Writing for Synthesis of Evidence in Empirical Software Engineering

Claes Wohlin
Blekinge Institute of Technology
SE-37136 Karlskrona
Sweden
+45-(0)455-385820
Claes.Wohlin@bth.se

ABSTRACT

Context: Systematic literature reviews have become common in software engineering in the last decade, but challenges remain.
Goal: Given the challenges, the objective is to describe improvement areas in writing primary studies, and hence provide a good basis for researchers aiming at synthesizing research evidence in a specific area.
Method: The results presented are based on a literature review with respect to synthesis of research results in software engineering with a particular focus on empirical software engineering. The literature review is complemented and exemplified with experiences from conducting systematic literature reviews and working with research methodologies in empirical software engineering.
Results: The paper presents three areas where improvements are needed to become more successful in synthesizing empirical evidence. These three areas are: terminology, paper content and reviewing.
Conclusion: It is concluded that it must be possible to improve the primary studies, but it requires that researchers start having synthesis in mind when writing their research papers.

Categories and Subject Descriptors

D.2 **[Software Engineering]**

General Terms

Documentation, Reliability, Experimentation, Standardization.

Keywords

Systematic literature reviews, research synthesis, systematic mapping studies, research methodology, guidelines, evidence.

1. INTRODUCTION

The number of systematic literature studies is increasing over the years as reported by da Silva et al. [4]. Given the increasing number of systematic literature reviews, the quality of them has been evaluated by Cruzes and Dybå [3]. Unfortunately, the authors conclude that many papers do not conduct any synthesis, and other syntheses are quite weak. According to the authors, only a few papers report a rigorous synthesis out 49 studies analysed. The authors conclude that limited attention is paid to synthesis. However, the ability to conduct a rigorous synthesis is heavily dependent on the primary studies and their quality that form the input to the synthesis. If the input to the synthesis is of low quality, then we cannot expect rigorous syntheses.

ESEM'14, September 18–19, 2014, Torino, Italy.

http://dx.doi.org/10.1145/2652524.2652559

Several authors report on the low quality in primary studies for synthesis purposes. This is further elaborated below in Section 2. Based on the observed low standard of syntheses presented in systematic literature reviews combined with the findings with respect to the quality of the primary studies, this paper presents some improvement areas with respect to primary studies. This relates to terminology, paper content and reviewing of papers.

The remainder of the paper is outlined as follows. Section 2 presents some of the related work in the area. In Section 3, an example is presented illustrating some of the challenges in analysing primary studies, in particular in relation to information missing in primary studies. Section 4 presents some improvement areas to hopefully increase the quality of primary studies with the objective to make them more suitable for synthesis in systematic literature reviews than all too many primary studies today.

2. RELATED WORK

Kitchenham et al. [9] introduced the concept of evidence-based software engineering ten years ago. The basis for synthesizing evidence from literature is to conduct systematic literature reviews as described in the guidelines by for example Kitchenham and Charters [10] and Wohlin [18]. However, synthesizing evidence has proven to be a challenge as reported by Guzman et al. [5]. Guzman et al. also point to the primary studies being a major concern when it comes to the ability to conduct synthesis. They mention aspects such availability, heterogeneity and quality. These conclusions are inline with those reported by Cruzes and Dybå [3] and da Silva et al. [4].

Hassler et al. [6] report on a workshop where experienced researchers discuss some of the challenges with respect to systematic literature reviews. With respect to quality assessment, the highest ranked issues is the quality of the primary studies, which is also the second highest concern overall reported in Hassler et al.

Brereton et al. [1] highlight the difficulty with primary studies when they state that the quality of the abstracts is too low, and hence in many cases it is impossible to judge the relevance of a paper for inclusion or exclusion in a systematic literature review based on the title and abstract.

The above observations are inline with our own experiences in conducting systematic literature reviews. Synthesis is a challenge, and it stems from both the challenges in performing synthesis as such, but without having high quality primary studies the prerequisites for conducting synthesis is not really in place. Thus, it can be concluded that higher quality primary studies are indeed needed. Next, this will be exemplified with an illustration from a systematic literature review in the area of global software engineering.

3. EXAMPLE

In 2010, Smite et al. [15] published a systematic literature review on empirical evidence with respect to global software engineering. The objective was to synthesize the evidence to better understand under which circumstances global software development works well and when not. This was perceived as important to be able to provide advice to industrial partners in a research project. Preceding the full-fledged systematic literature review was a pilot study resulting in some guidelines for reporting empirical research in global software engineering. It should be noted that Smite et al. [14] proposed the guidelines based on pilot study, which in itself indicates the state of the primary studies.

