CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The use of empirical methods in open source software research: Facts, trends and future directions
Authors	Stol, Klaas-Jan;Babar, Muhammad Ali;Russo, Barbara;Fitzgerald, Brian
Publication date	2019-05-18
Original Citation	Stol, K. -J., Babar, M. A., Russo, B. and Fitzgerald, B. (2009) 'The use of empirical methods in Open Source Software research: Facts, trends and future directions', 2009 ICSE Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development, Vancouver, BC, Canada, 18 May, pp. 19-24. doi: 10.1109/FLOSS.2009.5071355
Type of publication	Conference item
Link to publisher's version	https://doi.org/10.1109/FLOSS.2009.5071355 - 10.1109/FLOSS.2009.5071355
Rights	© 2009 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Download date	2026-10-15 01:25:08
Item downloaded from	https://hdl.handle.net/10468/15246

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The Use of Empirical Methods in Open Source Software Research: Facts, Trends and Future Directions

Klaas-Jan Stol[1], Muhammad Ali Babar[1], Barbara Russo[2] and Brian Fitzgerald[3]
[1]*Lero,* [1,3]*University of Limerick, Ireland,* [2]*Free University of Bolzano-Bozen, Italy*
[1]*{klaas-jan.stol, malibaba}@lero.ie,* [2]*barbara.russo@unibz.it,* [3]*brian.fitzgerald@ul.ie*

Abstract

Open Source Software (OSS) is a field of study with increasing interest of researchers. By its nature, OSS is especially suitable for empirical research. A great number of OSS related empirical studies have been conducted, but no effort has been made to systematically review the published evidence. This paper presents the results of a systematic review to investigate research topics and used methods in OSS related research. We present our results as facts and trends in this field and provide directions for future research.

1 Introduction

Over the last decade, Open Source Software (OSS) has received an increasing amount of attention of researchers. OSS as a field of study appears to be particularly suitable for empirical research, as there is an enormous amount of data freely available through public project repositories (e.g. SourceForge.net) and other public data (e.g. mailing lists). This is different from conducting empirical research in an industrial context, where data is collected from companies and practitioners, and can be an expensive undertaking.

Various researchers have reported on empirical research methods in Software Engineering (SE) [1-3]. Both researchers and practitioners need empirical methods to help evaluate and validate research results, as they provide evidence to support claims of superiority of tools, methods, etc. [3]. Kitchenham et al. [4] write how SE can benefit from such an evidence-based approach, and introduced Evidence-Based Software Engineering (EBSE). The core tool for EBSE is the Systematic Literature Review, also called Systematic Review (SR), which is a type of review "*aimed at gathering together and analysing all of the experimental results available for a topic in an objective, unbiased and consistent manner*" [5].

Although various researchers have provided overviews of 'what is known of OSS' [6-9], to the best of our knowledge no effort has been made to *systematically* review the empirical evidence in the field of OSS. Therefore, we decided to conduct a SR, focusing on the following research questions:

1. What topics have been empirical studied in OSS related research?
2. What OSS projects have been empirically studied?
3. What research methods and data collection methods have been used in OSS related research?

This paper proceeds as follows. Section 2 discusses our research method, Section 3 and 4 discuss 'Facts' and 'Trends', respectively. Section 5 presents our suggested research directions, and Section 6 concludes.

2 Research Methodology

We conducted a Systematic Review (SR) of the OSS literature by following the guidelines proposed by Kitchenham [10]. The SR involved two researchers: the principal reviewer and a secondary reviewer.

2.1 Systematic Review

A SR typically involves the following phases: planning a review, conducting the review, and reporting the review. Each phase has some steps, as listed below:

1. **Planning the review:** identify the need for a review; develop and validate a protocol.
2. **Conducting the review:** identify primary studies; select primary studies; assess the quality; extract data; synthesise the data.
3. **Reporting.** Report the results.

Before starting the review, a protocol is developed. The protocol defines the data sources, search criteria, inclusion and exclusion criteria, what data to extract and how to synthesise the data.

