DOI: http://dx.doi.org/10.15446/ing.investig.v35n3.46616

Protocol adaptations to conduct systematic literature reviews in software engineering: A chronological study

Adaptaciones al protocolo para realizar revisiones sistemáticas de literatura en ingeniería de software: Un estudio cronológico

S. Sepúlveda[1], and A. Cravero[2]

ABSTRACT

Systematic literature reviews (SLR) have reached a considerable level of adoption in Software Engineering. However, protocol adaptations for its implementation remain tangentially addressed. This work provides a chronological study for the use and adaptation of the SLR protocol, including its current status. A systematic literature search was performed, reviewing a set of twelve articles published between 2004 and 2013, and selected in accordance with the inclusion and exclusion criteria and using digital data sources recognized by the SE community. A chronological study that includes the current state of the protocol adaptations to conduct SLR in SE is provided. The results indicate areas where the quantity and quality of investigations needs to be increased, and the identification and also the main proposals providing adaptations for the protocol conducting SLR in SE.

Keywords: Systematic literature review, software engineering, chronological study.

RESUMEN

Las Revisiones Sistemáticas de Literatura (RSL) han alcanzado un nivel considerable de adopción en la Ingeniería de Software (IS). Sin embargo, las adaptaciones del protocolo para su aplicación siguen siendo abordadas tangencialmente. Este trabajo proporciona un marco cronológico del uso y adaptación de este protocolo. Se realizó una búsqueda sistemática de literatura, revisando un conjunto de doce artículos, publicados entre los años 2004 y 2013, utilizando fuentes de datos digitales reconocidos por la comunidad de IS. Se proporciona un estudio cronológico que incluye el estado actual de las adaptaciones de protocolos para llevar a cabo una RSL en IS. Los resultados indican áreas en las que la cantidad y la calidad de las investigaciones deben ser aumentadas, y la identificación de las principales propuestas que ofrecen adaptaciones para el protocolo de realización de RSL en la IS.

Palabras clave: Revisión sistemática de literatura, ingeniería de software, estudio cronológico.

Received: October 21st 2014
Accepted: July 26th 2015

Introduction

The importance of research in software engineering (SE) aims to produce knowledge based on the scientific method. This has become one of the main challenges in strengthening the foundations of SE as a discipline on its path to total maturity (Rodriguez 2005). Different types of experimental studies can be used in SE (Wohlin *et al*. 2006), some proposals to support the fulfillment of these studies can be found in Wohlin *et al*. (2000).

Researchers have applied primary studies to improve the knowledge of SE (Basili *et al*. 1999) in order to support the processes related to SE technologies, mainly those related to appraising the technology (Shull *et al*. 2001). Secondary studies are designed to make feasible the comparisons between individual investigations, scientifically selected within a series of primary studies that can support the creation of an evidence-based body of knowledge (Kitchenham *et al*. 2009).

Evidence-based Software Engineering (EBSE) is designed to provide the means to obtain the best current evidence, integrating practical experience and human values into the decision-making for software development and maintenance (Dybå *et al*. 2005). EBSE considers five steps (Sackett *et al*. 1996): (i) convert the need for information into questions and answers, (ii) identify the best evidence to answer these questions, (iii) assess the critical evidence (validity and utility), (iv) put the results of this evaluation into practice in SE and (v) evaluate the yield of this implementation.

The preferred method for steps (ii) and (iii) is the systematic literature review (SLR) (Da-Silva*et al*. 2011). Unlike a peer review, a SLR is a rigorous methodological review; it aims to provide all the existing evidence on a research question and also to support the development of evidence-based directives for practitioners (Kitchenham *et al*. 2007).