Going back to the systematic literature review [15], four aspects perceived as important to be able to synthesize evidence were:

- Global arrangements
  This refers to the setup of the collaboration. It includes aspects such as whether or not the arrangement is offshore insourcing or offshore outsourcing [16], and whether it was professionals in industry or student projects. This is important to know. It is quite different to conduct outsourcing to another company or working with another site of your own company, although in a different location, which may be far away. In 5% of the papers this was unclear. This is also closely related to the terminology used to describe the global setting as exemplified in Section 4.1.

- Perspectives of analysed studies
  The perspectives relate to whether the paper discusses collaboration in general or if the analysis is conducted from a specific perspective, for example from an originator of the global collaboration or the supplier of the work. For example, when subcontracting to a company located far away, it is important to know whether the study is conducted from the perspective of the originator or the company being subcontracted to. In 10% of the papers, this was unclear.

- Success evaluation
  If building on evidence from literature, it is of course important to know whether the actual global collaboration was successful or not. Unfortunately, 20% of the papers do not report on whether the collaboration was successful or not.

- Reasons for starting global collaboration
  Different companies have different reasons to setup global collaborations, and hence a company considering going more global would of course primarily like to build on the experiences from those that went global for similar reasons. However, this is only reported (or it is irrelevant) in about half of the papers. 47% of the papers did not report the reason for going global despite it being judged as relevant by the researchers [15].

The limited information points first of all to the need for authors to improve their presentations, but reviewers knowledgeable in the area should also be able to identify that some important aspects are missing in the presentation, although the primary responsibility lies with the authors.

In summary, the above shows that the context for many of the different studies is unclear, which complement the findings reported in Section 2. Assuming that the reporting of the four aspects above is independent, then only 36% of the papers report on all four ((1-0.05)*(1-0.10)*(1-0.20)*(1-0.47) = 36%).

Given that many sources point to the need of higher quality primary studies, a number of aspects to improve with respect to primary studies are presented next.

4. IMPROVEMENTS FOR SYNTHESIS

Unfortunately, papers are all too often written with out really having synthesis in mind. It is time to change. If assuming that authors would like others to find their papers, read them, and cite them, then it is time to become more consistent in the way research in software engineering is presented. Improvements with respect to primary studies may be divided into three areas: terminology, paper content and reviewing.

4.1 Terminology

Terminology is far from easy. This can be divided into two main problems areas: 1) the same word is used, but the interpretation is not the same, and 2) different words are used, but the same concept is meant. An example of the former became evident when trying to compare two systematic maps on the same topic [17]. The authors of the two maps both looked at testing of software product-lines. However, they used the word testing slightly differently, which meant that some papers included in one of the maps were considered non-relevant by the authors of the other systematic map. An example of the use of different words can be taken from the area of global software development. When conducting two systematic searches [7] for papers with respect to global software development and agile development with the objective of comparing different search procedures, it was found that the authors of one paper used the term “cross-continent” development. When conducting a snowballing search [7], the paper was found and it was clear that it should be included. However, it was not found when searching in the databases, since the used wording is very uncommon, and hence it was not included in the search strings. This illustrates some of the difficulties with terminology. Jalali and Wohlin [8] present the actual results from the systematic literature review. The following is needed:

- More consistent use of terminology – look at standards, and other papers in the area to ensure a more consistent usage of terminology,

- Terminology and taxonomies are needed to support other researchers. It is important that terminology and taxonomies are community based and not pushed onto the community. Smite et al. [16] present one example when they made a survey among experts in the area to try to come to a consensus regarding words to use.

4.2 Paper content

It may seem obvious that papers must have the right content, but given the example in Section 3, it is apparent that in too many cases authors do not write for synthesis, and hence they do not provide the information actually needed. Shaw [13] provides some general guidelines for writing research papers in software

engineering. However, there are some specific needs that deserve to be highlighted from a synthesis point of view. First, the information must be such that it becomes easy to judge whether a paper should be included or excluded in a systematic literature review. Secondly, if a paper is included then it should be easy to extract the information needed, and finally it must be possible to synthesize the actual data extracted (with data extracted from other papers). Thus, it is both writing for understandability of the paper being examined, and for being able to conduct the synthesis. These aspects are of course not independent. The needs are as follows:

