4], which shows the OSAMI-Commons project that defines a cross platform of an open-service ecosystem.

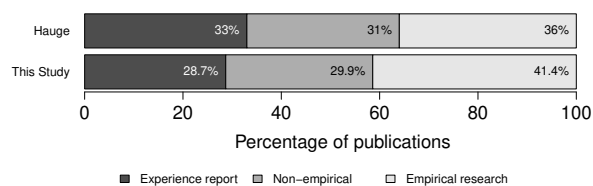


Figure 17: Comparison with respect to the type of research

In contrast to other mapping studies, we did not find a correlation with $p < 0.05$ between types of papers and continents. However, we did find a correlation between *Publication year* and *Paper type*. The number of empirical research papers has been increasing (4 between 2003-2008 and 30 between 2009-2015) ($p=0.025$). This can be interpreted as a sign of increasing maturity of the OSSECO field (Manikas 2016a).

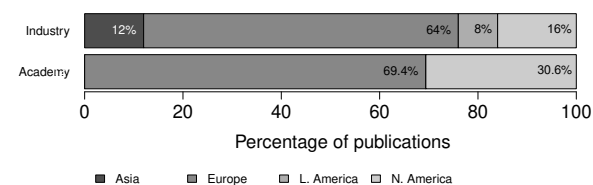


Figure 18: Affiliation per continent

5.2. RQ2. What is an OSSECO?

RQ2.1 What definitions are related to the OSSECO definition?

To discuss this RQ, we split this section into two parts. In the first part, we define a genealogical tree of the definitions that are related to the OSSECO definition. In the second part we analyze the common elements across these definitions.

OSSECO related definition evolution.

In order to clarify the relationships between the definitions related to OSSECOs and to contribute to the understanding of the OSSECO phenomenon, we wanted to picture their chronological evolution. Moreover, we attempted to depict the research in this field that we found in our mapping study. Thus, we built a genealogical tree with the ecosystem definitions, their relationships, and their predominance in the OSSECO community (see Fig. 19). The figure can be read as follows: (a) from left to right, the figure shows the evolution of the OSSECO definition over time; (b) from right to left, the figure shows the inheritance relationships between the different ecosystem authors definitions; (c) from top to bottom, the figure shows the evolution of each ecosystem definition; and, (d) each node in the figure shows the first author and the number of citations per publication. This number corresponds to the papers (from our set of primary studies) that cited that publication for the definition used in their research work. This means that references for other purposes were not taken into account.

Figure 19 shows that SECO is the most frequently referenced definition in our set of primary studies. Furthermore, there are several references to the BECO definition. Wynn [R34] references the BECO and DBECO articles in his paper and Jansen et al. [R40] references the SECO, BECO and DEBECO definitions. This indicates that the OSSECO is a specialization of these definitions.

The genealogical tree shows that there are several direct and indirect relationships among all of the definitions related to OSSECO. In our systematic mapping, we found few papers that try to adapt ecosystem theories to the OSSECO domain (i.e., [R34] uses the entropy concept and [R82] uses the predators and prey concept). The rest of the papers simply use the ecosystem definitions (i.e., BECO, DBECO, SECO, OSSECO) to identify the actors, the relationships, and the specific environment of a specific OSSECO (e.g., [R8] for Nagios, [R12, R87] for Ruby, [R17] for Eclipse). Furthermore, we found three papers that use the *health* metaphor to analyze OSSECOs (i.e., [R52, R56, R62]). However,

similarly to Mens et al. (2014), we did not find the application of theories, models, or ideas from ECOs to the domain of OSSECOs, despite the fact that ECOs have been studied for many decades. It is a challenge for OSSECO researchers to transpose theories and ideas from ECOs (e.g., systems dynamics modelling, general system theory) to OSSECOs.

Common elements.

Fig. 20 shows a conceptual map that represents the relationships between the five OSSECO related definitions and their terms. We found that there are common elements across definitions:

- A community of actors (i.e., complex organisms in ECO, business world organisms in BECOs and DBECO, and collections of products, projects, software solutions, and businesses in SECOs and OSSECOs).
- A set of relationships.
- An environment (i.e., economic communities in BECOs, open socio-technical systems in DBECO, shared market and technological platforms in SECOs and OSSECOs).

The ecosystem metaphor is useful for explaining the dynamics of complex systems such as business, digital, and software systems. The software ecosystem metaphor was coined 13 years ago by Messerschmitt and Szyperski (2003), reflecting and incorporating software technology into BECO. However, we only found one study that discusses the metaphor in depth (Mens et al. 2014). Most of the papers have only adopted common definitions of SECO or related definitions (see Section 4.2). In our opinion, in the near future, most SECOs, BECOs, and DBECO will be more open to become closer to OSSECOs and share some of their features. This is because SECOs are strongly related to BECOs and DBECO and *openness* is not only a desirable characteristic of SECOs but a vital characteristic as well. Furthermore, every software platform at the centre of an ecosystem has to have some degree of openness (Hoving et al. 2013).

Finally, we find that there is currently a consensus among SECO researchers for two SECO definitions: business-centric definition of Jansen et al. (2009b) and the platform-centric definition of Lungu (2008). In our opinion, a commonly accepted definition of SECO is important in order to improve the communication between SECO researchers and practitioners and thereby reduce the subjective and ambiguous notions of SECOs.

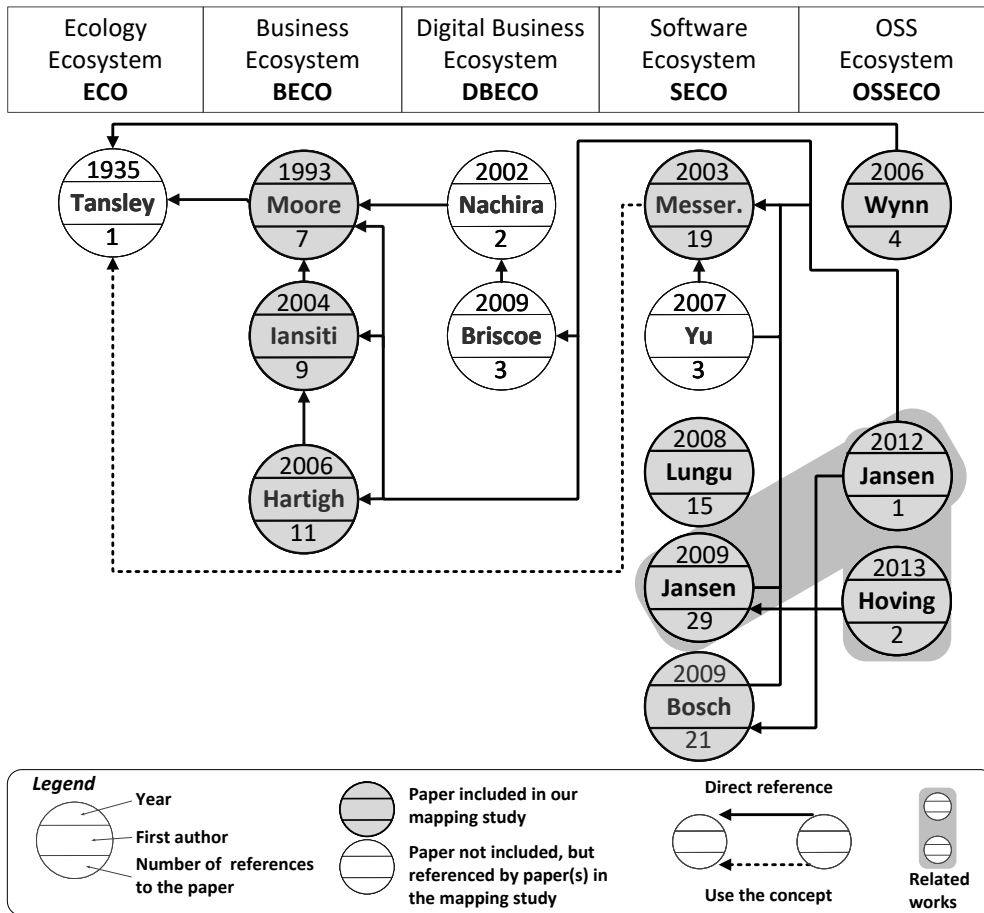


Figure 19: Evolution of OSSECO definition

RQ2.2 Are there specific definitions of OSSECO?

OSSECOs are understood from two perspectives: (1) an ecosystem perspective, where OSSECOs are a network of actors, organizations and companies with symbiotic relationships that can be studied from a business-goal point of view; (2) a project-community perspective that focuses on technical and social aspects of a set of software projects and their communities [R53, R68]. We found that the three main authors of the clusters study OSSECOs from a project-community perspective. However, in their most recent work, they make a call to action for future research in OSSECOs from an ecosystem perspective [R53, R56]. Table 11 shows the classification of the three main author-coauthor clusters and the two OSSECO perspectives.

Some authors argue that OSSECOs are probably the most complex type of SECO [R74]. However, we found very few definitions of OSSECO in our study. Specif-

ically, in the definitions of OSSECO (see Table 9), we found the following as common elements:

- A set of heterogeneous units (e.g., organizations, software projects and services).
- Symbiotic relationships among units (e.g., capital, projects, components).
- An open perspective in a shared market (e.g., to provide support to OSS, to add contributions without barriers, to provide freely available results for everyone).

We did not find any explicit reference to OSSECO communities in the definitions. This is surprising because the OSSECO communities are one of the most important differentiators between OSSECOs and other types of ecosystems.

Finally, we distilled an OSSECO definition (see Section 6.1) because the three above-mentioned definitions

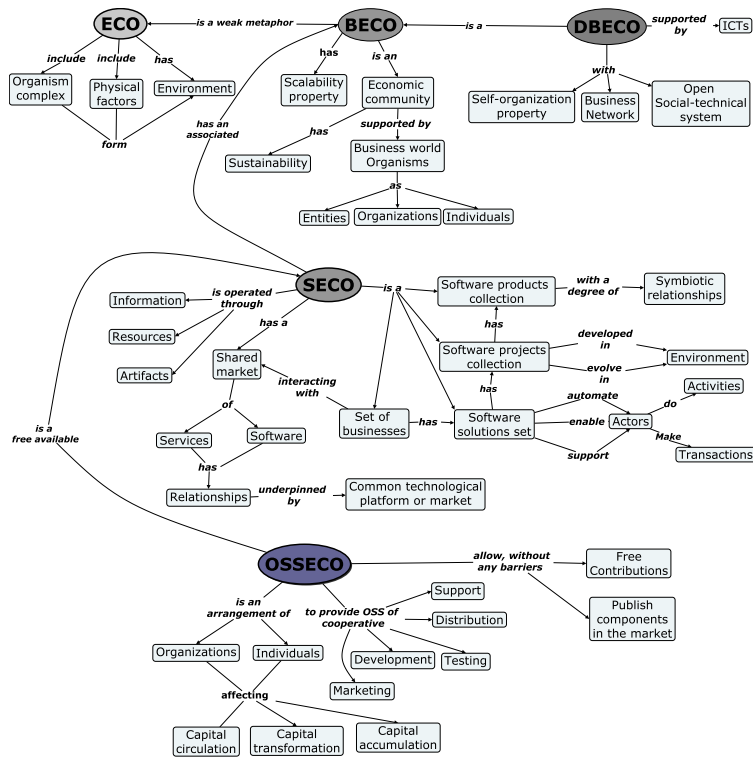


Figure 20: Conceptual map of OSSECO

(see Table 9) have a lack of specificity in the particularities of OSSECOs (i.e., OSS community, open-common platform).

RQ2.3 What elements belong to an OSSECO?

In order to validate our manual content analysis approach to collect data for answering RQ2.3 (see Section 4.2), we compared its outcome (65 terms belonging to the OSSECOs identified) to that of a computational approach. To this end, we used a text-mining approach based on co-occurrence and term frequency analysis as defined by Salton and Buckley (1988). In their work, the content of a document is represented as a vector space, i.e., $D = (w_1, w_2, \dots, w_k)$ where w_k represents the weight of term k in document D that is calculated upon the term occurrences (tf) and the inverse document frequency (idf)¹⁹. This method allowed us to identify the importance of each term in the corpus²⁰. Different terms have different importance in a text, and so w_i is an indicator that represents how much the term t_i contributes to the

semantics of document D . This approach is different from the one described in Manikas (2016a), who identifies the keywords of the set of papers. However, he took these words from the keyword field of each paper. We are taking the terms from the entire text of the paper.