[1] Samuel Sepúlveda Cuevas: ScD. Candidate in Informatics Application, Universidad de Alicante, Spain. Affiliation: Universidad de la Frontera, Chile. E-mail: samuel.sepulveda@ceisufro.cl

[2] Ania Cravero Leal: ScD. in Computer Information Systems, Atlantic International University, United States. Affiliation: Universidad de la Frontera, Chile. E-mail: ania.cravero@ufrontera.cl

How to cite: Sepúlveda, S., & Cravero, A. (2015). Protocol adaptations to conduct systematic literature reviews in software engineering: A chronological study. *Ingeniería e Investigación, 35(3)*, 84-91.
DOI: http://dx.doi.org/10.15446/ing.investig.v35n3.46616

It was Kitchenham (2004) who adopted the protocol to implement a SLR from medicine to SE. Later, the protocol was updated using concepts from the social sciences (Kitchenham *et al.* 2007). In addition, SLRs require an extra effort that must be planned prior to execution, and the entire process must be documented (Biolchini et al. 2005). This indicates the need to put the efforts into its planning and execution, so as to guide researchers in carrying it out. Therefore, SLR protocol adaptations to SE must be considered. Additionally, in a study from Kitchenham (2013) regarding using SLRs in SE, she concludes that the three most significant problems are: (i) digital libraries are not well suited to complex automated searches, (ii) the time and effort needed for SLRs and (iii) the quality assessment of papers based on different research methods.

The aim of this work is to account for the adaptations made to the protocol used to conduct SLR in SE, also providing a chronological study that includes its current status. This article may be of interest to researchers planning to conduct additional studies, as well as to practitioners and new researchers who wish to approach SLRs as a relevant source of information in SE.

The structure of the article presents the main steps of the conducted methodology in section 2. In section 3 the selected works are reviewed in detail. Results and discussion are presented in section 4. Finally, the main conclusions of this work are presented in section 5.

Methodology

A systematic literature search was conducted, compiling background on change proposals for protocol to conduct SLRs in SE. We speak of a systematic search and not a SLR as defined by Kitchenham (2007), because we did not strictly follow all the steps defined in the protocol for its implementation (i.e. we did not do a quality assessment or a classification of works).

Research Questions (RQs): The RQs to be answered for this work are presented in Table 1.

Table 1. RQs to be answered.

RQ	Aim
RQ1: What changes are proposed to the original protocol defined by B. Kitchenham?	We want to give an account of the changes made to the original SLR protocol in SE. It would be possible to observe how the community has adapted this protocol in practice.
RQ2: What sections of the SLR protocol have been more modified, and what did these modifications consider?	We want to know the protocol sections that have been modified and what kinds of changes have been incorporated. This would show the protocol sections most controversial for the SE community and their adaptions.
RQ3: What has happened with proposals for changes to the SLR protocol through time?	We want to know when these proposals have originated. This would show the temporal evolution of the SLR protocol.

Searching for works: To answer the RQs, the systematic search was based on identifying adaptations made to the protocol to conduct SLRs in SE. This search covers the period between 2004 –date on which Kitchenham (2004) adapts the protocol used in medicine– and 2013.

We use some of the sources most frequently used by the SE community (Brereton et al. 2007). In our case we consulted: IEEE Xplore, ACM Digital Library and Science Direct. The used search string was: ("systematic literature review" OR "systematic review") AND ("software engineering") AND ("guidelines" OR "protocols" OR "lessons" OR "study" OR "proposals").

Selecting works: Once the data sources and search string were defined, all those works that reported changes to the protocol on conducting SLR were reviewed. This included reading the methodology used, the steps carried out and the results obtained.

Inclusion and exclusion criteria: The following criteria allow us to determine the relevance of the works collected.

a. *Inclusion criteria:* all the works regarding SLR in SE that specifically mention aspects about modifications of the protocol to carry out a SLR, i.e. how to conduct a SLR and the stages/activities this entails.

b. *Exclusion criteria:* all the works containing SLR topics, but that do not suggest proposals on how to carry out or modify the defined protocol to develop SLRs in SE.

Initially 31 works were compiled, and 12 were finally selected. These are the ones analyzed and described in detail in the next section.

Considering that inclusion and exclusion of the works was done by reviewing and interpreting the text (which is potentially ambiguous), the reliability between the reviewers was calculated using *Cohen's Kappa statistic* (Gwet 2002). The results, after two failed attempts, were satisfactory (K = 0.851). This result indicates that the scale presented in Clark et al. (2004) provides a basis for criteria that is clear enough, which does not induce significant divergences among measurers. In addition, for those cases where the reviewers had doubts about including or not a work, this was subjected to an individual review and then a decision was made by group consensus.