- **Titles** should be as precise and informative as possible. The authors should think "synthesis". If conducting a synthesis in the area of my paper, would the researchers 1) find it and 2) be able to include it based on the title.
- **Keywords** describing and positioning the paper are important. The standards from publishers are often not very helpful for a systematic literature review. If allowed, it is important to add keywords useful for a systematic literature review that could complement the title.
- The **abstract** should be written last and very carefully. The authors should have synthesis in mind when writing the abstract. Potentially, structured abstracts may be used as suggested by Budgen et al. [2], and even if not writing a structured abstract formally, it is recommended to have the headings in mind when writing the abstract. The journal of Information and Software Technology uses the following five headings in their structured abstracts: context, objective, method, results and conclusion. These five subheadings have been used in the abstract of this paper too.
- In many cases, researchers perform selective reading when conducting systematic literature review. This means that the researcher focuses on some part of a paper to decide whether to include or exclude a paper. Depending on the research question in the systematic literature review, different parts of the primary studies/papers may be read. However, in particularly important parts are:
  - **Introduction** – it should clearly position the paper and help the reader to understand the contribution of the paper and its context.
  - **Research question** – the question(s) must be clearly formulated and cover the main contribution of the paper.
  - **Research method** – the method used conveys important information to the reader. Unfortunately, researchers do not use the terminology consistently, for example "case study" is often a misused research method. In many papers, the authors actually mean "example" or "illustration" and not a case study in the way it is defined, for example, by Yin [20] or Runeson and Höst [12]. Wohlin and Aurum [19] present a more in-depth discussion about the choice of research methods in empirical software engineering.
  - **Context** – the contextual information is very important when synthesizing research. Some results may only be valid in a certain context and hence it must be clearly presented. Petersen and Wohlin presented guidelines for context descriptions [11]. It is particularly important to include information about the study object and subjects.
  - **Research results** – the actual results are important to make clear, since they often form the basis for the data extraction with respect to the findings. The results must be clearly presented and any statistical analysis or other type of data analysis method must be clear to the reader to allow for synthesis.
  - **Validity threats** – validity is an important concern, since it gives some hints about, for example, the generalizability of the results. This means that it helps the researcher performing the systematic literature review to better understand the usefulness of the findings in a synthesis situation.
  - **Conclusions** – these must be clear and summarize the paper in a consistent and comprehensive way.
  - **References** – references form an important source of information when conducting a systematic literature review. From the reference list, it is possible to conduct a backward snowballing [18] to identify more relevant papers. Authors should ensure that references to seminal papers and original sources of information are listed in the reference list. Furthermore, the point of reference should explicitly state why a specific paper is referenced, and not only list papers in the area. The latter is a key point, since it would provide the reader of the paper with important information with respect to the likelihood of the referenced papers also being important or not to include in the systematic literature review.

In summary, the requirements on authors are high to enable other researchers to effectively include the right papers in a systematic literature review. Authors must have synthesis in mind when writing their papers. It would benefit all parties involved, including the authors of the primary studies, the researchers conducting the systematic literature review and the research community as a whole.

4.3 Reviewing

Reviewing is an important activity with respect to research publications. Before submitting a paper for external review, authors ought to review their own paper. This is a good standard practice, but it would be good to not only doing a general read-through, but actually have some specific items to consider. A checklist is recommended. One of the major items in the checklist ought to be to evaluate the writing in the paper from a synthesis point of view.

Once a paper is submitted to a journal or conference, those involved in the evaluation of the submission should have synthesis in mind. Thus, editors, conference program chairs, and reviewers should as part of the evaluation consider whether or not the paper provides sufficient details, use a correct terminology and is presented in such a way that the paper will be useful for other researchers conducting a systematic literature review.

In summary, a large responsibility must be put on those reviewing the primary studies. If researchers are concluding that the quality of the primary studies is insufficient, as reported by Hassler et al. [6], then the research community must raise the level themselves by conducting better reviews. The same researchers most often publish primary studies, review papers and conduct systematic literature reviews.

5. CONCLUSIONS

Anyone having conducted a systematic literature review is well aware that it is a large and challenging undertaking. Several steps in the process are challenging. Perhaps the most challenging step is the synthesis of evidence. It is challenging for several different reason. The actual synthesis as such is difficult, but the key problem making it difficult is the primary studies. Two recent papers [5] and [6] present similar findings, Guzman et al. [5] identified the primary studies and their quality as the main challenge with respect to synthesis, and Hassler et al. [6] report that one of the main barriers for conducting systematic literature studies is the quality of the primary studies.

