

# Reporting Empirical Research in Open Source Software: The State of Practice

Klaas-Jan Stol and Muhammad Ali Babar  
Lero—The Irish Software Engineering Research Centre  
University of Limerick  
Limerick, Ireland  
{klaas-jan.stol, muhammad.alibabar}@lero.ie  
WWW home page: <http://www.lero.ie>

**Abstract. Background:** The number of reported empirical studies of Open Source Software (OSS) has continuously been increasing. However, there has been no effort to systematically review the state of the practice of reporting empirical studies of OSS with respect to the recommended standards of performing and reporting empirical studies in software engineering. It is important to understand, how to report empirical studies of OSS in order to make them useful for practitioners and researchers.

**Research aim:** The aim of our research is to gain insights in the state of the practice of reporting empirical studies of OSS in order to identify the gaps to be filled for improving the quality of evidence being provided for OSS.

**Method:** To that end, we decided to systematically review the empirical studies of OSS. A total of 63 papers reporting empirical studies were selected from the four editions of the Proceedings of the International Conference on Open Source Systems. The data were extracted and synthesised from the selected papers for analysis.

**Results and conclusions:** We have found that the quality of the reported OSS-related empirical studies needs to be significantly improved. Based on the results of our systematic review and general principles of reporting good empirical research, we present a set of guidelines for reporting OSS-related empirical studies. The suggested guidelines are expected to help the research community to improve the quality of reported studies.

**Keywords:** empirical research, open source, reporting guidelines

## 1 Introduction

Since the introduction of the term ‘Open Source Software’ (OSS) in February 1998 [1], OSS has received an increasing amount of attention. The OSS as a field of study appears to be particularly suitable for empirical research, as there is an enormous amount of data freely and easily available through public project repositories, such as SourceForge.net, public mailing lists, and other public data such as IRC (Internet

Relay Chat) logs. This scenario is different from conducting empirical research in an industrial context, where data is collected from companies and practitioners, and can be considered an expensive undertaking.

Hence, it should not be a surprise that a lot of OSS-related empirical research has been reported. Other fields within Software Engineering, where data collection is more costly and difficult seem to have much less empirical studies. For instance, a recent literature review of empirical research in Agile Software Development found only 36 empirical papers [2]. A higher but still limited number of empirical studies was found in [3], which reported a review of empirical research in Global Software Engineering (GSE). The researchers found 12 (18.2%) empirical papers from a set of 66 papers and posters.

Of great importance in empirical research is the quality of the reported studies [4, 5]. Researchers have described the standard of empirical research in Software Engineering being ‘poor’ [4]. The strength of evidence in empirical studies on agile software development was found to be ‘very low’ [2]. In order to improve the quality of empirical research in SE, several researchers have proposed guidelines for conducting and reporting empirical research in software engineering [4, 5]. We believe that the quality of a reported empirical study affects the impact of the evidence. For researchers and practitioners, it is important to understand what the current state of reporting empirical research of OSS is, so that presented evidence can be interpreted correctly. As the community is growing and empirical research is increasing, we believe that it is important to start a community-wide debate on how OSS-related empirical research results should be reported. To the best of our knowledge, no effort has been made to systematically review the state of practice of reporting empirical studies of OSS. The research reported in this paper has been motivated by the following research questions:

1. What kind of empirical studies have been conducted in OSS-related research?
2. What is the quality of reporting empirical studies in OSS-related research?
3. How can the quality of reporting empirical research of OSS be improved?

In order to answer these research questions, we have conducted a Systematic Literature Review (SLR), also called Systematic Review (SR), of the research papers published in the proceedings of the four editions of the International Conference on Open Source Systems. Out of a set of 219 papers, we have classified 63 papers (28.8%) as empirical studies. Based on the analysis of the data extracted from the selected papers, we have identified four general categories of OSS-related empirical studies. Our results also show that there is a need for improving the quality of reporting empirical studies of OSS. To help improve the state of practice, we have made a few recommendations based on the standards and guidelines proposed to conduct empirical research in software engineering [4-6].