In order to get w_i and compare our terms with the most weighted terms in the corpus, we used the R text mining package (Feinerer, 2015) and followed the steps from Narang (2015): (a) we obtained a document term matrix of 23617 columns (i.e., terms) and 87 rows (i.e., documents); (b) we calculated the weight for all terms in the document term matrix as defined by Salton and Buckley (i.e., $tf_{i,j} \cdot idf_i$); (c) we sorted the list of terms by weight; and (d) we searched the position of each of the 64 terms found in Section 4.2 on the list of sort weighted terms. Table 12 shows the distribution of the number of OSSECO terms across the weighted interval ranking.

Table 12 shows that 48% of the OSSECO terms that we manually identified appear in the top 100 of the weight matrix terms. It also reveals that 77% of these terms are among the 500 most ranked terms in the corpus. This may indicate that the use of well-known terms is significant in the OSSECO research community. We

¹⁹The idf varies inversely with the number of documents N to which a term is assigned.

²⁰The corpus is a set of documents on which to perform the text analysis.

Table 11: Classifications of authors clusters

Cluster	Ecosystem perspective	Community perspective
Jansen et al.	Implements frameworks for OSSECOs health measuring [R62].	Provides studies of data repositories of particular OSSECOs like Ruby and Debian [R12, R30]. Gives a set of models for ecosystem governance and OSSECO enterprise [R36, R41].
Mens et al.	Provides a framework for analyzing OSSECO communities [R53].	Implements tools for modelling OSSECO communities and projects [R10, R19]
Lanza et al.	Provides a view of SECOs as a collection of software projects developed within and across organizational boundaries[R39].	Implements tools for visualizing OSSECO projects [R39]

Table 12: Distribution of OSSECO terms

Interval rank	OSSECO terms	%	% Accum
1-100	31	48%	48%
101-200	9	14%	62%
201-300	5	8%	70%
301-400	3	5%	75%
401-500	1	2%	77%
1000-2000	5	8%	85%
2001-5000	4	6%	91%
upper 5000	3	9%	100%

found that 70 terms in the top 100 of the weight matrix terms do not appear on our list of the 64 terms identified. The reason is that they are mostly common terms in the software engineering domain or general words (e.g., *syntax*, *error*, *analysis*, *software*, *systemic*, *component*, etc.).

In order to analyze the OSSECO terms below the rank of 500 (e.g., *wishlist*, *vocabulary*, *entropy*, *bug reporter*, *sub community*, *adopter*, *IRC*, *bug fixer*, and *passive user*), we calculated the *keyness*²¹ of the 64 OSSECO terms. To this end, we used the Scott and Tribble approach (Scott and Tribble, 2006) to calculate *keyness* using log-likelihood tests. This is a statistical function used for comparing word frequencies of linguistic features in two or more corpora (Rayson et al., 2004). In this work, the OSSECO corpus is the sub-corpus, and the corpus academic vocabulary list of contemporary academic English (consisting of 190.000 documents) was used as a reference corpus (Davies, 2015).

We found that there is a large disparity in values. The term with the highest *keyness* value is project (31694) and the one with the lowest *keyness* is super repository (33.76). All of the terms in the group with the lowest weight are among the 20 ones with the lowest *keyness*. However, all of the OSSECO terms are *pos-*

²¹Keyness is a term used in linguistics to describe the quality a word or phrase has of being key in its context. Keywords are items of unusual frequency in a given sub-corpus in comparison with a reference corpus Scott (1997).

itively key, meaning that they occur more often than would be expected by chance in comparison with the reference corpus (Taljard and De Schryver, 2002). In other words, the OSSECO research community tends to overuse the terms related to OSSECOs more than the academic community in general. This may indicate that the research community is able to create a common vocabulary, which could represent a first step towards an ontology of OSSECOs. In Section 6, we present a taxonomy as a first step towards such an OSSECO ontology. We think that the OSSECO ontology is necessary in order to allow semantic interoperability between the distributed and heterogeneous OSSECO actors.

RQ2.4 What instances of OSSECOs have been reported in the literature?

Table B.17 (in Appendix B) shows that most of the papers found in this review are about OSSECO instances. This is because OSSECOs have several kinds of data sources such as: project sites, ecosystem hubs, and aggregation sites (German et al., 2013). These data sources are freely available and tend to contain the entire history of all OSSECO projects, community relationships, and their artefacts. In addition, OSSECO researchers also use and develop dedicated tools to get a better insight into how the ecosystem surrounding an OSS project affects its evolution (van Angeren et al., 2011). We can conclude that, because of the openness of the OSSECO repositories, they are ideal for statistical and network analysis research.

In our study, Eclipse was the predominant OSSECO studied. It was analyzed from different perspectives (e.g., OSSECO licensing models [R9], co-creation process in OSSECOs [R17], global SECOs [R28], OSSECO co-evolution [R43] and OSSECOs marketplaces [R68]). We compared our result with two previous mapping studies (i.e., Manikas and Hansen 2013, and Ameller et al. 2015). In their work, Eclipse was the most referenced OSSECO, 16% and 41.7% respectively. Eclipse’s popularity among researchers may be

due to the less restrictive Eclipse Public License [R1], the Eclipse incubation programs [R13], the common development infrastructure, the possibilities of co-creation and co-evolution with relevant partners, among other important aspects.

5.3. RQ3. Which representations have been proposed for OSSECOs?

RQ3.1 Which primary studies use models to represent OSSECOs?

Fig. 21 shows the numbers and percentage of papers that use OSSECO models in the secondary studies of Hanssen and Dybå (2012); Barbosa et al. (2013); Manikas (2016a); Axelsson and Skoglund (2016), and this study. Significant differences in the five studies can be observed. However, this might be due to the fact that we used a more flexible criterion for paper classification (i.e., we selected a paper if it had any OSSECO model). In contrast, Manikas and Barbosa were more restrictive in their criteria (i.e., Manikas selected papers with empirical and analytical models, Barbosa selected papers with software product line development models and OSS development models, all of the papers from Axelsson have qualitative or descriptive models, and Hanssen identified papers describing and modelling ecosystems). Nevertheless, we agree with Manikas when he argues that there is a lack of papers using models based on automatic or mathematical manipulation for solving a specific problem and there is an excess of papers using ad hoc models.

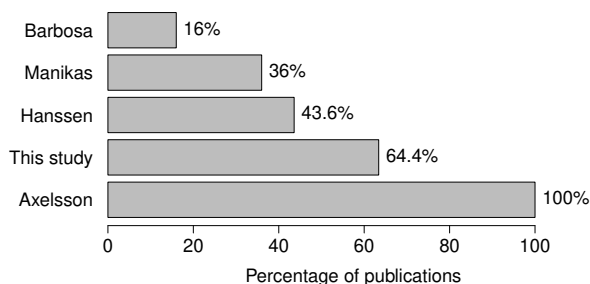


Figure 21: Comparison of publications with SECO models

RQ3.2 Which of the proposed models, if any, are specific for OSSECOs?

OSSECO modelling has emerged as an important research area in software engineering (Handoyo et al., 2013). In our literature review we identified several specific OSSECO models and meta-models to describe and analyze the complex relationships between members in

specific OSSECO case studies. However, there is no unified model language for OSSECOs. We found that researchers of OSSECOs used several types of modelling techniques that are specifically adapted for only one or a few research studies. Nonetheless, there is still a need for modelling OSSECOs due to the following: (a) Complexity- Since SECOs have several type of actors, resources, implicit boundaries, shared market, licenses issues, etc; they are complex artifacts (Mens et al., 2014) and we need to understand them; (b) Traceability- Since the software industry is constantly evolving and is currently undergoing rapid changes (Yu and Stephanie, 2011), it is important to understand OSSECO evolution by analyzing its historical data sources, and (c) Communication- Because of the complex network of symbiotic relationships between entire social actors, open source communities and commercial software companies, etc. (Wynn Jr. et al., 2008), the heterogeneity of OSSECO stakeholders will require a common language to facilitate communication. In other words, OSSECO modelling needs to be complemented by more research efforts that focus on providing model-based approaches to describe and analyze OSSECOs.

RQ3.3 Which notation and guidelines have been used for modelling OSSECOs?

We found several notations for modelling OSSECOs. However, all of them adapt available modelling techniques or use ad hoc models to support their works without proposing new modelling techniques. We think that the development of new modelling techniques for OSSECOs is important because it has evolved from different domains (i.e., ecosystem and BECO). These domains are not directly related to the software engineering base of knowledge nor have the software modelling techniques been designed for the complex elements and symbiotic relationships of software ecosystems. Furthermore, there is a small but growing line of recent research efforts that is specifically focused on providing model-based approaches to describe and analyze SECOs (H. Sadi and Yu, 2015). These conditions are necessary for modelling OSSECOs in a systematic way. In addition, they allow abstracting and reasoning about OSSECOs (Christensen et al., 2014). Table 13 shows the contingency table for the type of ecosystem and models used.

RQ3.4 What type of analysis was conducted using the models identified in RQ3.3?

Jansen et al. (2015) identify three important uses of SECO modelling, one of which is SECO analysis. How-

Table 13: Contingency table for ecosystem and models

	BECO	DBECO	ECO	OSSECO	SECO
Ad hoc	8 (14.5%)	1 (1.8%)	0 (0%)	0 (0%)	9 (16.4%)
Class Diagram	1 (1.8%)	0 (0%)	0 (0%)	1 (1.8%)	2 (3.6%)
Conceptual Map	2 (3.6%)	1 (1.8%)	0 (0%)	1 (1.8%)	1 (1.8%)
iStar	3 (5.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Mathematical	1 (1.8%)	0 (1.8%)	2 (3.6%)	0 (0%)	2 (3.6%)
Metamodel	1 (1.8%)	0 (0%)	0 (0%)	0 (0%)	3 (5.5%)
SNA	1 (1.8%)	0 (0%)	1 (1.8%)	0 (0%)	5 (9.1%)
Tabular	2 (3.6%)	0 (0%)	0 (0%)	4 (7.3%)	3 (5.5%)

ever, we found that most of the papers that use models have not conducted any OSSECO analysis (see Fig.12). In addition, the analysis techniques used in the remaining papers, such as mathematical, visual, statistical, and SNA techniques were used to analyze specific cases. They are insufficient when a more in-depth analysis is necessary. In agreement with other authors (Barbosa et al. 2013; Jansen et al. 2015; Manikas 2016a), we think that developing analysis and modelling techniques is one of the most important challenges in the OSSECO domain. We found a correlation between *Type of analysis* and *Model type*: 32.1% (18) of the papers that conducted some type of analysis use adhoc models.

5.4. Threats to validity

As in every empirical study, there are several threats that might negatively affect the validity of this systematic mapping. In the protocol, we identified and tried to mitigate them using four categories: construct validity, internal validity, external validity, and conclusion validity (see Wohlin et al. (2012)) for details of this classification).

5.4.1. Construct validity

The construct validity category includes three major threats. The first threat is that the research questions may not cover all the relevant aspects that characterize the existing research in our area of interest. To minimize this risk, we used a brainstorming technique with the participation of all the authors of the study to define them. The second threat is that the inclusion of all the relevant works in the field is not guaranteed. This threat was mitigated by combining several databases and manual searches to selected journals and conferences from previous literature reviews on software engineering and OSS. However, this issue may not have been solved since the problem goes beyond an accurate protocol and also concerns issues related to the paper (e.g., inaccurate abstracts). To mitigate this risk, we included the papers from two other literature reviews (Hanssen and Dybå, 2012; Barbosa et al., 2013; Manikas, 2016a; Axelsson and Skoglund, 2016) and all of the chapters of the only existing book that is centered on the study of SECOs

(Jansen et al., 2013). Finally, there is a risk of obtaining a biased selection. To mitigate this risk, inclusion and exclusion criteria guided the selection, and a multi-stage process involving more than one researcher for each paper was used to perform it.