This paper has tried to address the concern regarding primary studies by pinpointing three areas that need improvement with respect to primary studies to facilitate synthesis of empirical evidence in software engineering. The areas are: terminology, paper content and reviewing. Some experiences from conducting systematic literature reviews have been reported in the form of recommendations to ensure that higher quality primary studies are published in the future.

Finally, it is important to stress that this is a community problem. In many cases, it is the same authors writing primary studies that also conduct systematic literature reviews. Furthermore, the same researchers are also reviewers. Thus, the only way to get higher quality systematic literature reviews is that we, as a community, start thinking "synthesis" when writing research papers that later may be a primary study in someone's systematic literature review.

We must "***write for synthesis***" to be successful in synthesizing empirical evidence in software engineering.

6. REFERENCES

[1] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007.

[2] D. Budgen, B. A. Kitchenham, S. M. Charters, M. Turner, P. Brereton, and S. G. Linkman, "Presenting software engineering results using structured abstracts: A randomised experiment," *Empir. Softw. Eng.*, vol. 13, no. 4, pp. 435–468, 2008.

[3] D. S. Cruzes and T. Dybå, "Research synthesis in software engineering: A tertiary study," in *Information and Software Technology*, 2011, vol. 53, no. 5, pp. 440–455.

[4] F. Q. B. da Silva, A. L. M. Santos, S. Soares, A. C. C. Franca, C. V. F. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Inf. Softw. Technol.*, vol. 53, no. 9, pp. 899–913, 2011.

[5] L. Guzmán, C. Lampasona, C. Seaman, and D. Rombach, "Survey on Research Synthesis in Software Engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 2014, pp. 13–22.

[6] Edgar Hassler, Jeffrey C. Carver, Nicholas A. Kraft, and David Hale, "Outcomes of a Community Workshop to Identify and Rank Barriers to the Systematic Literature Review Process," in *Proceedings of the18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 2014, pp. 267–276.

[7] S. Jalali and C. Wohlin, "Systematic literature studies: database searches vs. backward snowballing," in *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, 2012, pp. 29–38.

[8] S. Jalali and C. Wohlin, "Global software engineering and agile practices: a systematic review," *J. Softw. Evol. Process*, vol. 24, no. 6, pp. 643–659, 2012.

[9] B. A. Kitchenham, T. Dybå, and M. Jørgensen, "Evidence-based software engineering," in *Proceedings of the 26th International Conference on Software Engineering*, 2004, pp. 273–281.

[10] B. A. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[11] K. Petersen and C. Wohlin, "Context in industrial software engineering research," in *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, 2009, pp. 401–404.

[12] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empir. Softw. Eng.*, vol. 14, no. 2, pp. 131–164, 2008.

[13] M. Shaw, "Writing Good Software Engineering Research Papers," in *Proceedings of 25th International Conference on Software Engineering (ICSE'03)*, 2003, pp. 726–736.

[14] D. Smite, C. Wohlin, R. Feldt, and T. Gorschek, "Reporting empirical research in global software engineering: a classification scheme," in *Global Software Engineering, 2008. ICGSE 2008. IEEE International Conference on*, 2008, pp. 173–181.

[15] D. Šmite, C. Wohlin, T. Gorschek, and R. Feldt, "Empirical evidence in global software engineering: a systematic review," *Empir. Softw. Eng.*, vol. 15, no. 1, pp. 91–118, 2010.

[16] D. Šmite, C. Wohlin, Z. Galvina, and R. Prikladnicki, "An empirically based terminology and taxonomy for global software engineering," *Empir. Softw. Eng.*, vol. 19, no. 1, pp. 105–153, 2014.

[17] C. Wohlin, P. Runeson, P. A. da Mota Silveira Neto, E. Engström, I. do Carmo Machado, and E. S. de Almeida, "On the Reliability of Mapping Studies in Software Engineering," *J. Syst. Softw.*, vol. 86, no. 10, pp. 2594–2610, 2013.

[18] C. Wohlin, "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 2014, pp. 321–330.

[19] C. Wohlin and A. Aurum, "Towards a Decision-making Structure for Selecting a Research Design in Empirical Software Engineering," *Empir. Softw. Eng.*, vol. 0, p. 0, 2014. Available from articles in press. DOI= http://dx.doi.org/10.1007/s10664-014-9319-7

[20] R. K. Yin, *Case Study Research: Design and Methods*, vol. 5, no. 5. Sage Publications, 2009, p. 219.