## 2 Research Methodology

We have conducted a Systematic Review (SR) of the OSS literature by following the guidelines proposed by Kitchenham in [7]. She describes a systematic review as: “A systematic literature review (often referred to as systematic review) is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest”. Our SR involved two researchers: the principal reviewer (first author) and a secondary reviewer (the first author’s supervisor).

### 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 in Table 1. The reasons for conducting this SR have partially been addressed in the introduction and stem from the perceived need for systematically extracting and synthesising, and critically analysing the literature published on empirical studies of OSS. In this paper, we present results from the first stage of our SR that is based on the papers published in the four editions of the OSS Conference. We believe that this series of conferences is the most representative venue for publishing OSS-related papers.

**Table 1.** Phases and steps of a Systematic Literature Review

Phase No.	Phase	Steps
1	Planning the review	Identify the need for a review Develop and validate a review protocol
2	Conducting the review	Identify primary studies Select primary studies Assess the quality of primary studies Extract data Synthesize data
3	Reporting the results	Write the report

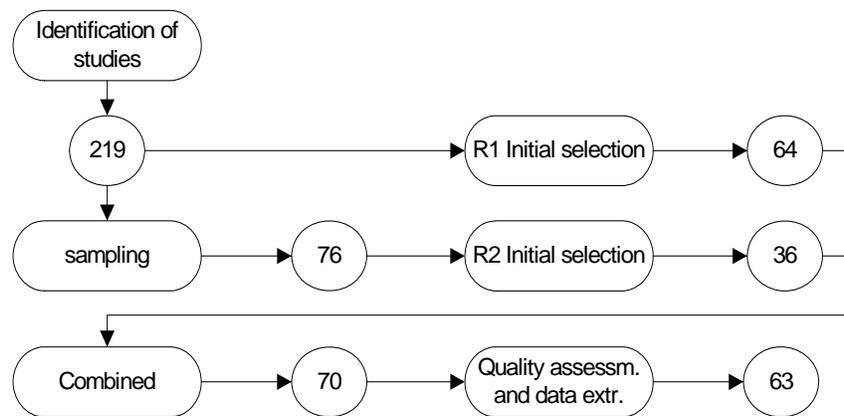
In a SR, researchers usually search all the relevant digital libraries using a set of well-constructed search strings that are expected to yield as many relevant results as possible. Defining these search strings is therefore extremely important. Since the scope of this SR is limited to the four conference proceedings, we decided to manually scan these proceedings in order to select all relevant papers.

Since not all studies are usually relevant to the SR being carried out, researchers need to define inclusion and exclusion criteria [7]. We also defined criteria for including and excluding papers in our review prior to conducting the papers selection process. We decided to include all papers that presented some empirical evidence in the context of OSS research. We decided to include only papers published in English, which was why we excluded the papers written in Italian published in the first OSS conference. Studies without empirical evidence, including tutorials,

posters, panel sessions, workshop briefs, experience or “lesson learned” reports were also excluded.

## 2.2 Study selection

The study selection in a SR is a multistage process. Figure 1 shows the selection process as performed in our study. First, the data sources for relevant papers are identified. It has been mentioned that we limited our scope to the OSS conference proceedings. These proceedings were searched manually by the first researcher. Based on the criteria presented above, 64 studies were initially included from a total number of 219 papers. The second researcher performed a cross-check on a random selection of 76 papers from 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” [8]. Before the next stage, all disagreements were resolved by discussion and 70 papers were selected.



**Fig. 1.** Flow diagram of the selection process. R1 refers to the first researcher, R2 to the second researcher.

## 2.3 Quality assessment

After the initial selection of the papers, the next step of a SR is to perform a quality assessment, after which data is extracted from the final selection of the papers. During the quality assessment, papers may be excluded depending on the minimum quality threshold defined by researchers. After the quality assessment, data is extracted from only those papers that are considered to be of sufficient quality. Following the approach in [9], we extracted the 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 just after the quality assessment, the details of the paper are still fresh in a researcher's memory. For the sake of clarity of our discussion, however, this paper describes the quality assessment and data extraction as separate steps. Moreover, the results of the quality assessment of the selected papers were also used to answer the question about the quality of reporting empirical studies of OSS.