5.4.2. Internal validity

There are two threats to internal validity in this systematic mapping. The first threat is that most of the papers do not provide accurate definitions or references for the OSSECO term. For instance, several papers use definitions related to SECOs and they study OSS communities or OSS projects (see Fig. 7 and Table B.17). The second threat is related to the identification of values for classification criteria: for some of the criteria to classify the papers, the possible values were not obvious. with regard to OSSECO related definitions, one author identified the possible values, and the list of definitions was discussed and analysed closely by all of the authors of the paper. Furthermore, we calculated a word frequency table from the documents and added other 16 new OSSECO related definitions. With regard to OSSECO models, we found a lack of modelling techniques to represent OSSECOs. We decided to identify the different ecosystem definitions of the authors and classify the techniques and notations used in each paper to model ecosystems. This process minimized the risk because several papers use these techniques to model OSS communities or OSS projects.

5.4.3. External validity

Since our results are within the scope of OSSECOs and we do not attempt to generalize conclusions beyond this scope, external validity threats do not apply.

5.4.4. Conclusion validity

Conclusion validity is concerned about whether the research performed is reproducible by other researchers with similar results. In this regard, we have explicitly described all of the steps performed in the systematic mapping by detailing the procedure as defined in the research method (See Section 3). We have also created an online document with details that are not central to the paper but that are necessary to ensure reproducibility and provide evidence about our findings.

6. Further work

The analysis of the results allows us to state that OSSECO is a growing research area in software engineering [R16, R49, R50]. Due to this, there are several

new research opportunities in the empirical examination, modelling, analysis, measuring, quality evaluation, etc. of OSSECOs. Along with this argumentation, in this section we provide two initial proposals to improve the current structure of the knowledge on OSSECOs: a definition for OSSECOs and a taxonomy of OSSECO-related terms.

6.1. The OSSECO definition

In any domain, the concept of ecosystem can be difficult to define clearly. This is true even among scholars in ecology, its native discipline. According to our study, there is a relation between BECOs, SECOs, and OSSECOs (see Fig. 19). In particular, the difference between OSSECOs and just SECOs is made explicit by Manikas' literature review itself Manikas (2016b), where is stated that *“one of the most common differentiation of ecosystem types is the separation between ecosystems that are driven or supported by free and open source software (FOSS) and ecosystems that are driven or supported by proprietary software”*.

In order make explicit the differences between OSSECOs and the other SECO types, we highlight some major findings:

- *Software development process:* As Weber (2004) states: *“The essence of open source, is not the software. It is the process by which software is created”*. In OSSECOs the development process is decentralized and collaborative, *“programs must be broken down into discrete modules so that different people can work on different modules at different times without loss of coherence to the whole”* [R21]. This allows sharing source code between projects and knowledge. In addition, it allows co-evolve the OSSECO community with its associated project [R29]. On the other hand, proprietary SECOs limit access to proprietary information, such as source code Manikas (2016b).
- *Project contributions and collaboration:* According to Manikas (2016b), contribution is not one of the most common components in proprietary SECOs, also *“proprietary SECO studies lack deeper investigation of technical and collaborative aspects”*. On the contrary, in OSSECOs, the OSS community is a key stone actor in the ecosystem [R1], [R40] being the role of its contributors very relevant in it. Actually, the number of active contributors is one of the most important indicators of the OSSECO health and quality [R8], [R11], [R14], [R62].

- *Governance:* Traditionally proprietary SECOs have a closed environment in which a single entity, like a company or a corporation, holds the power to make decisions about what is approved and disapproved to be added into the software code base and what to implement next [R20]. This means that the proprietary entity is the center of the ecosystem [R32]. In OSSECOs the decision power does not belong to a company or corporation. Instead, there are OSSECOs where it is the community of contributors that determines *“which contributions are accepted into the source code base and where the software is headed. Individual developers, the committers, and not a specific company, make decisions about the software”* [R20] while there are also OSSECOs where benevolent dictatorships *“own”* or *“control”* the projects evolution [R20].
- *Co-* concepts:* OSSECOs provides new capabilities for creation, innovation and developing that exceed the benefits offered by proprietary SECOs [R15]. The co-* concepts refer to competence and collaboration between different entities about a specific topic: Co-evolving [R16], co-operation [R13], co-develop [R13] and co-creation [R17]. These topics will be difficult, if not impossible, to implement in proprietary SECOs. Because, as Manikas (2016b) states: *“proprietary SECO studies lack deeper investigation of technical and collaborative aspects”*.

By combining the definitions of SECO, BECO, and DBECO that we found in our mapping study and taking into account the differences between OSSECOs and the other SECO types, we define an OSSECO as: *a SECO placed in a heterogeneous environment, whose boundary is a set of niche players and whose keystone player is an OSS community around a set of projects in an open-common platform*. Table 14 details the OSSECO definition. The first column shows the breakdown of the OSSECO definition. The second column describes the definition-related elements in an OSSECO. The third column references the source of the definition component. Finally, the last column contains specific examples.

6.2. Taxonomy of the OSSECOs terms

In this subsection, we present an initial taxonomy composed of the terms that we found in our review. In order to do this, we applied the second cycle of our data extraction process to the results from Section 4.2. We

Table 14: Breakdown of OSSECO definition

OSSECO definition break down	Description	Source	Examples
<i>a SECO placed in a heterogeneous environment</i>	In OSSECO is an economical social and technical environment	Iansiti and Levin BECO	Other OSSECOs, commercial SECOs, Government, Market rules, synaptical relationships, etc.
<i>whose boundary is a set of niche players</i>	In OSSECO there is more than one.	Jansen et al. SECO	Partners, Re-sellers, Platform provider, etc.
<i>and whose keystone player is an OSS community around a set of projects in a open-common platform</i>	In OSSECO keystone players drive platform technologies and the standards (Aarnoutse et al. 2014).	Lungu et al. SECO	Contributors, passive users, data sources, etc.

then grouped the OSSECO terms into three dimensions, which we had presented in a previous work (Franco-Bedoya et al., 2014): (a) the software platform which groups the terms related to the technology or market around which the ecosystem is built; (b) the OSS community, which groups the terms related to the community (or set of communities) of the ecosystem; (c) the ecosystem network, which groups the terms related to the ecosystem as a network of elements, such as projects or companies. These categories are related to the SECO viewpoints defined in Section 2.2. In addition, we divided the categories into subcategories based on the categories from Bosch (2009), the levels from Jansen et al. (2009b), and the dimensions from dos Santos and Werner (2011). This taxonomy, which is presented in Fig. 22, aims to serve as the starting point for establishing a common terminology for OSSECO²².

Many of the terms in the taxonomy are not exclusive to OSSECOs; however, many of them exhibit some characteristics that are specific to the OSSECO domain. For example:

- A *network boundary* around an open or semi-open *platform boundary* has the potential for numerous benefits, including enhanced *adopter* offers through the use of innovation potential in the OSSECO [R1].
- Unlike to other software distribution paradigms, *source code* is usually available from OSSECO *repositories*. This facilitates some software quality practices like peer reviews (Axelsson and Skoglund, 2016).
- In OSSECOs, the relationships between *keystone players* (e.g., the OSS community) and *niche players* (e.g., partners, providers, adopters) are under

an OSS *license* schema. It is sometimes difficult to control because there are different licenses with specific characteristics that are not always compatible (Dai et al., 2011).

- In OSSECOs, the *OSS community* usually dominates the development instead of an individual organization (this could happen indirectly because sometimes a community is influenced by a single organization indirectly) (Hanssen and Dybå, 2012). The community defines a *roadmap* that guides the development.
- The OSSECOs typically provide access to all data *repositories* related to their evolution (i.e., how software changes over time) (Goeminne and Mens, 2013). Also, another feature of software *repositories* is the option to fork or copy a whole *OSS project* and start a different *forge* of the project (Manikas, 2016a).

Table B.19 (in Appendix B) shows a general description of each taxonomy term according to the primary studies. We are currently developing an OSSECO ontology based on the taxonomy presented here. This ontology is intended to support QuESo, a framework for the representation, synthesis, analysis, evaluation, and evolution of OSSECOs [23]. For this purpose, we are improving the OSSECO taxonomy by standardizing and extending the set of terms, and we are using OWL to describe the relationships among the concepts. Finally, we are defining a set of axioms and inference rules to represent the meaning of these concepts in a formal way to support reasoning.

7. Research roadmap for OSSECOs

In this section, we outline a research roadmap for OSSECOs. First, we compile the few studies in the broader area of SECOs that have identified research

²²In Fig. 22, the references to Jansen et al. (2009b), dos Santos and Werner (2011) and Bosch (2009) are abbreviated with the name of the first author for the sake of brevity.

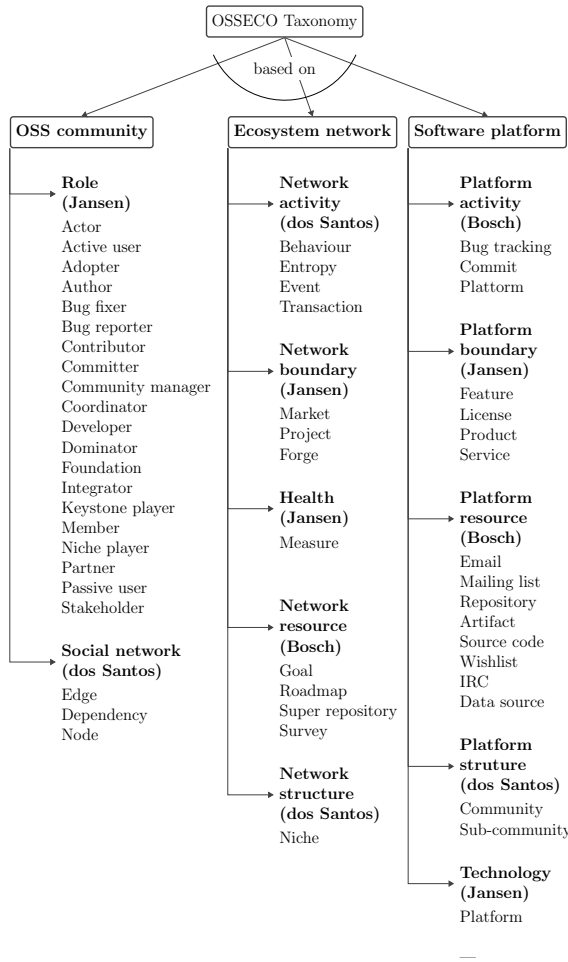


Figure 22: OSSECO taxonomy: an initial proposal

challenges. The first study in that direction was Jansen et al. (2009a), who mentioned several challenges, notably characterization and modelling of SECOs. Barbosa et al. (2013) identified eight major fields within the software ecosystem domain; it is worth mentioning that one of them is the further study of OSSECOs. Hanssen and Dybå (2012) uncovered several theoretical challenges about SECO, which are specifically related to socio-technical theory. Finally, Manikas (2016a) proposed two approaches to address complexity and theory building in SECOs. Table 15 summarizes these SECO challenges and the papers in our set of primary studies that addressed them in the context of OSSECOs.

The primary studies listed in Table 15 only provided partial answers to the fundamental questions behind these four challenges. Furthermore, some other aspects were not mentioned in the four papers on SECO challenges, but they do appear in some of the primary stud-

Table 15: SECO challenges

Authors	Challenge	Addressed by
Jansen et al. (2009a)	Characterisation and modelling of SECOs	R2, R10, R16, R87
	Developing Policies and strategies within SECOs for SECO orchestration	R4, R47, R49, R56, R88
	Determining a strategy to thrive and make profit in an SSN	R13, R18, R20
Barbosa et al. (2013)	How quality can be measured per developer	R19, R51, R61
	How relationships are formed between developers	R39, R59, R63, R85, R86, R87
	How conflicts are solved in OSS ecosystems	R40, R73
	How decisions are made in SECO and how can be measured in code changes	R53, R55, R64
Hanssen and Dybå (2012)	Socio-technical theory	R37, R54, R77, R86, R87
	Related theory of organizational ecology	R34, R38, R76
Manikas (2016a)	Software ecosystem scoping	None (It is from 2016)
	Theory building	None (It is from 2016)

ies that we have surveyed. As a result of both observations, we outline the following research agenda:

- *OSSECO modelling and analysis.* Different authors mention the need to tackle the lack of a universally accepted set of modelling methods because this is hampering the advancement of software ecosystem research (Barbosa et al., 2013; H. Sadi and Yu, 2015). In addition, modelling large networks, scalable model visualization tools, and the study of ecosystem evolution and dynamics are some of the major challenges within the SECO domain (Jansen et al., 2015).
- *Socio-technical theory.* The field of SECOs is missing an established theoretical background (Manikas, 2016a). Socio-technical theory addresses important concepts such as organizational control, ecosystem self-regulation, network organization, the role of technology, and the sharing of values. These concepts are relevant in order to understand OSSECOs as the interplay between the social system and the technical system (Hanssen and Dybå 2012).
- *Ecosystem knowledge.* OSSECO data sources provide access to a variety of information about OSSECO evolution. However, information about social behaviour in the ecosystem must be taking into account. In order to do this, machine learning, text mining, case-based reasoning, and other techniques [R65, R57] can be used to identify social issues such as no implicit relationships, community sentiment analysis, cross-references between OSSECOs, among many other challenges [R23, R85].
- *OSSECO quality.* The quality of OSSECOs affects organizations, adopters, software developers and the OSSECO itself. However, quality management and operationalization of software ecosystems is still an immature discipline. In addition, OSSECO quality is quite different from the standard ones (e.g., ISO/IEC 25010, in terms of production

process, community, distribution methods, license types, social organization, support, etc (Aversano and Tortorella, 2011)). Therefore, OSS quality models emerged due to the inability of traditional quality models to measure these unique OSS features (Adewumi et al., 2013). These quality models in OSS projects can be the basis of OSSECO quality models (Franco-Bedoya et al., 2014).