2.2 Study selection

The study selection in a SR is a multistage process. First, the data sources for relevant papers are

978-1-4244-3720-7/09/$25.00 © 2009 IEEE

identified. We selected the four editions of the International Conference on Open Source Systems (hereafter referred to as 'OSS Conference') as data source. We believe that this series of conferences is the most representative venue for OSS related papers.

Selecting studies for inclusion in the review can be done either manually or through an automatic search. A manual selection consists of scanning a pre-selected set of papers, in our case the proceedings of the OSS Conference. During inspection, studies are either included or excluded according to the criteria defined in the protocol. When performing an automatic search, a set of carefully constructed search strings is used to search in digital libraries. The resulting papers are then inspected for inclusion or exclusion.

For this review we decided to manually scan the proceedings, as the total number of studies (219) was limited. As mentioned, we defined criteria for including and excluding papers in the protocol. We decided to include all papers that present some empirical evidence in the context of OSS research. Only papers published in English were included, excluding several papers written in Italian in the first edition of the OSS Conference. Studies without empirical evidence, such as tutorials, posters, and so on, were also excluded.

The proceedings of the four OSS conferences were manually searched by the primary researcher (first author of this paper). Based on the criteria defined above, 64 studies were initially included from a total number of 219 papers. The secondary researcher (the second author of this paper) performed a cross-check on a random selection of 76 papers from the total of 219 papers. From this sample, 36 were found eligible for inclusion. There were eight disagreements which means there was a Kappa coefficient of agreement of 0.79, which can be considered "substantial agreement" [11]. Before going to the next phase, all disagreements were resolved by discussion, resulting in a set of 70 selected papers.

2.3 Quality assessment

After the initial selection of the papers, the next step of a SR is to perform a quality assessment. During the quality assessment, papers may be excluded depending on the minimum quality threshold defined by researchers in the protocol. After the quality assessment, only those papers that are considered to be of sufficient quality are included in the data extraction phase. During the quality assessment of the papers included in our SR, it was found that some papers did not actually present empirical research, or presented empirical results of tools evaluation. Others did not provide any conclusions, and two papers presented data based on search results from web search engines such as Google. After discussing these papers, we decided to exclude these papers because they were not expected to provide any useful data. Seven papers were excluded, leaving a total of 63 for the data extraction phase. The list of included papers in our review is available at http://staff.lero.ie/stol/.

Following the approach in [12], we extracted data immediately after performing the quality assessment. It was felt that for both the quality assessment and the data extraction steps, a paper must be read in relatively close detail. When doing the data extraction immediately after the quality assessment, the details of the paper are still fresh in a researcher's memory.

2.4 Data extraction and synthesis

During the data extraction, we enlisted keywords for all studies that summarised the focus of the studies. Based on these keywords, we found that many studies were related or similar to a certain extent. Based on this finding, we grouped similar keywords together, which resulted in a categorisation of the studies, based on studied topic.

In the following sections we will present this classification and other findings of our review. Following an approach by Zhang et al. [13], we have organised our results as 'facts', 'trends', after which we discuss some future research directions in this area.

3 Facts

3.1 Studied topics

We identified a categorisation of the studied topics within OSS research. Four categories were identified, which are discussed next.

"OSS Communities": hereafter referred to as category '*C*', is the largest category, with 25 studies (39.7%). Studies in this category have investigated aspects of OSS project communities. Typical studied aspects are: the lifecycle of communities, social ties and structure among OSS project participants, coordination and communication within communities and motivation of participants to join a project.

"Development and Maintenance": this category, hereafter referred to as category '*D*', contains 13 studies (20.6%). Papers in this category have studied aspects related to the development and maintenance practices in OSS projects. Studied aspects in this category are amongst others: requirements analysis, distributed development, defect lifecycle, defect fixing process in OSS, and sprint-driven development.

"Diffusion and Adoption": this category, hereafter referred to as category '*A*', has 18 studies (28.6%). These studies have investigated the spread of use and adoption of OSS by organisations. Examples of studied topics in this category are: incentives for companies to use OSS, migration to and adoption of OSS, interest in and perceptions of OSS, and industry involvement in OSS.