**Table 2.** Quality assessment criteria; questions marked with (\*) were also used for data extraction

No.	Questions
1.	Does the study report any empirical research?
2.	Was there a clear description of the motivation for the research? <sup>1*</sup>
3.	Was there a clear description of the aim of the research? *
4.	Was there a clear description of the study context? *
5.	Was there an adequate justification for the research design? * Does the paper explain why the research is done this way?
6.	Was there an adequate description of the studied sample? *
7.	Was there an adequate justification for the selected sample?
8.	Was there a clear description of the data collection?
9.	Was there a clear description of the data analysis?
10.	Were the findings of the study clearly stated? *
11.	Was sufficient data presented to support the stated findings?
12.	Was the relationship between the researcher and the studied sample considered?
13.	Were limitations of the research adequately described? * This refers to limitations of the conducted study.

Kitchenham states that there are many published quality checklists for different types of empirical studies and has accumulated a list of questions organized per study type [4]. We took her advice to adopt Fink's suggestion described in [10] to select those quality assessment criteria that are most appropriate in the context of the posed research questions. We performed the quality assessment using a checklist of 13 criteria, listed in Table 2. These criteria were selected from the criteria presented in [2, 4, 5]. Some of these criteria (e.g. 1, 2 and 3) were presented as a single criterion, which can cause problems when using a binary value scale (i.e., Yes, No); in order to avoid this problem, we enumerated these sub-criteria separately.

We are aware that such a quality assessment is highly subjective. Our SR included both quantitative and qualitative empirical studies; in the latter case, assessing whether the paper presented sufficient data is particularly subjective. Nevertheless, the quality assessment was performed as objectively as possible, and despite the potential for subjectivity, we believe that the quality assessment has provided us with a global impression of the quality of the studies included in our SR.

<sup>1</sup> The single fact that "no research has been done" on a particular research question is considered a weak motivation, as this does not explain *why* the topic is interesting.

During the quality assessment, it was found that some papers included in the first stage 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, both researchers decided to exclude these papers because they were not expected to provide useful data. Seven papers were excluded at this point, leaving a total of 63 papers for the next phase of data extraction. The list of the papers included in this SR is available at: <http://staff.lero.ie/stol>.

## 2.4 Data extraction and synthesis

The objective of the data extraction activity is to extract the relevant data from the selected papers for synthesis and analysis in the following steps. One of the key tasks in conducting a SR is to design and evaluate a suitable data extraction form based on the research questions to be answered. The data extraction form is used for extracting and capturing the data. We designed a data extraction form based on the items marked with (\*) in Table 2 as well as some data extraction items found in [3] and [2] (such as statement of contributions to literature, stated hypothesis if any, used metrics, and study focus). The first author extracted the data from the selected papers and stored the data in a spreadsheet for analysis.

Given the variety of research methods used and types of data presented in the selected studies, a meta-analytical approach was not considered appropriate for data synthesis. We decided to inspect the extracted data for similarities in order to define how results could be compared. We focused on analysing the following data as they appear to be the most suitable to characterize OSS-related empirical research:

- Study focus – the aspect of OSS being investigated, for instance, collaboration or bug fixing.
- Studied projects – the project or community under investigation, if the sample was small enough to mention this. In case of a sample of thousands of projects, this is not applicable.
- Sample size – the number of projects/communities being studied. In many cases this is 1, 2 or 3, but it can be as high as 80,000 projects.
- Research approach – the research approach taken in the study, such as case study and survey.

In order to categorize the reported empirical studies of OSS, we analysed these four types of data as follows. We identified and enlisted all keywords from the studies, which summarised the focus of the studies. Based on these keywords, we found that many studies were related or similar to a certain extent. Therefore, similar keywords were grouped together, which resulted in a classification of studies.