- *OSSECO monitoring.* The assessment of OSSECO health is usually realized by tools for a specific community or a specific platform. For instance, there are several solutions in the literature for the monitoring and analysis of specific OSS communities by accessing their available data repositories directly [R1, R4, R13, R83]. There is a need to implement frameworks that are able to: (a) monitor a list of OSSECO quality sub-characteristics over time; (b) link the gathered values with adopter needs by operationalizing quality requirements; and (c) engineer a portfolio of web services that support OSSECOs.

8. Conclusions

This paper has reported a systematic mapping in the field of OSSECOs with the goal of identifying and examining the state of the art on this topic. We designed and followed a rigorous protocol, which uncovered up to 82 papers from a gross total of 652, to answer the different research questions that we identified. We may consider the answers to these questions as the main outcome of this paper.

8.1. RQ1. What are the demographic characteristics of the studies about OSSECOs?

RQ1.1 In which type of sources are articles mostly published? Our results have revealed that research on OSSECOs is mostly published in conference proceedings. The approximate ratio of publication in journals with respect to conferences is 1 to 2. This indicates that OSSECOs are considered to be a valuable software engineering research topic.

RQ1.2 How has the number of publications evolved over the years? OSSECOs have been an increasingly addressed research topic since 2006. Publication peaks occurred in 2011 and 2013. There is evidence that OSSECOs have become an established research domain.

RQ1.3 How are papers geographically distributed? The results in this study suggest that the current output of OSSECO papers is strongly supported by European and North American researchers. However, in the

last four years, authors from other continents have been contributing with publications related to the OSSECO topic. This review shows that the United States and The Netherlands are currently the leading countries in terms of undertaking OSSECOs.

RQ1.4 Who are the predominant researchers? We observed that six authors have been the predominant researchers in OSSECOs. These authors and their clusters account for a considerable fraction of all papers covered in this systematic mapping.

RQ1.5 How are publications distributed between academy and industry? It is no surprise that the publications written only by academic authors by far outnumber papers that have at least one industry author.

RQ1.6 What type of papers are published? Although there are more empirical research papers than papers from other categories (i.e., experience reports and non-empirical papers), the difference is not significant.

8.2. RQ2. What is an OSSECO?

RQ2.1 What definitions are related to the OSSECO definition? Regarding the definitions related to OSSECOs, we encountered five major concepts (i.e., ECO, BECO, DBECO, SECO, and OSSECO), and we built a genealogical tree with their evolution.

RQ2.2 Are there specific definitions of OSSECO? Our results show that there are only three definitions of OSSECOs. This paper proposes a definition of OSSECOs, integrating the different definitions related to OSSECOs: *a SECO placed in a heterogeneous environment, whose boundary is a set of niche players and whose keystone player is an OSS community around a set of projects in an open-common platform.*

RQ2.3 What elements belong to an OSSECO? We obtained up to 64 elements belong to OSSECOs in our review. Among them, project, community, and source code are the most used. Furthermore, we sketched a taxonomy with three categories (i.e., OSS community, ecosystem network, and software platform) to classify the OSSECOs terms.

RQ2.4 What instances of OSSECOs have been reported in the literature? We identified 27 instances of OSSECOs that appear in our systematic mapping. Among them, Eclipse and GNOME are the most frequently used.

8.3. RQ3. Which representations have been proposed for OSSECOs?

RQ3.1 Which primary studies use models to represent OSSECOs? Our study showed that most of the papers adapt available modelling techniques or use ad hoc

models to support their works, without proposing new modelling techniques.

RQ3.2 Which of the proposed models, if any, are specific for OSSECOs? None of the primary studies developed a new technique, notation, or guidelines for modelling OSSECOs.

RQ3.3 Which notation and guidelines have been used for modelling OSSECOs? We found a lack of specific modelling techniques for OSSECOs. However, we identified several modelling techniques to describe them in general. The most commonly applied notations were: ad hoc, tabular, and conceptual maps. Other OSSECOs were modelled using class diagrams, metamodels, or mathematical models.

RQ3.4 What type of analysis was conducted using the models identified in RQ3.3? We found that most of the papers using models for OSSECOs do not conduct any OSSECO analysis. In addition, the analysis techniques used in the remaining papers, such as mathematical, visual, statistical, and SNA were used to analyze specific cases.

Acknowledgments

This work was developed between the years 2013 and 2016 and was supported by the RISCOSS project, which funded by the EC 7th Framework Programme FP7/2007-2013 under agreement number 318249. We would also like to thank the contribution of the EOSSAC project, which funded by the Ministry of Economy and Competitiveness of the Spanish government (TIN2013-44641-P). Oscar Franco-Bedoya has a grant from Universidad Nacional de Colombia.

References

- Aarnoutse, F., Renes, C., Snijders, R., Jansen, S., 2014. The Reality of an Associate Model: Comparing Partner Activity in the Eclipse Ecosystem. In: Proceedings of the ECSAW-2014. ACM, pp. 8:1–8:6.
- Abbasi, A., Altmann, J., Jan 2011. On the correlation between research performance and social network analysis measures applied to research collaboration networks. In: Proceedings of the 44th HICSS. pp. 1–10.
- Adewumi, A., Misra, S., Omoregbe, N., 2013. A Review of Models for Evaluating Quality in Open Source Software. In: Proceedings of the IEEECS. Vol. 4. pp. 88–92.
- Ameller, D., Burgués, X., Collell, O., Costal, D., Franch, X., Papazoglou, M. P., 2015. Development of Service-Oriented Architectures Using Model-Driven Development: A Mapping Study. Information and Software Technology 62, 42 – 66.
URL 10.1016/j.infsof.2015.02.006
- Aversano, L., Tortorella, M., 2011. Evaluating the Quality of Free/Open Source Systems: A Case Study. In: Filipe, J., Cordeiro, J. (Eds.), Proceedings of the 12th ICEIS. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119–134.
- Axelsson, J., Skoglund, M., 2016. Quality Assurance in Software Ecosystems: A Systematic Literature Mapping and Research Agenda. Journal of Systems and Software 114, 69 – 81.
URL <http://dx.doi.org/10.1016/j.jss.2015.12.020>
- Barbosa, O., Alves, C., 2011. A Systematic Mapping Study on Software Ecosystems. In: Proceedings of the 3rd IWSECO. pp. 15–26.
- Barbosa, O., dos Santos, R. P., Alves, C., Werner, C., Jansen, S., 2013. Systematic Mapping Study on Software Ecosystems from a Three-Dimensional Perspective., Ch. 4. Vol. 1 of Jansen et al. (2013), pp. 59–81.
- Berger, T., Pfeiffer, R.-H., Tartler, R., Dienst, S., Czarnecki, K., Wsowski, A., She, S., 2014. Variability Mechanisms in Software Ecosystems. Information and Software Technology 56 (11), 1520 – 1535, special issue on Software Ecosystems.
- Bosch, J., 2009. From Software Product Lines To Software Ecosystems. In: Proceedings of the 13th SPLC. pp. 111–119.
- Bosch, J., 2012. Software Ecosystems - Implications for Strategy, Business Model and Architecture. Tutorial in the 34th ICSE.
- Briscoe, G., 2009. Digital Ecosystems. Ph.D. thesis, Imperial College London.
- Buford, J., Mahajan, K., Krishnaswamy, V., Dec 2011. Federated Enterprise and Cloud-based Collaboration Services. In: Proceedings of the 5th IEEE-IMSAA. pp. 1–6.
- Chen, W., Hirschheim, R., 2004. A Paradigmatic and Methodological Examination of Information Systems Research From 1991 to 2001. Information systems journal 14 (3), 197–235.
URL 10.1111/j.1365-2575.2004.00173.x
- Cheong, F., Corbitt, B., 2009. A Social Network Analysis of the Co-authorship Network of the Pacific Asia Conference on Information Systems from 1993 to 2008. pp. 1–23.
- Christensen, H. B., Hansen, K. M., Kyng, M., Manikas, K., 2014. Analysis and Design of Software Ecosystem Architectures : Towards the 4S Telemedicine Ecosystem. Information and Software Technology 56 (11), 1476 – 1492, special issue on Software Ecosystems.
URL 10.1016/j.infsof.2014.05.002
- Dai, N., Thronicke, W., Lopez, A. R., Latasa, F. C., Zeeb, E., Fiehe, C., Litvina, A., Krueger, J., Dohndorf, O., Agudo, I., et al., 2011. OSAMI Commons - An open dynamic services platform for ambient intelligence. In: Proceedings of the 16th ETFA. IEEE, pp. 1–10.
- Davies, M., 2015. Academic Vocabulary Lists. Corpus of Contemporary American English. <http://www.academicvocabulary.info/>, accessed 2 Mayo 2015.
- den Hartigh, E., Visscher, W., Tol, M., Salas, A. J., 2013. Measuring the Health of a Business Ecosystem, Ch. 11. Vol. 1 of Jansen et al. (2013), pp. 221–246.
- Dhungana, D., Groher, I., Schludermann, E., Biffl, S., 2010. Software Ecosystems vs. Natural Ecosystems: Learning from the Ingenious Mind of Nature. In: Proceedings of the 4th ECSA. ACM, pp. 96–102.
- dos Santos, R. P., Werner, C. M. L., 2011. A Proposal for Software Ecosystems Engineering. In: Proceedings of the 3th IWSECO. pp. 40–51.
- Driver, M., 2013. Hype Cycle for Open-Source Software. Tech. Rep. G00252545, Gartner.
- Eclipse Foundation, 2015. Eclipse Community. <http://www.eclipse.org/community>, accessed 30 Nov 2015.
- Feinerer, I., 2015. Introduction to the tm Package Text Mining in R. URL <https://cran.r-project.org/web/packages/tm/index.html>
- Figay, N., Ghodous, P., Nov 2009. FLOSS as Enterprise Application Interoperability Enabler. In: Proceedings of 5th SITIS. pp. 435–442.
- Fitzgerald, B., Agerfalk, P. J., 2008. Outsourcing to an Unknown Workforce: Exploring Opensourcing as a Global Sourcing Strat-