Data extraction and synthesis: Regarding the works that submit a proposal for changes to the SLR protocol in SE, in a previous work we established which would be the activities in the protocol to look for changes in each one of the protocol stages (Sepúlveda and Cravero 2013).

Proposals for changes to the SLR protocol

This section analyzes the selected works. Table 2 shows the three stages of the protocol, their activities and an identifier for each one of them. We chose these activities by reviewing literature evidence that shows changes to the SLR protocol.

Selected Works

Having selected the works, 12 were found to present proposals for modifications to the protocol to conduct SLRs in SE. Table 3 shows the #Id, title, authors and year of each selected work.

Table 2. Stages and activities reviewed.

Planning	A1: Definition of the Research Questions (RQ)
Implementation	A2: Identification of the relevant works
	A3: Selection of the relevant works
	A4: Quality evaluation of selected works
	A5: Data extraction
	A6: Data synthesis
Documentation	A7: Report of the results

Table 3. Selected works summary.

#Id.	Title	Authors	Year
T1	Procedures for performing systematic reviews	Kitchenham, B.	2004
T2	Análisis y revisión de la literatura en el contexto de proyectos de fin de carrera: Una propuesta	Caro, M.A., Ríos, A.R. et al.	2005
T3	A systematic review process to software engineering	Mian, P., Conte, T. et al.	2005
T4	Systematic Mapping Studies in Software Engineering	Petersen, K., Feldt, R. et al.	2007
T5	Guidelines for performing Systematic Literature Reviews in Software Engineering	Kitchenham, B., Charters, S. et al.	2007
T6	Proposal of a Review Process of Empirical Studies in Software Engineering	Grimán, A. And Juristo, N.	2007
T7	Lessons from applying the systematic literature review process within the software engineering domain	Brereton, P., Kitchenham, B. et al.	2007
T8	Applying systematic reviews to diverse study types: An experience report	Dybå, T., Dingsøyr, T. et al.	2007
T9	Experiences using systematic review guidelines	Staples, M. and Niazi, M.	2007
T10	Refining the systematic literature review process-two participant-observer case studies	Kitchenham, B., Brereton, P. et al.	2010
T11	Identifying relevant studies in software engineering	Zhang, H., Babar, M.A. et al.	2011
T12	Using the 5W+1H Model in Reporting Systematic Literature Review: A Case Study on Software Testing for Cloud Computing	Changjiang, J. and Yuen-Tak, Y.	2013

Stages of a SLR-SE analysis

In order to conduct a detailed analysis of each stage and activities according to the protocol, a table was designed for each activity reviewed. Due to extension issues, an example of these tables is shown on Table 4. All the Tables can be seen in Sepúlveda and Cravero (2013).

Table 4. Proposals for the "Report of the results (A7)" stage.

Changes - Proposals	Code	Id
Proposes formats to publish the results (structure and contents of a report).	KitA7	T1
Results of the final protocol must be reported, which includes reviews/changes regarding the process and explaining the nature of changes to the original protocol.	StaA7	T9
Make a detailed record of decisions taken during the process. To establish a mechanism that allows SLR results to be published (more extensive than traditional papers) or use of appendices in electronic repositories.	BreA7	T7

a. Planning stage and activities reviewed

For the planning stage we review the activity "Definition of the RQ (A1)". Seven works were selected for this stage and activity A1 (#Id. T1, T3, T5, T6, T7, T9, T12). The proposals suggest: (i) guidelines to help define best RQs, (ii) guidelines to check that defined RQs are indeed the most appropriate and (iii) that RQs are not defined *a priori*, but rather defined as a greater knowledge of the subject being gained.

b. Implementation stage and activities reviewed

For implementation stage we review the activities A2 to A6, according to Table 1.