To analyse the studied projects or communities, all were listed and grouped, and then the groups were sorted on number of occurrences. The sample sizes were analysed, and sorted on size. For the research approach, we listed and grouped all approaches as they were reported in the studies. We intend to do a more extensive

data analysis when we extend this review in future. In the next section, we present and discuss the results of our SR.

### 3 Results

#### 3.1 Categorising empirical studies in OSS

As we have mentioned, during the data extraction phase we identified the papers based on the similarities, topics and research content in order to categorise the reported empirical studies of OSS. Based on our analysis of the data for similarities and differences of the studied topics and research, we listed keywords of all the included studies, and classified the studies into groups based on their related keywords. Our analysis found that the studies could be classified into four categories. The identified categories are:

- **OSS communities:** The studies investigating practices and participants in OSS communities, and including the communication, collaboration, social networking aspects.
- **OSS development and maintenance:** The studies investigating practices of development and maintenance in OSS.
- **Diffusion and adoption of OSS:** The studies investigating perceptions, factors for adoption by industry and usage of open source in organisations.
- **Characteristics of OSS:** The studies investigating properties of OSS as a whole, such as growth and evolution.

The first category is ‘OSS communities’; this was also the largest category, with 25 studies (39.7%). Most of the studied aspects in this category are: social networks of communities, lifecycle and evolution of communities and communication within communities. The second category is ‘OSS development and maintenance’. Thirteen studies (20.6%) were classified in this category. The focus of these studies was mainly on practices and issues in OSS development. The third category, containing 18 studies (28.6%) is related to ‘Diffusion and adoption of OSS’. In this category, the studied aspects are: perceptions of OSS, incentives to adopt OSS, migration to OSS and usage of OSS. The remaining 7 studies (11.1%) could all be classified in the fourth category that we call: ‘Characteristics of OSS’. These papers report on different characteristics of OSS in general, and typically have large sample sizes. The typical aspects studied in these papers are the growth and evolution of OSS, and the quality of OSS.

We have analysed the evolution of the distribution of studies in these categories. Figure 2 shows this distribution graphically. It was interesting to compare the distribution of empirical studies with the themes of the published proceedings. In 2005, the distribution of studies over the categories appears to be uniform. This fact seems to match the fact that 2005’s conference did not have a focus on a particular aspect of OSS. In 2006, the conference did not have a particular theme either, but the

figure shows that the majority of studies investigated in OSS communities. In 2007, the theme was “Development, Adoption and Innovation”. However, the figure does not reflect this; a majority of studies was still focusing on OSS communities. There were even less studies in the category Development and Maintenance than in 2006. The category Diffusion and Adoption has an equal number of studies as in 2005 and 2006. In 2008 the conference theme was “Development, Communities and Quality”. That year presented more studies on Characteristics of OSS (which includes studies on quality), and a larger number of studies on Development and Maintenance. The category Communities, on the other hand, was smaller than the two previous years.