- egy. *MIS Quarterly* 32 (2), 385–409, p1.
- Foulonneau, M., Pawelzik, R., Grégoire, B., Donak, O., 2013. Analyzing the Open Source Communities' Lifecycle with Communication Data. In: *Proceedings of the 5th MEDES*. ACM, pp. 340–344.
- Franco-Bedoya, O., Ameller, D., Costal, D., Franch, X., 2014. QuESo: A Quality Model for Open Source Software Ecosystems. In: *Proceedings of the 9th ICSOFT-EA*. pp. 209–221.
- Gentleman, R., Ihaka, R., 2015. The R project for Statistical Computing. <https://www.r-project.org>, accessed 30 Nov 2015.
- German, D., Adams, B., Hassan, A., March 2013. The Evolution of the R Software Ecosystem. In: *Proceedings of the 17th CSMR*. pp. 243–252.
- Goeminne, M., Feb 2014. Understanding the Evolution of Sociotechnical Aspects in Open Source Ecosystems. In: *Proceedings of IEEE-(CSMR-WCRE)*. pp. 473–476, p60.
- Goeminne, M., Mens, T., 2010. A Framework for Analysing and Visualising Open Source Software Ecosystems. In: *Proceedings of IWPSE-EVOL*. ACM, pp. 42–47.
- Goeminne, M., Mens, T., 2013. Analyzing Ecosystems for Open Source Software Developer Communities, Ch. 12. Vol. 1 of Jansen et al. (2013), pp. 247–275.
- Golbeck, J., 2013. *Analyzing the Social Web*, 1st Edition. Elsevier, Waltham, USA.
- Gousios, G., Spinellis, D., June 2012. GHTorrent: Github's Data from a Firehose. In: *Proceedings of the 9th MSR*. pp. 12–21.
- H. Sadi, M., Yu, E., 2015. Designing Software Ecosystems: How Can Modeling Techniques Help? In: *Proceedings of 16th BPMDS*. pp. 360–375.
- Handoyo, E., Jansen, S., Brinkkemper, S., 2013. Software Ecosystem Modeling: The Value Chains. In: *Proceedings of the 5th MEDES*. pp. 17–24.
- Hanssen, G. K., Dybå, T., 2012. Theoretical Foundations of Software Ecosystems. In: *Proceedings of the 4th IWSECO-ICSOB*. pp. 6–17.
- Hauge, Ø., Ayala, C., Conradi, R., 2010. Adoption of Open Source Software in Software-intensive Organizations—A Systematic Literature Review. *Information and Software Technology* 52 (11), 1133–1154.
URL 10.1016/j.infsof.2010.05.008
- Hoving, R., Slot, G., Jansen, S., 2013. Python: Characteristics Identification of a Free Open Source Software Ecosystem. In: *Proceedings in the 7th IEEE-DEST*. pp. 13–18.
- Iansiti, M., Levien, R., 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*, 1st Edition. Harvard Business Press, Boston, USA.
- Jabeur, L. B., Tamine, L., Boughanem, M., 2010. A Social Model for Literature Access: Towards a Weighted Social Network of Authors. In: *Proceedings of 9th RIAO*. pp. 32–39.
- Jansen, S., 2007. *Customer Configuration Updating in a Software Supply Network*. IKS Dissertation Series, Utrecht University.
- Jansen, S., Brinkkemper, S., Cusumano, M., 2013. *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*, 1st Edition. Edward Elgar Publishing, Cheltenham, UK.
- Jansen, S., Finkelstein, A., Brinkkemper, S., 2009a. A Sense of Community: A Research Agenda for Software Ecosystems. In: *Proceedings of the 31st ICSE*. pp. 187–190.
- Jansen, S., Finkelstein, A., Finkelstein, A., 2009b. Business Network Management as a Survival Strategy: A Tale of Two Software Ecosystems. In: *Proceedings of the 1st IWSECO*. pp. 187–190.
- Jansen, S., Handoyo, E., Alves, C., 2015. Scientists' Needs in Modelling Software Ecosystems. In: *Proceedings of the ECSAW EC-SAW '15*. ACM, pp. 44:1–44:6.
- Jansen, S., Souer, J., Luinenburg, L., Brinkkemper, S., 2012. Shades of gray: Opening up a software producing organization with the open software enterprise model. *Journal of Systems and Software* 85 (7), 1495–1510.
URL 10.1016/j.jss.2011.12.007
- Joshua, J., Alao, D., Okolie, S., Awodele, O., 2013. Software Ecosystem: Features, Benefits and Challenges. *International Journal of Advanced Computer Science & Applications* 4 (8).
- Kabbedijk, J., Jansen, S., 2011. Steering Insight: An Exploration of the Ruby Software Ecosystem. In: *Proceedings of the 2nd ICSOB*. Springer, pp. 44–55.
- Kamei, Y., Matsumoto, S., Maeshima, H., Onishi, Y., Ohira, M., Matsumoto, K.-i., 2008. Analysis of Coordination Between Developers and Users in the Apache Community. In: *Proceedings of the 4th OSS*. pp. 81–92.
- Kapitsaki, G. M., Tselikas, N. D., Foukarakis, I. E., 2015. An Insight Into License Tools for Open Source Software Systems. *Journal of Systems and Software* 102, 72 – 87.
URL 10.1016/j.jss.2014.12.050
- Kitchenham, B., Charters, S., 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Tech. Rep. EBSE-2007-01, Keele University.
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., Linkman, S., 2009. Systematic Literature Reviews in Software Engineering—A Systematic Literature Review. *Information and Software Technology* 51 (1), 7–15.
URL 10.1016/j.infsof.2008.09.009
- Li, W., Wang, X., Yu, M., Sept 2010. A Research on Collaboration Knowledge Construction in the Virtual Learning Community by Social Network Analysis. In: *Proceedings of ICEIT*. Vol. 2. pp. V2–323–V2–327.
- Li, Y., Tan, C.-H., Yang, X., 2013. It is All About What We Have: A Discriminant Analysis of Organizations' Decision to Adopt Open Source Software. *Decision Support Systems* 56, 56 – 62.
URL 10.1016/j.dss.2013.05.006
- López, L., Costal, D., Ayala, C. P., Franch, X., Annosi, M. C., Glott, R., Haaland, K., 2015. Adoption of {OSS} components: A Goal-oriented Approach. *Data & Knowledge Engineering* 99, 17 – 38.
URL <http://dx.doi.org/10.1016/j.datak.2015.06.007>
- Lungu, M., 2008. Towards Reverse Engineering Software Ecosystems. In: *Proceedings of the IEEE-ICSM*. pp. 428–431.
- Manikas, K., 2016a. Revisiting Software Ecosystems Research: A Longitudinal Literature Study. *Journal of Systems and Software* 117, 84 – 103.
URL <http://dx.doi.org/10.1016/j.jss.2016.02.003>
- Manikas, K., Jun 2016b. Supporting the Evolution of Research in Software Ecosystems: Reviewing the Empirical Literature. In: *Proceedings of the 7th ICSOB 2016*. pp. 63–78.
- Manikas, K., Hansen, K. M., 2013a. Reviewing the Health of Software Ecosystems—A Conceptual Framework Proposal. In: *Proceedings of the 5th IWSECO-ICSOB*. pp. 33–44.
- Manikas, K., Hansen, K. M., 2013b. Software Ecosystems: A Systematic Literature Review. *Journal of Systems and Software* 86 (5), 1294 – 1306.
URL 10.1016/j.jss.2012.12.026
- Mens, T., Claes, M., Grosjean, P., Serebrenik, A., 2014. *Studying Evolving Software Ecosystems Based on Ecological Models*, 1st Edition. Vol. 1. Springer Berlin Heidelberg, Berlin, Germany, Ch. 10, pp. 297–326.
- Mens, T., Goeminne, M., 2011. Analysing the Evolution of Social Aspects of Open Source Software Ecosystems. In: *Proceedings of the 3rd IWSECO*. pp. 1–14.
- Messerschmitt, D. G., Szyperki, C., June 2003. *Software Ecosystem: Understanding an Indispensable Technology and Industry*, 1st Edition. Vol. 1 of MIT Press Books. The MIT Press, Cambridge MA, USA.
- Miles, M. B., Huberman, A. M., Johnny, S., 2014. *Qualitative Data*

- Analysis: A Methods Sourcebook, 3rd Edition. SAGE Publications, California, USA.
- Mizushima, K., Ikawa, Y., 2011. A structure of Co-creation in an Open Source Software Ecosystem: A Case Study of the Eclipse Community. In: Proceedings of PICMET. pp. 1–8.
- Montesi, M., Lago, P., 2008. Software Engineering Article Types: An Analysis of the Literature. *Journal of Systems and Software* 81 (10), 1694 – 1714.
URL [10.1016/j.jss.2007.11.723](https://doi.org/10.1016/j.jss.2007.11.723)
- Moore, J. F., 1993. Predators and Prey: A New Ecology of Competition. *Harvard business review* 71 (3), 75–83.
URL blogs.harvard.edu/jim/files/2010/04/Predators-and-Prey.pdf
- Morgan, L., Feller, J., Finnegan, P., 2013. Exploring Value Networks. *European Journal of Information Systems* 22 (5), 569–588.
URL [10.1057/ejis.2012.44](https://doi.org/10.1057/ejis.2012.44)
- Nachira, F., Chiozza, E., Ithonen, H., Cunningham, F., 2002. Towards a Network of Digital Business Ecosystem. Tech. rep., European Commission DG INFSO.
- Nachira, F., Dini, P., Andrea, N., 2007. A Network of Digital Business Ecosystems for Europe: Roots, Processes and Perspectives, 1st Edition. European Commission, in association and with the support of the FP6 projects DBE and OPAALS, Luxembourg, Luxembourg, pp. 1–20.
- Narang, T., 2015. Hierarchical Clustering of Documents: A Brief Study and Implementation in Matlab. In: Proceedings of the 3rd ICETESMA. pp. 153–159.
- Newman, M. E. J., Jun 2001. Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality. *Phys. Rev. E* 64, 016132.
URL [10.1103/PhysRevE.64.016132](https://doi.org/10.1103/PhysRevE.64.016132)
- Oriol, M., Franco-Bedoya, O., Franch, X., Marco, J., May 2014. Assessing open source communities' health using service oriented computing concepts. In: Proceedings of the 8th IEEE RCIS. pp. 1–6.
- Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank Citation Ranking: Bringing Order to the Web. Tech. Rep. 1999-66, Stanford InfoLab.
- Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M., 2008. Systematic Mapping Studies in Software Engineering. In: Proceedings of the 12th EASE. pp. 68–77.
- Petersen, K., Vakkalanka, S., Kuzniarz, L., 2015. Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update. *Information and Software Technology* 64, 1–18.
URL [dx.doi.org/10.1016/j.infsof.2015.03.007](https://doi.org/10.1016/j.infsof.2015.03.007)
- Rayson, P., Berridge, D., Francis, B., 2004. Extending the Cochran Rule for the Comparison of Word Frequencies Between Corpora. In: Proceedings of 7th. pp. 926–936.
- Riehle, D., 2007. The Economic Motivation of Open Source Software: Stakeholder Perspectives. *IEEE Computer* 40 (4), 25–32.
URL [10.1109/MC.2007.147](https://doi.org/10.1109/MC.2007.147)
- Ruíz, F., 2012. La Investigación en Informática. en España Análisis Bibliométrico. *Novática* 85 (215), 54–58.
- Salton, G., Buckley, C., 1988. Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management* 24 (5), 513 – 523.
URL [dx.doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Scacchi, W., 2007. Free/Open Source Software Development : Recent Research Results and Emerging Opportunities. In: Proceedings of the 6th SIGSOFT. ACM, pp. 459–468, p24-P36.
- Scacchi, W., Alspaugh, T. A., 2012. Understanding the role of licenses and evolution in open architecture software ecosystems. *Journal of Systems and Software* 85 (7), 1479–1494.
URL [10.1016/j.jss.2012.03.033](https://doi.org/10.1016/j.jss.2012.03.033)
- SCIMago, September 15, 2015. Journal & country rank scimago. (2007). Retrieved.
URL <http://www.scimagojr.com>
- Scott, M., 1997. PC Analysis of Key Words and Key Key Words. *System* 25 (2), 233–245.
URL [10.1016/S0346-251X\(97\)00011-0](https://doi.org/10.1016/S0346-251X(97)00011-0)
- Scott, M., Tribble, C., 2006. Textual Patterns: Key Words and Corpus Analysis in Language Education, 1st Edition. Vol. 22. John Benjamins Publishing, Amsterdam, The Netherlands.
- Smith, M., 2014. NodeXL Network Overview, Discovery and Exploration for Excel. <http://nodexl.codeplex.com/>, accessed November 30 2015.
- Spauwen, R., Jansen, S., 2013. Towards the Roles and Motives of Open Source Software Developers. In: Proceedings of 5th IWSECO. p. 62.
- Squire, M., Williams, D., 2012. Describing the Software Forge Ecosystem. In: Proceedings of the 45th HICSS. IEEE, pp. 3416–3425.
- Stanley, J., Briscoe, G., 2010. The ABC of Digital Business Ecosystems. *Computer, Media and Telecommunications Law* 15 (1), 1–24.
URL arxiv.org/abs/1005.1899
- Syed, S., Jansen, S., 2013. On Clusters in Open Source Ecosystems. In: Proceedings of the 5th IWSECO. pp. 19–32.
- Taljard, E., De Schryver, G.-M., 2002. Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho. *Lexikos* 12 (1).
URL [10.5788/12-0-760](https://doi.org/10.5788/12-0-760)
- Tansley, A. G., 1935. The Use and Abuse of Vegetational Concepts and Terms. *Ecology* 16 (3), 284–307.
URL doi.org/10.2307/1930070
- Toivanen, T., Mazhelis, O., Luoma, E., 2015. Network Analysis of Platform Ecosystems: The Case of Internet of Things Ecosystem.
- Uden, L., Damiani, E., Gianini, G., Ceravolo, P., 2007. Activity Theory for OSS Ecosystems. In: Proceedings of the DEST. Inaugural IEEE-IES. IEEE, pp. 223–228.
- van Angeren, J., Kabbedijk, J., Jansen, S., Popp, K., 2011. A Survey of Associate Models Used Within Large Software Ecosystems. In: Proceedings of the 3rd CEUR-WS. pp. 27–39.
- Van Lingen, S., Palomba, A., Lucassen, G., 2013. On the Software Ecosystem Health of Open Source Content Management Systems. In: Proceedings of the 5th IWSECO. p. 38.
- Viljainen, M., Kauppinen, M., 2013. Framing Management Practices for Keystones in Platform Ecosystem, Ch. 7. Vol. 1 of Jansen et al. (2013), pp. 121–137.
- Weber, S., 2004. The Success Of Open Source. Harvard University Press.
- Weiss, M., 2011. Economics of Collectives. In: Proceedings of the 15th SPLC. ACM, pp. 39:1–39:8.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., Wesslén, A., 2012. Experimentation in Software Engineering, 1st Edition. Springer Science & Business Media, New York, USA.
- Wynn Jr., D., Boudreau, M.-C., Watson, R. T., 2008. Resilience of Professional Open Source Ecosystems. In: Proceedings of ECIS. AISeL, p. Paper 102.
- Xie, T., Zimmermann, T., van Deursen, A., 2013. Introduction to the Special Issue on Mining Software Repositories. *Empirical Software Engineering* 18 (6), 1043–1046.
- Yu, E., Stephanie, D., 2011. Understanding Software Ecosystems: A Strategic Modeling Approach. In: Proceedings of the 3rd IWSECO. pp. 65–76.
- Yu, L., Sept 2013. The Market-Driven Software Ecosystem. *IT Professional* 15 (5), 46–50.
URL [10.1109/MITP.2012.102](https://doi.org/10.1109/MITP.2012.102)
- Yu, L., Ramaswamy, S., Bush, J., Oct 2007. Software Evolvability: An Ecosystem Point of View. In: Proceedings of the 3rd Workshop