"Characteristics of OSS": the last category, hereafter referred to as category '*O*', has 7 studies (11.1%). These studies study aspects of OSS as a whole, and typically have large samples of hundreds or thousands of projects. These studies analyse aspects such as growth, evolution, innovation and quality of OSS. Given that it is relatively easy to acquire data of large samples of OSS projects through public repositories such as SourceForge, it is remarkable that there have been relatively few studies in this category.

3.2 Studied OSS projects

Regarding studied OSS projects, the 63 studies included in our review can be roughly divided into three groups:

1. the papers that study a limited number of projects that are explicitly enumerated (*n*=27);
2. the papers that study a large number of projects (typically hundreds or thousands), which are not enumerated (*n*=13);
3. the papers that do not study OSS projects at all, but instead focus on aspects such as migration/adoption of OSS, collaboration, release management, and so on (*n*=23).

Based on this grouping, we made this distinction for each of the four categories that we identified. This is shown in Figure 1, which shows that the distribution is fairly uniform in category '*D*', whereas most studies in category '*A*' do not investigate OSS projects at all. On the other hand, most studies in category '*C*' study one or more particular OSS projects.

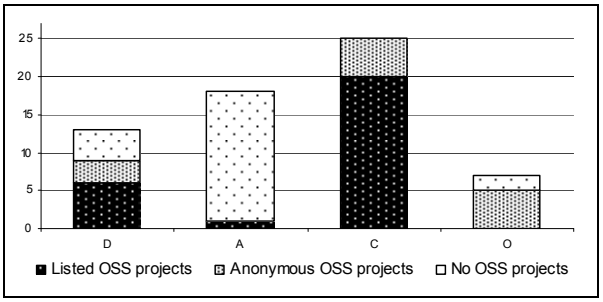

Figure 1: Three groups of studies: 1: enumerated projects; 2: anonymous projects; 3: no projects, specified per study category.

Category '*O*', finally, contains mainly studies that have studied large samples of projects that are not explicitly enlisted.

In order to understand the types of software that have been studied, we applied the software classification scheme from SourceForge.net. Naturally, this analysis could only be done for the first group of 27 studies, which explicitly enumerate the OSS projects being studied. The classification is shown in Figure 2. The figure shows that most studied OSS projects fall in the categories "system" and "internet".

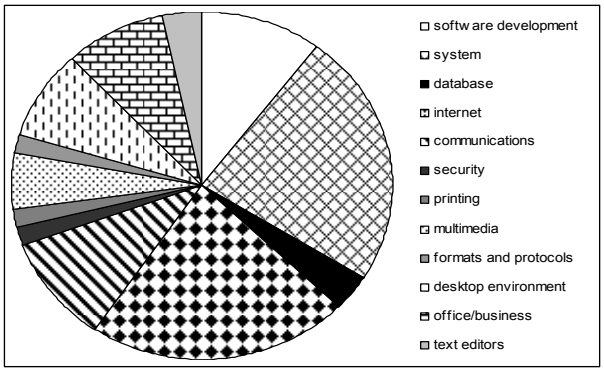

Figure 2. Types of studied OSS projects, using the classification from SourceForge (n=59).

In the 27 studies that explicitly list the studied OSS projects, we identified 34 unique projects. A number of these 34 projects seem of particular interest to OSS researchers, as they have been investigated in different studies. Also, a number of studies investigated several OSS projects. For the analysis of type of software, we counted all projects in all studies (including duplicates), resulting in a total count of 59.

Table 1. Most studied OSS projects

Project name	Times studied	% of total
Apache web server	7	11.7%
Mozilla web browser/Firefox	5	6.7%
Debian	4	6.8%
KDE, Linux kernel, GNOME, Gaim	3	5.0%
PostgreSQL, LUGs, Fire, Evolution	2	3.3%

The most studied projects are listed in Table 1. The table shows that the Apache web server project is of most interest, having been studied seven times. Five studies have investigated the Mozilla web browser/"Firefox". The Debian project has been studied four times, whereas the KDE, GNOME, Gaim

and Linux kernel projects have each been studied three times. PostgreSQL, Evolution, Linux User Groups (LUGs) and Fire are all projects that have been studied two times. From the 27 papers that explicitly listed the studied projects, 22 studied one or more of the 11 projects listed in Table 1.