- Identification of relevant works (A2): Six works were selected for A2 (#Id. T1, T2, T5, T6, T7, T11). The proposals suggest: (i) identification and selection of relevant data sources, (ii) definition and justification of a systematic search strategy according to the defined RQs and (iii) identification of categories for classification of the works identified (Nabi and Mullins 2011).
- Selection of relevant works (A3): Eight works were selected for A3 (#Id. T1, T2, T3, T5, T6, T7, T9, T10). The proposals suggest: (i) definition of guidelines to establish the inclusion/exclusion criteria, (ii) guidelines to resolve disagreements between reviewers when selecting works, (iii) use of peer review to avoid bias when selecting a work and (iv) review of other elements of the paper such as the conclusions, because abstracts are usually of low quality.
- Quality evaluation of selected works (A4): Six works were selected for A4 (#Id. T1, T5, T6, T8, T9, T10). The proposals suggest: (i) guidelines and framework to evaluate the quality of the selected work, (ii) use of checklists with defined factors to evaluate the quality of the work and (iii) participation of multiple evaluators and discussion rounds to reach a consensus on criteria.
- Data extraction (A5): Eight works were selected for A5 (#Id. T1, T2, T3, T4, T5, T6, T7, T9). The proposals suggest: (i) design and use of forms to record data, (ii) use of software tools to support the documentation of data, (iii) use of peer review and (iv) recording the section of the article where the selected data is found.
- Data synthesis (A6): Seven works were selected for A6 (#Id. T1, T2, T3, T4, T6, T7, T9). The proposals suggest: (i) guidelines for synthesizing data, (ii) summary with statistical results from quantitative and qualitative data, and (iii) use of tables and databases to facilitate data queries and analysis.

c. Documentation stage and reviewed activities

For the documentation stage, reviewing the activity report of the results was considered (A7).

- Report of the results (A7): Three works were selected for A7 (#Id. T1, T7, T9). The proposals suggest: (i) formats and guidelines to publish results and (ii) the reviews and decisions made during the process must be reported.

d. Comments to the stages and activities reviewed

Having reviewed the three stages and the identified proposals for each one, we can say that they focus essentially on

defining guidelines for: (i) supporting the definition of RQs, (ii) defining inclusion/exclusion criteria, (iii) synthesizing data, and (iv) publishing the results. The details are presented in Sepúlveda and Cravero (2013).

Results and Discussion

Next, the results and findings are discussed. The main threats to the validity of this study are also presented.

The final selection included 12 works between 2004 and 2013. We think that the specificity of the topic has caused the sample to be rather small, and due to this same specificity, the review provides a reliable overall view of the state of research in this area.

Answering the RQs

RQ1: What changes are proposed to the original protocol defined by B. Kitchenham?

The original protocol for conducting SLRs in SE was defined by #Id T1. Later works were published proposing changes to it, in one or more activities for the three stages.

Generally, we can say that the proposals for changes to the SLR protocol in SE focus essentially on defining guidelines for: (i) supporting the definition of the RQs; (ii) identifying and selecting relevant data sources as well as the definition of a search strategy aligned with the RQs and classification of the identified works by category; (iii) defining the inclusion/exclusion criteria, the solution of disagreements between reviewers when selecting works and the caution in using only abstracts due to their low quality; (iv) evaluating the quality of the selected works and participation of several evaluators and how to reach a consensus on the criteria; (v) synthesizing the data, obtaining statistical results from quantitative and qualitative data and using tables and databases to facilitate the analysis; and finally (vi) publishing the results, reporting the reviews and decisions taken in the process.

RQ2: What sections of the protocol have been more modified and what did these modifications consider?

From the point of view of the stages, the greatest number of proposals is for the implementation stage, which concentrates 37 proposals (80%). With respect to the activities, three were identified with the greatest number of proposals: identification of relevant works (A2), selection of relevant works (A3) and data extraction (A5) with 8 proposals for each one of them (17% in each case). The documentation stage presents only 3 proposals (7%). Finally, it is worth noting that some works not only present changes to the protocol, but also define different stages being executed in a different order compared to the other proposals. An example is #Id. T6. Table 5 summarizes the proposed changes to the protocol for each activity identified.

Table 5. Changes to the protocol, activities and proposals.