on Software Evolvability, pp. 75–80.

Appendix A. Systematic Mapping References

References

- R1 Fitzgerald, B., Agerfalk, P. J., 2008. Outsourcing to an Unknown Workforce: Exploring Opensourcing as a Global Sourcing Strategy. *MIS Quarterly* 32 (2), 385–409, p1.
- R2 Alves, A. M., Pessoa, M., Salviano, C. F., 2011. Towards a Systemic Maturity Model for Public Software Ecosystems. In: *Proceedings of the 116th SPICE*. Springer, pp. 145–156.
- R3 Cataldo, M., Herbsleb, J. J. D., 2010. Architecting in software ecosystems: interface translucence as an enabler for scalable collaboration. In: *Proceedings of the 4th ECSA*. ACM, pp. 65–72.
- R4 Dai, N., Thronicke, W., Lopez, A. R., Latasa, F. C., Zeeb, E., Fiehe, C., Litvina, A., Krueger, J., Dohndorf, O., Agudo, I., et al., 2011. OSAMI Commons - An open dynamic services platform for ambient intelligence. In: *Proceedings of the 16th ETFA*. IEEE, pp. 1–10.
- R5 Darking, M., Dini, P., Whitley, E., 2006. The challenge of building public technology infrastructure: issues of governance and sustainability in a digital business ecosystem. In: *Proceedings of the 14th ECIS*. AISeL, p. Paper 47.
- R6 Economides, N., Katsamakas, E., 2006. *The Economics of Open Source Software Development*. Elsevier B.V., Ch. Linux vs. Windows: A comparison of application and platform innovation incentives for open source and proprietary software platforms, pp. 207–218.
- R7 Figay, N., Ghodous, P., Nov 2009. Floss as enterprise application interoperability enabler. In: *Proceedings of 5th SITIS*. pp. 435–442.
- R8 Gamalielsson, J., Lundell, B., Lings, B., 2010. The Nagios community: An extended quantitative analysis. In: *Proceedings of the 6th OSS*. Springer, pp. 85–96.
- R9 Gamalielsson, J., Lundell, B., Mattsson, A., 2011. Open Source Software for Model Driven Development: A Case Study. In: *Proceedings of the 7th OSS*. Springer, pp. 348–367.
- R10 Goeminne, M., Mens, T., 2010. A framework for analysing and visualising open source software ecosystems. In: *Proceedings of IWPSE-EVOL*. ACM, pp. 42–47.
- R11 Jergensen, C., Sarma, A., 2011. The onion patch: migration in open source ecosystems. In: *Proceedings of the 19th SIGSOFT and 13th ESEC/FSE*. ACM, pp. 70–80.
- R12 Kabbedijk, J., Jansen, S., 2011. Steering insight: An exploration of the ruby software ecosystem. In: *Proceedings of the 2nd ICSOB*. Springer, pp. 44–55.
- R13 Kilamo, T., Hammouda, I., Mikkonen, T., Aaltonen, T., 2012. From proprietary to open source-Growing an open source ecosystem. *Journal of Systems and Software* 85 (7), 1467–1478.
- R14 Lungu, M., Lanza, M., Gırba, T., Robbes, R., 2010. The small project observatory: Visualizing software ecosystems. *Science of Computer Programming* 75 (4), 264–275.
- R15 Mattmann, C. A., Downs, R. R., Ramirez, P. M., Goodale, C., Hart, A. F., 2012. Developing an open source strategy for NASA earth science data systems. In: *Proceedings of the 13th IRI*. IEEE, pp. 687–693.
- R16 Mens, T., Doctors, L., Habra, N., Vanderose, B., Kamseu, F., 2011. QUALGEN: Modeling and Analysing the Quality of Evolving Software Systems. In: *Proceedings of the 15th CSMR*. IEEE, pp. 351–354.
- R17 Mizushima, K., Ikawa, Y., 2011. A structure of co-creation in an open source software ecosystem: A case study of the eclipse community. In: *Proceedings of PICMET*. pp. 1–8.
- R18 Olsson, U., Borjesson, M., 2009. Leveraging open source in commercial service layer development-a case study. In: *Proceedings of the 13th ICIN*. IEEE, pp. 1–6.
- R19 Pérez, J., Deshayes, R., Goeminne, M., Mens, T., 2012. *SECONDA: Software Ecosystem Analysis Dashboard*. In: *Proceedings of the 16th CSMR*. IEEE, pp. 527–530.
- R20 Riehle, D., 2007. The economic motivation of open source software: Stakeholder perspectives. *IEEE Computer* 40 (4), 25–32.
- R21 Samuelson, P., 2006. IBM’s pragmatic embrace of open source. *Communications of the ACM* 49, 21–25.
- R22 Sanz-Salinas, J.-G., Montesinos-Lajara, M., 2009. Current Panorama of the FOSS4G Ecosystem. *UPGRADE* 10 (2), 43.
- R23 Scacchi, W., 2007. Free/Open Source Software Development : Recent Research Results and Emerging Opportunities. In: *Proceedings of the 6th SIGSOFT*. ACM, pp. 459–468, p24-P36.
- R24 Scacchi, W., Alspaugh, T. A., 2012. Understanding the role of licenses and evolution in open architecture software ecosystems. *Journal of Systems and Software* 85 (7), 1479–1494.
- R25 Shao, J., Kuk, G., Anand, S., Morley, J. G., Jackson, M. J., Mitchell, T., 2012. Mapping Collaboration in Open Source Geospatial Ecosystem. *Transactions in GIS* 16 (4), 581–597.
- R26 Squire, M., Williams, D., 2012. Describing the software forge ecosystem. In: *Proceedings of the 45th HICSS*. IEEE, pp. 3416–3425.
- R27 Stanley, J., Briscoe, G., 2010. The ABC of digital business ecosystems. *Computer, Media and Telecommunications Law* 15 (1), 1–24.
- R28 Stevens, G., Draxler, S., 2010. Appropriation of the eclipse ecosystem: Local integration of global network production. In: *Proc. of COOP*. Springer, pp. 287–308.
- R29 Uden, L., Damiani, E., Gianini, G., Ceravolo, P., 2007. Activity theory for OSS ecosystems. In: *Proceedings of the DEST*. Inaugural IEEE-IES. IEEE, pp. 223–228.
- R30 Ververs, E., van Bommel, R., Jansen, S., 2011. Influences on developer participation in the Debian software ecosystem. In: *Proceedings of the MEDES*. ACM, pp. 89–93.
- R31 Weiss, M., 2011. Economics of collectives. In: *Proceedings of the 15th SPLC*. ACM, pp. 39:1–39:8.
- R32 Wynn Jr., D., Boudreau, M.-C., Watson, R. T., 2008. Resilience of professional open source ecosystems. In: *Proceedings of ECIS*. AISeL, p. Paper 102.
- R33 Yamakami, T., 2011. A three-layer view model of OSS: Toward understanding of diversity of OSS. In: *Proceedings of the 13th ICACT*. IEEE, pp. 1190–1194.
- R34 Yu, L., Cawley, J., Ramaswamy, S., 2012. Entropy-Based Study of Components in Open-Source Software Ecosystems. *INFOCOMP Journal of Computer Science* 11 (1), 22–31.
- R35 Donald Wynn, J., 2007. *Emerging Free and Open Source Software Practices*. Idea Group Publishing, Ch. Assessing the Health of an Open Source Ecosystem, pp. 238–258.
- R36 van Angeren, J., Kabbedijk, J., Jansen, S., Popp, K., 2011. A survey of associate models used within large software ecosystems. In: *Proceedings of the 3rd CEUR-WS*. pp. 27–39.
- R37 Goeminne, M., Mens, T., 2013. *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*, Ch. Analyzing ecosystems for open source software developer communities. In: Jansen et al. (2013), pp. 247–275.
- R38 Yu, L., Ramaswamy, S., Bush, J., 2008. Symbiosis and Software Evolvability. *IT Professional* 10 (4), 56–62.
- R39 Lungu, M., Malnati, J., Lanza, M., 2009. Visualizing gnome with the small project observatory. In: *Proceedings of the 6th MSR*. IEEE, pp. 103–106.
- R40 Draxler, S., Jung, A., Boden, A., Stevens, G., 2011. Workplace warriors: identifying team practices of appropriation in software ecosystems. In: *Proceedings of the 4th CHASE*. ACM, pp. 57–60.
- R41 Jansen, S., Souer, J., Luinenburg, L., Brinkkemper, S., 2012. Shades of gray: Opening up a software producing organization with the open software enterprise model. *Journal of Systems and Software* 85 (7), 1495–1510.