3.3 Research methods

A variety of research methods has been used in the 63 reviewed studies. We recorded the research methods as they were stated by the authors of the papers included in our review. In some cases, the research method was not explicitly claimed, but merely mentioned casually. For instance, one paper mentions the word 'case study' to introduce a software library that is being investigated. In the conclusion, the authors state that 'more case studies' should be conducted. We recorded this as 'case study' methodology without any critique. The methods are listed in Table 2 for each of the four categories that were identified in Section 3.1.

Table 2. Research methods per category of study topics.

Research meth.	Cat. *C*	Cat. *D*	Cat. *A*	Cat. *O*
Case study	8	5	6	1
Survey	6	1	7	1
Quant. Analysis	12	3	1	5
Grounded theory	-	1	2	-
Action research	-	1	-	-
Experiment	-	1	1	-
Ethnography	-	1	-	-
Field study	-	-	2	-
Not specified	2	-	-	-
Avg. #meth/study	1.20	1.00	1.06	1.00

The last row contains the average number of research methods per study; only some studies in category '*C*' seem to use multiple methods. The case study method was used quite a number of times in each category, except in category '*O*'. The survey method was mostly used in the categories '*C*' and '*A*'. A quantitative method was used in all categories, but particularly in category '*C*'. We considered a topological network analysis to be a quantitative analysis. Other research methods were also applied a few times. Grounded theory approach was applied three times in total, action research only once, an experiment was conducted twice, and an ethnographical approach was taken once. Only two field studies were conducted. From two papers it was not clear what research method was used.

4 Trends

4.1 Study categories

We further analysed the study distribution over the four years of OSS conference's life using the four categories (i.e. all 63 studies) identified in Section 3.1.

Figure 3 shows this distribution.

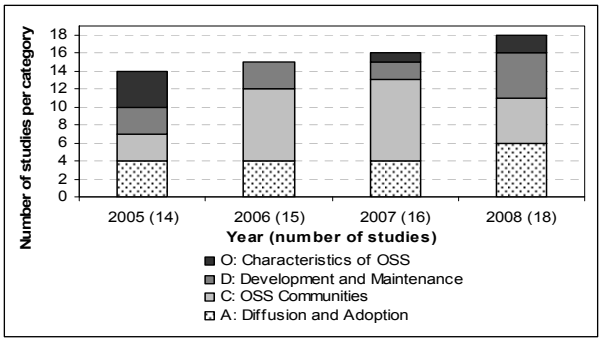

Figure 3. Evolution of reviewed studies per category, per year of the OSS Conference.

It is interesting to compare the distribution of empirical studies with the themes of the four conferences. In the 2005, the distribution of studies over the categories appears to be uniform, which seems to match the fact that 2005's conference did not have a focused theme. In 2006 the conference did not have a particular theme either, but the figure shows that the majority of the studies investigated fall into category '*C*'. In 2007, the theme was "Development, Adoption and Innovation". However, the figure does not reflect this; a majority of studies was still studying OSS communities. There were even less studies in category '*D*' than in 2006. Although 2007's conference theme was "Development, Adoption and Innovation", an equal number of studies in category '*A*' were published as in 2005 and 2006. In 2008 the conference theme was "Development, Communities and Quality". The conference presented more studies in category '*O*' (which includes studies on quality), and a larger number of studies in category '*D*'. Category '*C*', on the other hand, was smaller than the two previous years.

4.2 Studied software

We analysed how the types of open source software projects evolved over the four years of the OSS conference's lifetime. Again, we only consider the 27 studies that explicitly enumerated the OSS projects under investigation (see Section 3.2). The results are shown in Figure 4. The colours correspond to the

legend shown in Figure 2. As shown in the first pie-chart, the distribution of the types of software was uniform over five different types ("database", "system", "internet", "software development", and "formats and protocols"). However, this is because there were only five different projects studied in 2005, each studied once.