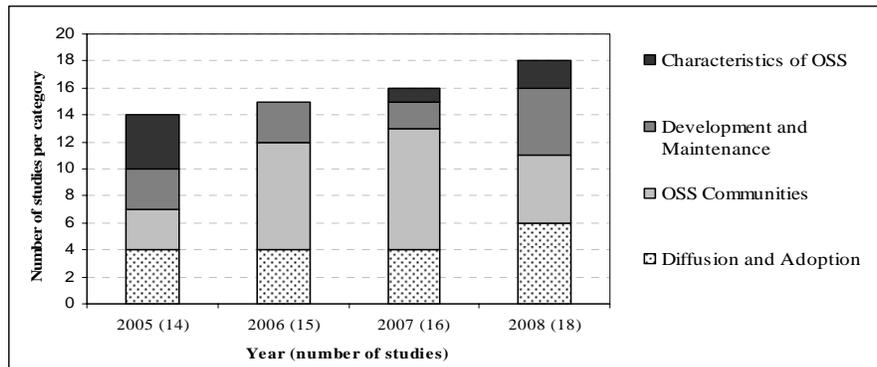


Fig. 2. Evolution of distribution of studies over identified study categories

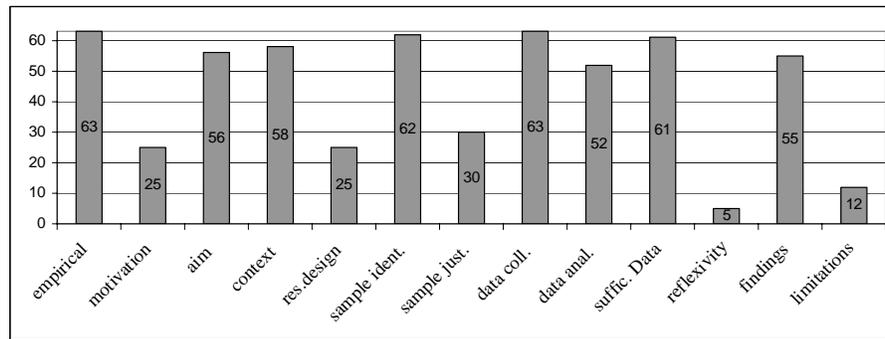
### 3.2 Results of quality assessment

The selected papers were subjected to a Quality Assessment (QA) using the checklist presented in Section 2.3. We found that a significant number of papers scored very poorly on a number of quality criteria. The results of the quality assessment are shown in Figure 3. The figure shows that 7 out of 13 quality criteria score quite well. It must be noted that the first criterion is whether the paper presents empirical data, and must therefore be true for each of the 63 included papers.

However, the score for the remaining five quality criteria is rather poor. It was found that in most cases, the motivation of the study was not clearly described. In many cases it was left to be implied or limited to a mere statement that “no research has been done” on the studied topic. We argue that this by itself is not a valid motivation because it does not explain to the reader why the researched topic is interesting in the first place. Most of the studies in this SR did not provide any justification for using a particular research design and only mentioned the name of the research approach used (such as survey, interview, mining mailing lists, and metrics).

It was also found that almost all studies (62) identified the sample, but more than half of the studies did not give any justification about the kind of sample used. In all

but five studies, the authors did not state their relationship to the studied subjects (reflexivity). Only 12 papers reported the limitations of the studies, which is an important aspect of reporting empirical research [4, 5, 11]. Of the 63 studies, a third scored at least 10 out of 12 quality assessment points. More than half scored eight or nine points, and the remaining six studies scored less than eight.



**Fig. 3.** Results of the quality assessment; number of studies that scored per criterion

## 4 Discussion

The Software Engineering (SE) community has been emphasising the need for guidelines for reporting empirical research [5, 12-14]. Such guidelines allow for a systematic and standardised way of presenting empirical findings. This will help both researchers and practitioners in several ways, as suggested in [5]:

- easily find the right information;
- understand the context of a study;
- to assess the validity of the findings.

In [4], a set of preliminary guidelines for controlled experiments was presented. In [5], a number of published proposals are surveyed, aiming to derive a unified standard. It is a common practice in OSS research to use publicly and freely available data from OSS repositories, with SourceForge.net as a well-known source of data. In that sense, OSS is somewhat different from other research in SE, where data is often collected from companies or individuals. Instead of a scarcity of data, the OSS community has a different problem: making sense of the enormous amount of available data.

In Section 3, we have established that the reporting of OSS-related empirical studies can be much improved by having a suitable set of guidelines. Moreover, such guidelines are also expected to help researchers to design and execute better quality studies. In this section, we propose a set of guidelines for reporting empirical research in OSS. These guidelines have been adapted from a recent effort to create a

reporting classification scheme for GSE [5]. We expect that the use of these guidelines will improve the reporting of empirical studies in open source research, and this in turn will help evaluate the generalisability and applicability of findings from empirical research of OSS.