- R42 Neu, S., Lanza, M., Hattori, L., D'Ambros, M., 2011. Telling stories about GNOME with Complicity. In: Proceedings of the 6th VIS-SOFT. IEEE, pp. 1–8.
- R43 Draxler, S., Stevens, G., 2011. Supporting the Collaborative Appropriation of an Open Software Ecosystem. *Computer Supported Cooperative Work (CSCW)* 20 (4-5), 403–448.
- R44 Dhungana, D., Groher, I., Schludermann, E., Biff, S., 2010. Software ecosystems vs. natural ecosystems: learning from the ingenious mind of nature. In: Proceedings of the 4th ECSA. ACM, pp. 96–102.
- R45 Gurbani, V. K., Garvert, A., Herbsleb, J. D., 2010. Managing a corporate open source software asset. *Communications of the ACM* 53 (2), 155–159.
- R46 Alspaugh, T. A., Asuncion, H. U., Scacchi, W., 2009. The role of software licenses in open architecture ecosystems. In: Proceedings of the 1st IWSECO. pp. 4–18.
- R47 Morgan, L., Feller, J., Finnegan, P., 2013. Exploring value networks. *European Journal of Information Systems* 22 (5), 569–588.
- R48 Jansen, S., Cusumano, M. A., 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, Ch. Defining software ecosystems: a survey of software platforms and business network governance. In: Jansen et al. (2013), pp. 13–28.
- R49 Viljainen, M., Kauppinen, M., 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, Ch. Framing management practices for keystones in platform ecosystem. In: Jansen et al. (2013), pp. 121–137.
- R50 Jansen, S., Brinkkemper, S., Souer, J., Luinenburg, L., 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, Ch. The open software enterprise model: how open is my software business? In: Jansen et al. (2013), pp. 159–186.
- R51 Jansen, S., van Capelleveen, G., 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, Ch. Quality review and approval methods for extensions in software ecosystems. In: Jansen et al. (2013), pp. 187–217.
- R52 den Hartigh, E., Visscher, W., Tol, M., Salas, A. J., 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, Ch. Measuring the health of a business ecosystem. In: Jansen et al. (2013), pp. 221–246.
- R53 Mens, T., Goeminne, M., 2011. Analysing the Evolution of Social Aspects of Open Source Software Ecosystems. In: Proceedings of the 3rd IWSECO. pp. 1–14.
- R54 Goeminne, M., Feb 2014. Understanding the evolution of socio-technical aspects in open source ecosystems. In: Proceedings of IEEE-(CSMR-WCRE). pp. 473–476, p60.
- R55 Goeminne, M., Claes, M., Mens, T., 2013. A historical dataset for the gnome ecosystem. In: Proceedings of the 10th MSR. IEEE Press, Piscataway, NJ, USA, pp. 225–228.
- R56 Manikas, K., Hansen, K. M., 2013. Reviewing the health of software ecosystems—a conceptual framework proposal. In: proceedings of the 5th IWSECO-ICSOB. pp. 33–44.
- R57 Sowe, S., Zettsu, K., Murakami, Y., April 2013. A model for creating and sustaining information services platform communities: Lessons learnt from open source software. In: Proceedings of ITU-K. pp. 1–8.
- R58 Foulonneau, M., Pawelzik, R., Grégoire, B., Donak, O., 2013. Analyzing the open source communities' lifecycle with communication data. In: Proceedings of the 5th MEDES. ACM, pp. 340–344.
- R59 Teixeira, J., Lin, T., 2014. Collaboration in the open-source arena: The webkit case. In: Proceedings of the 52nd ACM Conference on Computers and People Research. ACM, New York, NY, USA, pp. 121–129.
- R60 Lucassen, G., van Rooij, K., Jansen, S., 2013. Ecosystem health of cloud paas providers. In: Herzwurm, G., Margaria, T. (Eds.), *Software Business. From Physical Products to Software Services and Solutions*. Vol. 150 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg, pp. 183–194.
- R61 Bettenburg, N., Hassan, A., Adams, B., German, D., 2013. Management of community contributions. *Empirical Software Engineering*, 1–38P67.
- R62 Jansen, S., 2014. Measuring the health of open source software ecosystems: Beyond the scope of project health. *Information and Software Technology* 56 (11), 1508 – 1519, special issue on Software Ecosystems P68.
- R63 Fagerholm, F., Johnson, P., Sanchez Guinea, A., Borenstein, J., Munch, J., Aug 2013. Onboarding in open source software projects: A preliminary analysis. In: *Global Software Engineering Workshops (ICGSEW)*, 2013 IEEE 8th International Conference on. pp. 5–10.
- R64 Vasilescu, B., Serebrenik, A., Goeminne, M., Mens, T., 2014. On the variation and specialisation of workload a case study of the gnome ecosystem community. *Empirical Software Engineering* 19 (4), 955–1008.
- R65 Hoving, R., Slot, G., Jansen, S., July 2013. Python: Characteristics identification of a free open source software ecosystem. In: Proceedings of the 7th IEEE-(DEST). pp. 13–18, p71.
- R67 German, D., Adams, B., Hassan, A., March 2013. The evolution of the r software ecosystem. In: *Software Maintenance and Reengineering (CSMR)*, 2013 17th European Conference on. pp. 243–252.
- R68 Yu, L., Sept 2013. The market-driven software ecosystem. *IT Professional* 15 (5), 46–50.
- R69 Aarmoutse, F., Renes, C., Snijders, R., Jansen, S., 2014. The Reality of an Associate Model: Comparing Partner Activity in the Eclipse Ecosystem. In: Proceedings of the ECSAW-2014. ACM, pp. 8:1–8:6.
- R70 Berger, T., Pfeiffer, R.-H., Tartler, R., Dienst, S., Czarnecki, K., Wowski, A., She, S., 2014. Variability Mechanisms in Software Ecosystems. *Information and Software Technology* 56 (11), 1520 – 1535, special issue on Software Ecosystems.
- R71 Bosch, J., Bosch-Sijtsema, P., 2014. ESAO: A Holistic Ecosystem-Driven Analysis Model. In: Proceedings of the 5th ICSOB. Cham, pp. 179–193.
- R72 Costal, D., López Cuesta, L., Franch Gutiérrez, J., 2015. Using Roles for OSS Adoption Strategy Models. In: Proceedings of the 8th i* Workshop-iStar. pp. 19–24.
- R73 Eckhardt, E., Kaats, E., Jansen, S., Alves, C., 2014. The Merits of a Meritocracy in Open Source Software Ecosystems. In: Proceedings of the ECSAW. ACM, pp. 7:1–7:6.
- R74 Eklund, U., Bosch, J., 2014. Architecture for Embedded Open Software Ecosystems. *Journal of Systems and Software* 92, 128 – 142.
- R75 Fotrousi, F., Fricker, S. A., Fiedler, M., Le-Gall, F., 2014. KPIs for Software Ecosystems: A Systematic Mapping Study. In: Proceedings of the 5th ICSOB 2014. pp. 194–211.
- R76 Mens, T., Claes, M., Grosjean, P., Serebrenik, A., 2014. Studying Evolving Software Ecosystems based on Ecological Models. pp. 297–326.
- R77 Teixeira, J., Robles, G., González-Barahona, J. M., 2015. Lessons Learned from Applying Social Network Analysis on an Industrial Free/Libre/Open Source Software Ecosystem. *Journal of Internet Services and Applications* 6 (1), 1–27.
- R78 Handoyo, E., Jansen, S., Brinkkemper, S., 2013. Software Ecosystem Roles Classification. In: Herzwurm, G., Margaria, T. (Eds.), Proceedings of the 4th ICSOB. Berlin, Heidelberg, pp. 212–216.
- R79 Bosch, J., 2009. From Software Product Lines to Software Ecosystems. In: Proceedings of the 13th SPLC. pp. 111–119.
- R80 Iansiti, M., Levien, R., 2004. Strategy as Ecology. *Harvard business review* 82 (3), 68–81.
- R81 Messerschmitt, D. G., Szyperski, C., June 2003. *Software Ecosystem: Understanding an Indispensable Technology and Industry*. Vol. 1 of MIT Press Books. The MIT Press.

- R82 Moore, J. F., 1993. Predators and Prey: A New Ecology of Competition. Harvard business review 71 (3), 75–83.
- R83 Yu, E., Stephanie, D., 2011. Understanding Software Ecosystems: A Strategic Modeling Approach. In: Proceedings of the 3rd IWSECO. pp. 65–76.
- R84 Kenett, R. S., Franch, X., Susi, A., Galanis, N., 2014. Adoption of Free Libre Open Source Software (FLOSS): A Risk Management Perspective. In: Proceedings of the 38th COMPSAC. pp. 171–180.
- R85 Tourani, P., Jiang, Y., Adams, B., 2014. Monitoring Sentiment in Open Source Mailing Lists: Exploratory Study on the Apache Ecosystem. In: Proceedings of the 24th CASCON.
- R86 Orucevic-Alagic, A., Hst, M., 2014. Network Analysis of a Large Scale Open Source Project. In: 2014 40th EUROMICRO. pp. 25–29.
- R87 Syeed, M. M. M., Hansen, K. M., Hammouda, I., Manikas, K., 2014. Socio-technical congruence in the ruby ecosystem. In: Proceedings of the OpenSym. OpenSym '14. pp. 2:1–2:9.
- R88 Cardoso, Jr., J. L., Costa, I., Andres, F., Barbin, S. E., 2015. Collective Intelligence Approach for Free Software Adoption by Municipalities. In: Proceedings of the 7th MEDES. pp. 193–200.
- R89 Jansen, S., Brinkkemper, S., Michael Cusumano, 2013. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry. Edward Elgar Publishing.

Appendix B. Tables

Table B.16: Overview of selected studies

ID.	Type	Year	Country	Cat.	Scope
R1	J	2008	Sweden	R	BECO
R2	C	2011	Brazil	E	DBECO
R3	C	2010	USA	N	
R4	C	2011	Turkey	E	BECO
R5	C	2006	UK	R	
R6	B	2006	USA	R	SECO
R7	C	2009	France	E	
R8	C	2010	Sweden	E	SECO
R9	C	2011	Sweden	R	
R10	C	2010	Belgium	N	SECO
R11	C	2011	USA	R	
R12	C	2011	Netherlands	R	SECO
R13	J	2012	Finland	R	BECO&SECO
R14	J	2010	Switzerland	N	SECO
R15	C	2012	USA	E	BECO
R16	C	2011	Belgium	N	BECO
R17	C	2011	Japan	R	BECO
R18	C	2009	Sweden	N	
R19	C	2012	Belgium	E	SECO
R20	J	2007	Germany	N	
R21	J	2006	USA	N	
R22	J	2009	Spain	N	SECO
R23	C	2007	USA	N	
R24	J	2012	USA	E	SECO
R25	J	2012	UK	E	SECO
R26	C	2012	USA	N	
R27	J	2010	UK	E	DBECO
R28	C	2010	Germany	R	SECO
R29	C	2007	UK	N	
R30	C	2011	Netherlands	R	
R31	C	2011	Canada	E	SECO
R32	C	2008	USA	R	BECO&SECO
R33	C	2011	Japan	N	BECO
R34	J	2012	USA	E	BECO&SECO
R35	B	2007	USA	N	SECO
R36	C	2011	Netherlands	R	BECO
R37	B	2013	Belgium	R	SECO
R38	J	2008	USA	N	BECO
R39	C	2009	Switzerland	E	SECO
R40	C	2011	Germany	N	
R41	J	2012	Netherlands	R	BECO
R42	C	2011	Switzerland	E	
R43	J	2011	Germany	R	
R44	C	2010	Ireland	E	SECO
R45	J	2010	USA	N	
R46	W	2009	USA	N	BECO
R47	J	2013	Ireland	N	BECO
R48	B	2013	Netherlands	R	BECO
R49	B	2013	Finland	N	BECO
R50	B	2013	Netherlands	R	
R51	B	2013	Netherlands	R	
R52	B	2013	Netherlands	N	BECO
R53	B	2013	Belgium	R	SECO
R54	C	2014	Belgium	E	
R55	C	2013	Belgium	E	
R56	W	2013	Denmark	N	SECO
R57	C	2013	Japan	E	SECO
R58	C	2013	Luxembourg	E	
R59	C	2014	Finland	R	BECO
R60	C	2013	Netherlands	E	
R61	J	2013	Canada	R	SECO
R62	J	2014	Netherlands	R	BECO
R63	W	2013	Finland	R	
R64	J	2014	Netherlands	R	
R65	C	2013	Netherlands	R	SECO
R67	C	2013	Canada	R	
R68	J	2013	USA	N	BECO
R69	W	2014	Netherlands	E	BECO
R70	J	2014	Canada	R	SECO
R71	C	2014	Sweden	E	BECO
R72	W	2015	Spain	E	BECO
R73	W	2014	Netherlands	R	
R74	J	2014	Sweden	R	SECO
R75	C	2014	Sweden	R	BECO
R76	B	2014	Belgium	R	ECO
R77	J	2015	Finland	R	BECO
R78	C	2013	Netherlands	R	
R79	C	2009	Sweden	N	BECO
R80	J	2004	USA	N	
R81	B	2003	UK	N	
R82	J	1993	USA	N	
R83	W	2011	Canada	E	BECO
R84	C	2014	Spain	E	BECO
R85	C	2014	Canada	E	SECO
R86	C	2014	Sweden	E	SECO
R87	C	2014	Finland	E	SECO
R88	C	2015	Brazil	E	SECO