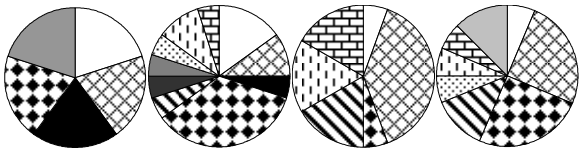

Figure 4. Distribution of studied software types over time. From left to right: 2005 (n=5); 2006 (n=20); 2007 (n=18); 2008 (n=16).

The second pie-chart representing 2006 shows a more varied distribution, with a majority of studies focusing on software projects in the "internet" category. The third pie-chart (2007) shows a very different distribution, with a majority of software projects in the "system" category. The categories "office/business", "desktop environment" and "communications" together make up half of the studied projects. The last pie-chart (2008) shows again a different distribution, now with half of the projects in the "internet" and "system" categories. The figures show that the categories "system" and "internet" are studied thoroughly throughout the OSS conference's editions.

4.3 Data collection methods

A plethora of data collection methods has been used in the 63 reviewed studies. In order to learn what different methods have been used and whether there has been any trend over time, we decided to analyse the data collection methods over time. The results are shown in Figure 5. We considered mailing lists, discussion lists, forums and IRC logs as one type of data collection method, as they all contain communication between an OSS community's members. Many studies use several data collection methods, in which case we counted each one of them. Figure 5 shows that the used data collection methods are varying per year. The use of questionnaires was high in 2005 and 2007, but not in 2006 and 2008. Use of community communication (mailing lists etc.) was low in 2005-2006, but was used more after that. Project data (repositories, websites, etc.) was used much in 2005, but was used less as a data source after that. Interviews were a large source of data in 2006-2007, but not in 2005 and 2008.

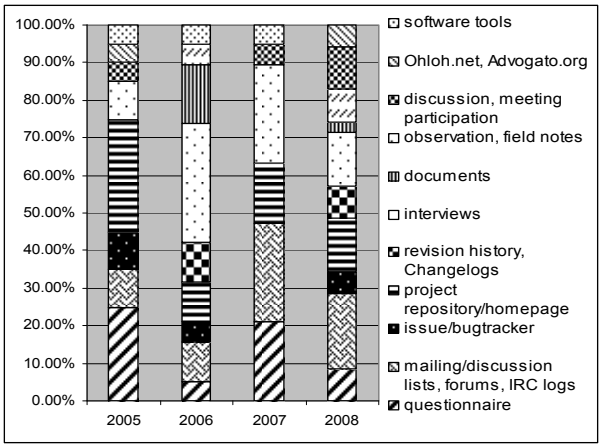

Figure 5. Data collection methods over time.

We have also analysed the most used data collection methods per category. The results were not surprising: studies in category '*C*' used mainly community communication data, such as mailing lists; studies in category '*D*' mainly used project repositories; studies in category '*A*' depended mainly on interviews, and studies in category '*O*', studying characteristics of OSS, thus collecting data of many projects, used primarily project repositories.

5 Research directions and limitations

Based on our findings, below we propose some directions for future research.

- *Aspects of OSS.* In Section 3.1 we have identified four different categories of studies. The numbers of studies per category indicate that most research has been done on OSS communities. Furthermore, we note that only 11.1% of the studies fall into the category 'Characteristics of OSS'. This is particularly remarkable, as empirical research in OSS is suitable for such type of studies, given the large amounts of available data in public repositories such as SourceForge.
- *Taxonomy of OSS research.* We note that our classification into four categories is very rough. However, we see this as a first step towards developing a taxonomy of research in OSS, as Glass et al. have done for SE [14].
- *OSS project diversity.* Most studies seem to investigate projects in the "system" and "internet" categories. The type of studied software may have an effect on the study outcome. Furthermore, a relatively small number of projects have been investigated several times. Future research should focus on a wider variety of projects so that the external validity of current results can be checked.

- *Research methods.* A variety of research methods has been used; however, it seems that case study, survey and quantitative analysis are most popular. As each research method has its benefits and limitations, it is important that other research methods are consistently applied as well, so that limitations of the previously mentioned research methods can be overcome. Most studies use only a single research method. However, the use of multiple research methods is expected to produce more reliable results. We also assert that OSS researchers can benefit from guidelines and best practices for producing empirical studies, such as is done in the field of SE. A first attempt towards this goal can be found in [15], in which the first two authors of this paper have proposed guidelines for reporting OSS related empirical studies.