## **4.1 Guidelines for reporting empirical research in OSS**

### **4.1.1 Motivation**

While there can be many aspects of OSS worth investigating, it should always be evident to a reader of a study of OSS as to why a certain research has been carried out. Jedlitschka and Pfahl state that the motivation section is to set the scope of the work, and provide readers with good reasons to read the remainder of the publication [5]. Our review has revealed that only 25 (39.7%) papers clearly stated the motivation for the undertaking the reported study. We believe that merely stating that the topic has not been investigated is a rather weak argument because this does not help to explain why the topic is interesting to researchers and practitioners. An explicitly stated motivation helps a reader to understand what is being studied. We suggest the motivation for conducting an empirical study of OSS be explicitly and clearly stated.

### **4.1.2 Research design**

Our SR has found that the majority of the studies do not clearly distinguish between the research method (e.g. action research, case study, ethnography, experiment [11]) and the data collection approach [6], or 'instrumentation' (e.g. interviews, questionnaire) [15]. A research method is usually selected based on its suitability to the problem being researched [16], which is why it is important to clearly report the research method used and justification and suitability for the choice. Such information is expected to help researchers and practitioners to assess the strength of evidence provided. In order to improve the reporting quality, we suggest that empirical research papers report both the applied research method and data collection techniques along with suitable justification.

### **4.1.3 Justification of research design**

It is also important that an appropriate research approach is taken in order to address the research question being investigated. It may not always be clear to the reader why a particular research approach was adopted. For both researchers and practitioners, it is quite helpful to know the reasons for using a certain research design in order to decide the relevance and reliability of the findings. We have found that less than half of the reviewed studies (39.7%) provided no justification for the research design used. This situation is quite disappointing as unless the justification of the used research method is known, it is very hard to justify the reliability of the findings. Therefore, we strongly encourage OSS researchers to provide sufficient justification for the research approach used.

#### **4.1.4 Sample description**

The sample sizes of the OSS projects used in the reported empirical studies vary from a single project to a very large number of projects. For example, one study used the data from more than 80,000 projects. The sample size of a study usually affects the reliability of the findings [16]. Additionally, it is also important to report the method (e.g., systematic random or convenience) used for selecting the sample from a sampling framework if there is one. That is why we assert that each study should report complete details about the sample size as well as the sampling method used in the study. In the case where a single or a few projects are studied, it is also important to report an accurate description of the project(s) under investigation. Such details can include: size of the OSS software (expressed as lines of code), size of community (expressed as number of active and inactive participants), and the domain of the OSS software (e.g. operating systems, desktop software, infrastructural such as web servers). Our review has revealed that 62 (98.4%) of the reviewed papers clearly identified the sample size used for the reported investigation.

However, there was hardly any study that reported the sampling methods used and justification for the choice of sampling method as emphasised next.

#### **4.1.5 Sample Justification**

In less than half of the reviewed studies (47.6%), the authors provided a justification for using a particular sample size. During the quality assessment, we did not consider the statement that a particular project is well-known or popular (i.e. Linux) as a valid justification. Justification of the studied sample (project) is important as it helps the reader to better understand what the researchers' aim was for the study. If a specific OSS project was chosen, then surely this was done because the researchers found the project interesting for a particular reason or they may have expected some interesting findings. Sharing such information with the readers helps to present a clear context of the study.

#### **4.1.6 Data collection**

The sample of projects selected for a study is usually the domain from which the actual data is gathered. In order to enable the readers to assess the amount of data gathered, a study should provide a clear description of the type and quantity of data (e.g. interviews, bug reports or mailing list posts). Reporting these details is expected to help evaluate the findings of a study. It can be argued that reporting the methods used for gathering and analysing data is important for evaluating the significance of the presented results of a study. Each data collection approach can help achieve certain research objectives and has its limitations [6], so a clear statement of the used data collection approach helps readers to understand the implications of the applied data collection approach.