Table B.17: SECOs instances and papers related

OSSECO	Papers
Eclipse	R9, R17, R28, R31, R34, R37, R40, R41, R43, R44, R48, R67, R68, R69, R73, R36
GNOME	R10, R11, R19, R39, R42, R53, R60, R62, R64, R76
Android	R48, R70, R86
Ubuntu	R16, R48, R50
Ruby	R12, R48, R87
Open Design Alliance ODA	R36, R41
Debian	R16, R30
Python	R62, R65
Wordpress	R48, R51
Brazilian Public Software (BPS)	R2, R88
Apache	R85
Ecos	R70
Evergreen	R35
FOSS4G	R22
Gurux Software	R13
Moodbile	R84
Nagios	R8
NASA Earth science	R15
OSAMI Consortium	R4
OSGeo	R25
OSMOSOFT	R47
OpenStack	R77
R	R67
Topcased	R9
Vaadin	R13
Webkit	R58

Table B.18: Model techniques and goals

ID	Model goal	Technique
R1	To use open source development model as a global sourcing strategy.	Tabular
R2	To characterize brazilian public software ecosystem.	Conceptual map
R4	To defining the foundations of a crossplatform open-services ecosystem.	Ad hoc
R6	To develop a model to compare the incentives to invest in operating system under open source and proprietary operating system.	Mathematical
R8	To elaborate approaches for how involvement of different roles can be analysed through quantitative analysis	SNA
R10	To automate the analysis of the evolution of software ecosystems	Tabular
R12	To presents an overview of the open source Ruby ecosystem	Metamodel
R13	To propose the OSCOMM framework for studying the problem of building open source communities.	Ad hoc
R14	To show how developers collaborate with each other within an software ecosystem across project boundaries.	Metamodel
R15	To model the NASA Earth science data systems ecosystem.	Conceptual
R16	To support modeling and evolution of quality from different points of view.	Metamodel
R17	To illustrate a co-creation process model of the Eclipse software ecosystem.	Ad hoc
R19	To study the GNOME ecosystem and developer community	Ad hoc
R22	To model the relationships between FOSS4G software ecosystem projects.	Ad hoc
R24	To model the software ecosystem that arise for open architecture systems.	Class diagram
R25	To map out the social history of collaborative activities within the OSGeo ecosystem	SNA
R27	To show the interactions in digital business ecosystems (as part of DBE European project).	Ad hoc
R28	To show the practices users have developed to manage the antagonism of maintaining a stable and productive working environment	Ad hoc
R31	To identify a model linking factors affecting the economics of collectives, and develop to economic outcomes.	Mathematical
R32	To model the relationships between FOSS4G software ecosystem projects.	Ad hoc
R33	To propose a three-layer view model of a software ecosystem.	Ad hoc
R34	To model Eclipse platform project ecosystem.	Mathematical
R35	To propose a framework for assessing the three dimensions of software ecosystem health.	Tabular
R36	To present a conceptual overview that describes the structure of an ecosystem associated model.	Class diagram
R37	To Analyse the evolution of social aspects of open source software ecosystems.	SNA
R38	To show some possible symbiotic relations between Linux and other software systems.	Ad hoc
R39	To present the software ecosystem metamodel that the small project observatory implements.	Metamodel
R41	To present the open software enterprise model that enable to establish the degree of openness of a software producing organization.	Tabular
R44	To propose a framework for sustainable software ecosystem management.	Tabular
R46	To propose a structure for modelling ecosystem software licenses.	Class diagram
R47	To construct a model to theorize how firms create and capture value from OSS.	Conceptual
R48	To propose a model for classifying software ecosystems.	Tabular
R49	To illustrate the management practices in technology and innovation management processes in software ecosystem.	Ad hoc
R52	To contribute to the concept of BECO health.	Tabular
R53	To propose a framework that enable the empirical study of OSS ecosystem and their developer communities.	Tabular
R56	To propose a software ecosystem health framework.	Ad hoc
R57	To present a model for creating and sustaining communities on the information services platform of Japan.	Ad hoc
R59	To observe how key events in the mobile device industry have affected the WebKit collaboration network over time	SNA
R61	To present a conceptual model of the collaboration management process in a OSS community.	Conceptual
R62	To propose a framework for the OSS ecosystem health operationalization.	Tabular
R65	To present an analysis of the Python egg software ecosystem.	Class diagram
R68	To analyse the market-driven view of an OSS ecosystem.	Ad hoc
R69	To analyse the partnership model of the Eclipse ecosystem and the activity of different types of partners.	Conceptual
R70	To address an exploratory study of the solutions to variability in software ecosystem.	Ad hoc
R71	To present ESAO model. It is focused on analysing and alignment between all the different ecosystem dimension.	Ad hoc
R72	To use i* roles in OSS adoption strategy models.	i*
R74	To present an open software ecosystem for embedded devices.	Ad hoc
R75	To present a study about Key Performance indicators (KPI) for software-based ecosystems.	Statistical
R76	To analyse the differences and analogies between natural ecosystems and software ecosystems.	Statistical
R77	To explore the role of groups, sub-communities and business models within a high-networked open source ecosystem.	SNA
R79	To present a software ecosystem taxonomy.	Tabular
R83	To illustrate how strategic modelling using the i* framework can help in analysing different configurations in the software industry.	i*
R84	To develop methodologies for managing risks of FLOSS adoption and deployment in various application domains.	i*
R86	To analyze committers networks.	SNA
R87	To verify whether the SECO context maintains the high socio-technical congruence levels observed in many smaller scale FLOSS projects.	SNA
R88	To propose a collective intelligence (CI) approach for improving the free software adoption by small and medium-sized municipalities in Brazil	Ad hoc

Table B.19: OSSECO taxonomy terms

Term	Description
Active User	Active users comprise occasional developers and users who report bugs, but do not fix them [R35].
Actor	Actors are either users or contributor [R29].
Adopter	Who do not contribute directly to the platform, but use it to develop tools [R31].
Artefact	The software project is defined as a structured collection of artefacts linked by derivations and produced to support/provide a collection of in use behaviours in order to satisfy a set of user requirements [R16].
Author	The author is the person that actually made the changes to the committed files [R19].
Behavior	No definition found in the primary papers.
Bug Fixer	Who Fixes reported bugs [R64].
Bug Reporter	Who Reports bugs [R64].
Bug Tracking	Track bug tracker activity (e.g. bugs opened, closed, statuses changed) bug tracker increase the source code centrality for a developer [R11].
Commit	These are pieces of atomic changes done on the source code [R10]. The developers commits to the project react not only the technical contributions but also the social and collaborative aspect of those contribution [R25].
Committer	Who contribute directly to the platform making changes in the OSS data sources[R69].
Community	OSS community is a social ecosystem on its own and in junction with other OSS communities. However, it differs from other social networks in its hierarchical structure. [R13]. Set of individual and shared resources of people's time, effort, attention, skill, sentiment (beliefs and values), and computing resources are part of the socio-technical web of FOSS [R23].
Community Manager	Who is the responsible of the OSS community governance [R41].
Configurator	No definition found in the primary papers.
Contributor	They contribute in some form to the OSS project [R20]. Contributors obtain private benefits from the development of shared assets that are not available to free riders, who only use the asset [R29].
Coordinator	No definition found in the primary papers.
Data Source	Code repository containing all versions of the source code, the bug tracker containing all feature requests and problem reports as well as all the resolution process, and the mailing list(s) containing all the mails exchanged among developers and between users and developer [R10, R53].
Dependency	It defines work interdependence among the ecosystem members [R3]. It is a symbiotic relationships between ecosystem actors [R38].
Developer	They contribute to OSS projects code for the personal gratification that comes from increasing their reputation among peers [R20]. They are primarily volunteers [R64].
Dominator	Is the actor that control the value capture and value creation of the ecosystem. [R56]
Edge	A-B in the network is created if an actor B replies to a message earlier sent by an actor A [R8]. Edges between ecosystem actors represent projects on which they collaborated [R14].
Email	No definition found in the primary papers.
Entropy	As a system is modified its disorder or entropy always increase. This is know as software entropy [R34].
Event	OSSECO organized events where stakeholders are brought together that share an interest in the total ecosystem [R62].
Feature	It identifies new functionality and enable develop the software in a common and creative way [R1].
Forge	Are Open Source Software (OSS) repositories designed to support teams doing software [R26]
Foundation	It is a democratic model based on voting rights, or a benign dictatorship (such as the Linux kernel), leadership will bean extremely important aspect of the ecosystem's development [R5]. Foundations provide financial, organizational, and legal support to the broader free in OSS [R25]. This economic community produces goods and services of value to customers, who are themselves members of the OSSECO [R27].
Goal	The goal was to provide stakeholders in OSSECO with insight into their ecosystem development and the most important metrics that indicate success in these ecosystems [R62].
Integrator	System integrators deliver solutions by selling a stack of hardware, software, and services as one product [R20].
IRC	It is a real-time chat [R11]
Keystone player	A keystone player is an actor in the ecosystem, whose contribution to the ecosystem stimulates the health ofthe entire ecosystem [R41].
License	OSS may be defined as software released under the terms of a license that basically allows the licensee to use, modify, and redistribute, either gratis or for a fee [R1].
Mailing List	It Contain all the mails exchanged among developers and between users and developer [R10].
Market	It is a phenomenon that occurs when the good is a shared resource such as a file format or software platform [R27]. The market as a regular player in a software ecosystem, assuming it plays a role similar to that of other players, such as developers and user. The market as the ecosystem's energy source, arguing that it plays a significantly different role from other players. It can directly or indirectly affect other players and determine the success of a software product [R68].
Measure	It is an indicator for OSS community health.
Member	It can start by directly contributing to code without prior socialization [R11]. It take part in the OSS community membership program [R36]. Customers, who are themselves members of the ecosystem, the member organisms also include suppliers, lead producers, competitors, and other stakeholders. [R48].
Niche	The software ecosystem niche in which a given OSSECO lies [R24] . The software supply networks that reveal which software ecosystem instances (or niches) each system exists within [R24].
Niche Player	Usually form the main volume of the ecosystem actors drawing value from the keystones. A niche player aims to separate from the other niche players by developing special functions [R56].
Node	Actors as nodes, tied or connected by one or more specific types of interdependencies [R25].
Partner	There are naturally business partners, industrial partners and similar interest groups participating outside the range of the model that are an integral part of an OSSECO [R13].
Passive User	Passive users are all remaining users who just use the system [R58].
Platform	It is set of software and services [R4] typically managed by an OSS community [R44].
Product	A product is a set of software intensive systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way [R14].
Project	OSS projects are typical environments in which SECOs develop around the community [R12].
Repository	It is a system which keeps and manages source codes [R17].
Roadmap	It defines planning and time-to-market (or more strictly time-to-technology) [R41].
Role	They representing the interaction mechanisms between the various actors that constitute a software system [R3]. Important roles in OSS are users, developers, core developers and project leaders [R8].
Service	These services are provided, the organization benefits from making explicit and sharing knowledge with partners, since the knowledge does not need to be made explicit when the organization provides these services again. [R46].
Social Network	Network derived from the data sources of an OSS project [R8].
Source Code	No definition found in the primary papers.
Stakeholder	There are three main groups of stakeholders: the publishing entity with its allocated resources for the project, the industrial partners and theirs developers, and finally existing opensource communities and other individuals [R13].
Sub Community	This OSS community may be subdivided in (possibly overlapping) sub-communities. For example, one can distinguish between the user community, containing all individuals who use an executable version of the software system, and the developer community, containing all individuals who are in charge of maintaining and improving this software system over time [R53].
Super Repository	It represents a collection of version control repositories of the projects of an OSSECO [R14].
Survey	No definition found in the primary papers.
Transaction	Transaction may be finished over a period of minutes, hours, or even days thus the term, from a computational perspective, of long-lived or long-running transaction [R27].
Wishlist	No definition found in the primary papers.