5.1 Limitations of this review

The present study has some limitations. Firstly, the scope of our review was limited to papers published in the conference proceedings of the four editions of the OSS Conference, implying a bias in the selection of the reviewed publications. We are planning to extend the scope of our review to include papers searched from all the relevant literature, by performing automatic searches.

The selection procedure for including studies was performed by a single researcher. To minimise the selection bias, a sample of the initial selection was cross-checked by a second researcher. Both researchers recorded reasons for inclusion and exclusion.

6 Conclusion and future work

This paper presents the results of a systematic review (SR) on OSS related empirical research. We plan to extend our review to include papers searched from all relevant literature. Our contributions include the identification of some facts and trends in OSS related research. Secondly, this work provides directions for future research. We anticipate that this paper contributes to the discussion of how the OSS research community can improve empirical research in this field, for instance by studying a wider variety of OSS projects and using multiple research methods to study the same phenomenon. We envision a community-wide effort to develop and evaluate a framework for conducting and reporting empirical studies of OSS. We believe that the workshop on Emerging Trends in FLOSS Research and Development can provide a platform to stimulate discussion and effort for developing such a framework.

Acknowledgements

This work is partially funded by IRCSET under grant no. RS/2008/134 and by Science Foundation Ireland grant 03/CE2/I303_1.

References

[1] Seaman, C.B., Qualitative Methods in Empirical Studies of Software Engineering. IEEE Trans. Softw. Eng., 1999. 25(4): p. 557-572.

[2] Easterbrook, S., et al., Selecting Empirical Methods for Software Engineering Research, in Guide to Advanced Empirical Software Engineering. 2008. p. 285-311.

[3] Wohlin, C., M. Höst, and K. Henningsson, Empirical Research Methods in Software Engineering, in Empirical Methods and Studies in Software Engineering. 2003. p. 7-23.

[4] Kitchenham, B.A., T. Dybå, and M. Jorgensen. Evidence-based software engineering. in Software Engineering, Proceedings. 26th International Conference on. 2004.

[5] EBSE Group. Evidence-Based Software Engineering. 2007 August 5, 2008 [cited 2009 January 14]; Available from: http://www.dur.ac.uk/ebse/home_first.php.

[6] Nelson, M., R. Sen, and C. Subramaniam, Understanding Open Source Software: A Research Classification Framework. Communications of the Association for Information Systems, 2006, 17: p. 1.

[7] Feller, J. and B. Fitzgerald, Understanding Open Source Software Development, 2002, Addison-Wesley.

[8] Scacchi, W. and M.V. Zelkowitz, Free/Open Source Software Development: Recent Research Results and Methods, in Advances in Computers. p. 243-269.

[9] Feller, J., et al., Perspectives on Free and Open Source Software. 2005, The MIT Press.

[10] Kitchenham, B., Guidelines for performing Systematic Literature Reviews in Software Engineering. Tech Report, Keele University, 2007.

[11] Landis, J.R. and G.G. Koch, The Measurement of Observer Agreement for Categorical Data. Biometrics, 1977, 33(1): p. 159-174.

[12] Staples, M. and M. Niazi, Systematic review of organizational motivations for adopting CMM-based SPI, Information and Software Technology, 2008, 50(7-8): p. 605-620.

[13] Zhang, H., B. Kitchenham, and D. Pfahl. Software Process Simulation Modeling: Facts, Trends and Directions. in Software Engineering Conference, 2008. APSEC '08. 15th Asia-Pacific. 2008.

[14] Glass, R.L., I. Vessey, and V. Ramesh, Research in software engineering: an analysis of the literature. Information and Software Technology, 2002, 44(8): p. 491-506.

[15] Stol, K. and M. Ali Babar, *Reporting Empirical Research in Open Source Software: The State of Practice*, *to appear in International Conference on Open Source Systems,* 2009, Springer.