#### **4.1.7 Research context**

Kitchenham et al. regard experimental context 'extremely important for software engineering research' [4]. Open source systems can be studied in different contexts. Firstly, researchers can use OSS for investigation from 'the sideline', where data is

gathered from open source repositories, such as SourceForge.net. In such a setting, data can be gathered in a non-obtrusive manner. On the other hand, when doing field research, researchers interact with the community or organisation directly and/or indirectly involved in the development or use of OSS. In these contexts, data is commonly gathered through surveys or interviews. There are mainly two different types of populations: OSS communities (i.e. OSS developers) and OSS users (individuals and organisations). Such difference is important because although companies can be contributing to OSS as well, they typically are the users of OSS. This context of the study is important for interpreting its findings. Therefore, it is vital to clearly report the context in which a study has been conducted.

Moreover, in order to categorise a study based on the aspect of OSS that is addressed (e.g. adoption, code quality, etc.), we suggest to clearly report the focus of the study. We believe that such information can greatly improve the identification of gaps in knowledge and commonalities of the studies. Furthermore, it also helps readers to better understand the issue being investigated.

#### **4.1.8 Reflexivity**

There can be instances where researchers conducting a study may have a particular relationship with the studied subject (project or community). For a reader of the reported study, it is important to understand this relationship, as it may affect the outcome of the study. Researchers may have been able to come to certain findings because of this relation. This should be considered in any attempt to replicate a study. In our SR, only five (7.9%) studies discussed the relation of the researchers with respect to the study subject. It is strongly recommended that such information be explicitly and completely reported in empirical study papers.

#### **4.1.9 Study limitations**

Each empirical study can face the risk of being affected by some validity threats (i.e., internal or external) or has some limitations. Our SR of empirical studies of OSS has found that only 12 (19.0%) studies discussed the limitations of the reported research. We argue that each empirical study should include an appropriate amount of discussion about the potential validity threats and limitations of the reported study, and any measures, if at all, taken to address some of the identified threats and limitations. Such information is important to help a reader to assess the credibility and validity of the findings. Kitchenham et al. describe a discussion on the limitations of a study as a responsibility of the researchers [4]. Typically, the internal and external validity should be discussed.

Internal validity refers to “the extent to which the design and conduct of the study are likely to prevent systematic error” [7]. Internal validity of a study means that the presented data supports a cause-effect relationship that is claimed in the study. In OSS-related research, this is important given that there is a large amount of data to be explored, and researchers may find many cause-effect relationships.

External validity refers to the generalisability of the findings outside of the studied context. In the context of OSS-related research, this is particularly important

in studies that focus on a single or a few OSS projects. What works in one OSS project may not be applicable to another.

#### **4.2 Implications for OSS research and practice**

The results of this SR have presented important information about the state of practice of reporting empirical research of OSS. First, it shows that there is a vital need to improve the quality of reporting empirical studies of OSS. We assert that an improvement in the empirical studies of OSS will help the community to better understand the results and limitations of the reported research.

We have presented a set of guidelines that are expected to help improve the quality of reported studies in OSS-related research. We do not claim that the set of guidelines we have proposed is exhaustive or complete. However, we believe that significant improvements can be made in the quality of reporting empirical research if the future papers on empirical studies of OSS provide all the information suggested by the guidelines. Furthermore, the results show that the empirical studies included in our SR can be classified into four categories. Such classification of empirical studies of OSS is expected to help researchers to put future research in the context of one of these categories. Furthermore, although the proceedings of the first conference on OSS show a more or less uniform distribution of studies over these categories, later editions of the conference proceedings show that there have been fewer studies in the category ‘Characteristics of OSS’. Studies in this category typically study properties of OSS based on large sample sizes, which implies a better generalisability of the results of these studies.

#### **4.3 Limitations of this review**

We are aware of some limitations of our study, which we discuss now. Firstly, the scope of our study was limited to papers published in the conference proceedings of the four editions of the International Conference on Open Source Systems. This means there is a bias in the selection of the reviewed publications. We are planning to do a full-scale SR of the empirical research in OSS that would include papers searched from all the relevant literature.

The selection procedure for including the studies is somewhat subjective. In order 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.