#Id	Planning	Implementation					Documentation
	A1	A2	A3	A4	A5	A6	A7
T1	√	√	√	√	√	√	√
T2		√	√		√	√	
T3	√	√	√		√	√	
T4		√			√	√	
T5	√	√	√	√	√		
T6	√	√	√	√	√	√	
T7	√	√	√		√	√	√
T8				√			
T9	√		√	√	√	√	√
T10			√	√			
T11		√					
T12	√						

RQ3: What happened with the proposed changes to the protocol through time?

From the stages and activities identified, as well as from the changes proposed for each one of these activities, a timeline has been prepared for each stage of the protocol (planning, implementation and documentation). We used the previously defined acronyms for each work reviewed.

Planning stage and proposed changes: As shown in Figure 1, the proposals defined for activity A1 are included between 2004 and 2013, totaling seven proposals; three of them are from 2007.

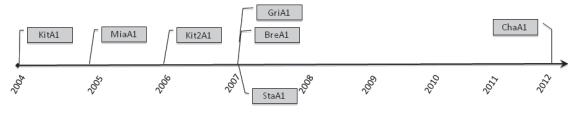

Figure 1. Proposals for changes to the SLR protocol for the planning stage.

Implementation stage and proposed changes: As shown in Figure 2, there are a large number of proposals defined for activities A2 to A6 included between 2004 and 2011, totaling thirty-seven proposals. Twenty of them are from 2007. In order to see what happens to each activity in greater detail, what follows is a breakdown of the analysis for each.

The proposals defined for activity A2 are included from 2004 to 2007 and 2011, totaling eight proposals, four of which are from 2007. The proposals defined for activity A3 are included from 2004 to 2007 and 2010, totaling eight proposals, four of which are from 2007. The proposals defined for activity A4 are included from 2004, 2006-2008 and 2010, totaling six proposals, two of which are from 2007. The proposals defined for activity A5 are included from 2004 to 2007, totaling eight proposals, five of which are from 2007. Finally the proposals defined for activity A6 are included from 2004 to 2005 and 2007, totaling seven proposals, five of which are from 2007.

Documentation stage and proposed changes: As shown in Figure 3, the proposals defined for activities A7 include the years 2004 and 2007, totaling three proposals, two of them are from 2007.

We can see from Figures 1-3: (i) as we expected, at the beginning everything was based on the proposal by (Kitchenham 2004); (ii) the greatest number of changes are concentrated in 2007, with a total of 25 proposals; (iii) the activities with the most proposed changes are A2, A3 and A5, which correspond to the implementation stage with a total of eight proposals each; (iv) the stage that seems to be the most stable is documentation/report, because between 2004 and 2011 only 3 change proposals are recorded; and (v) according to the changes after 2008, it could be argued that these are more focused on improving and controlling quality aspects of the SLRs. Figure 4 shows a quantification of the proposals for changes to the SLR protocol with respect to the year in which these were published, where it can be corroborated that the greatest number of proposals appears in 2007.

From the collected evidence, we can say that the greatest number of proposals regarding the original SLR protocol in SE appeared in 2007. This is consistent with a considerable increase in the number of SLRs published in the same year, which show a growth rate that is maintained up to the present day, but the changes proposed to the SLR protocol decay dramatically.

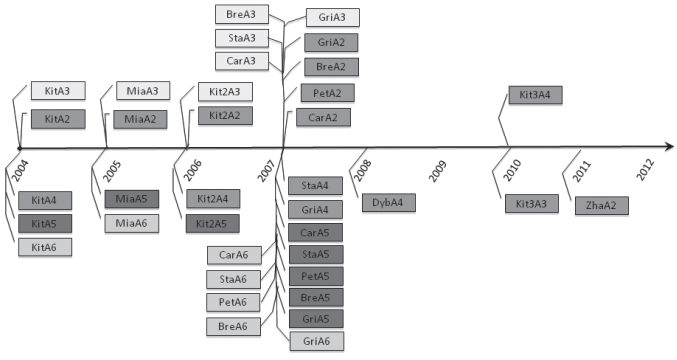

Figure 2. Proposals for changes to the SLR protocol for the implementation stage.