We acknowledge that the quality assessment is highly subjective. This was especially an issue as papers are not consistent in reporting the studies; some are very clear and explicit whereas others leave a lot of details implied. However, one of the goals of this research is to investigate how current research is reported, and how this can be improved.

## 5 Conclusion and future work

This paper presents the results of the first stage of a systematic literature review on OSS-related empirical research. This stage is limited to the studies reported in the four editions of the Open Source Systems conferences. From a set of 219 papers, we included 63 in our SR. We performed a quality assessment and extracted data from these 63 studies. We found that the selected studies could be classified into four categories: OSS communities, Development and Maintenance, Diffusion and Adoption, and Characteristics of OSS. Furthermore, our study has revealed that the quality of reported empirical research on OSS has significant room for improvement. To that end, we have proposed a set of guidelines for reporting empirical research on OSS. We claim that these guidelines can help the OSS research community to improve the quality of designing and reporting empirical studies.

We intend to extend our SR to include more studies by searching the well-known digital literature databases. We will also extend our data analysis as we plan to discover the trends and future directions of OSS research.

## Acknowledgements

This work is partially funded by IRCSET under grant no. RS/2008/134 and by Science Foundation Ireland grant 03/CE2/I303\_1 to Lero—The Irish Software Engineering Research Centre ([www.lero.ie](http://www.lero.ie)).

## References

- [1] Open Source Initiative. History of the OSI. Last accessed on October 16, 2008, Available from: <http://www.opensource.org/history>.
- [2] T. Dyba and T. Dingsoyr, Empirical Studies of Agile Software Development: A Systematic Review, *Information and Software Technology*, 2008. **50**(9-10): pp. 833-859.
- [3] D. Šmite, C. Wohlin, R. Feldt, and T. Gorschek, Reporting Empirical Research in Global Software Engineering: a Classification Scheme, *International Conference on Global Software Engineering*, 2008.
- [4] B.A. Kitchenham, et al., Preliminary guidelines for empirical research in software engineering, *IEEE Transactions on Software Engineering*, , 2002. **28**(8): pp. 721-734.
- [5] A. Jedlitschka and D. Pfahl, Reporting Guidelines for Controlled Experiments in Software Engineering, *Proceedings of the International Symposium on Empirical Software Engineering*, 2005.
- [6] T.C. Lethbridge, S.E. Sim, and J. Singer, Studying Software Engineers: Data Collection Techniques for Software Field Studies, *Empirical Software Engineering*, 2005. **10**: pp. 311-341.
- [7] B. Kitchenham and S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, *Tech Report EBSE-2007-1*, Keele University, UK, 2007.

- [8] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics*, 1977. **33**(1): pp. 159-174.
- [9] M. Staples and M. Niazi, Systematic review of organizational motivation for adopting CMM-based SPI, *Information and Software Technology*, 2008. **50**(7-8): pp. 605-620.
- [10] A. Fink, *Conducting Research Literature Reviews: From Internet to Paper*. 2005: Sage Publication, Inc.
- [11] C. Wohlin, et al., *Experimentation in Software Engineering: An Introduction*. 2000: Kluwer Academic Publications.
- [12] B. Kitchenham, et al., Evaluating guidelines for reporting empirical software engineering studies, *Empirical Software Engineering*, 2008. **13**(1): pp. 219-221.
- [13] D. Sjoberg, T. Dyba, and M. Jorgensen, The future of empirical methods in software engineering research, in *Proceedings of the International Conference on Software Engineering*, Future of Software Engineering Track. 2007.
- [14] M. Host and P. Runeson, Checklists for Software Engineering Case Study Research, *Proceedings of the First International Symposium on Empirical Software Engineering and Measurement*, 2007.
- [15] B. Kitchenham and S.L. Pfleeger, Principles of Survey Research, Parts 1 to 6, *Software Engineering Notes*, 2001-2002.
- [16] J. Miller, et al., Statistical power and its subcomponents - missing and misunderstood concepts in empirical software engineering research, *Information and Software Technology*, 1997. **39**: pp. 285-295.