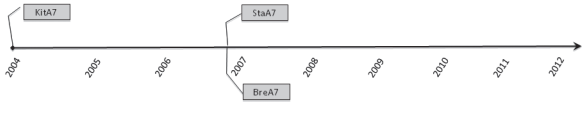

Figure 3. Proposals for changes to the SLR protocol for the documentation/report stage.

The twelve works selected present 46 proposals. Twenty-five proposals were published in 2007, which means 54% of the proposals are concentrated in this year.

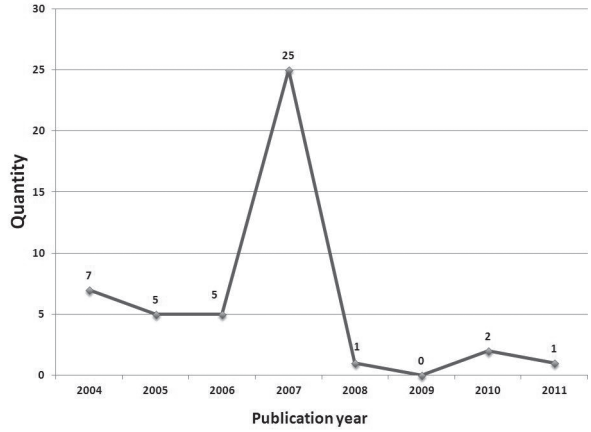

Figure 4. Annual number of proposals for changes to SLR protocol.

Meaning of the findings and results

From the collected data, we can state that SLR is a subject that has gained relevance in the SE community, which translates into an increasing number of articles in specialty journals and conferences, as well as an increasing number of experiences of application/adoption in the industry. Nevertheless, we detected some relevant aspects where there is a considerable lack of both theoretical and empirical contributions, and some areas where it is possible to make contributions to the community, such as the implementation of: (i) tertiary studies that allow the real state of the quality of SLRs conducted in SE to be visualized; (ii) studies that make it possible to verify whether there is indeed a stabilization of the protocol for conducting SLRs; (iii) use of empirical evidence to establish how this protocol is used and adapted; and (iv) studies to establish the level of adoption and adaptation of SLRs in the industry.

The collected data shows the significant increase in protocol proposals in 2007-2008, but this number has fallen drastically. This makes us think that the protocol to implement SLRs in SE has generally attained a certain acceptance and stability within the SE community. Then, the emphasis of the community is migrating toward improving the quality of primary studies. We do not have the arguments and it is beyond the scope of this work to verify whether these hypotheses are true or false. It can give rise to a new type of research for SLR and SE according to the quality of primary works and the need to establish more tertiary studies that are dedicated to reviewing the quality of secondary studies. An example of this are Kitchenham et al. (2012) and Zhang and Ali-Babar (2013), which report on issues, activities, working groups characterization and quality of SLRs and primary works collected.

In addition, if we observe the authors and co-authors of each one of the twelve selected works, in 50% of these a subset of six researchers is involved. Therefore, we can say that there is a group concerned with improving the processes and performance of SLRs in SE. The case of B. Kitchenham

stands out; besides having adapted the protocol to develop SLR in SE, she is present in four of the twelve works, and in three of them as the main author. We can say that she and her research group are leading the work in terms of SLR research in SE.

Finally, besides the changes to the protocol that we evidenced, some authors make a set of recommendations to improve the SLRs in SE. Next, we present a summary of these recommendations, identifying: item, authors and proposals of each one of them.

Abstract: The low quality and how the abstracts are considered as a key element in selecting works (Staples and Niazi, 2007). Also, it is recommended using the structured abstract and suggest it as an important source of information and to emphasize the abstract as the only section of the publication that is accessible free of charge (Jedlitschka and Pfahl, 2005). For more recommendations on using structured abstracts see (Budgen *et al.*, 2008).

Search: Searching for relevant works using digital sources in SE community makes it necessary to use different search strings, try them out and evaluate the results (Chen et al. 2009, Kitchenham *et al.* 2007.). Search engines do not support the use of search strings to conduct SLRs (Staples and Niazi 2007). Using a glossary of terms from the experience based medicine can be helpful for those initiating SLRs (Kitchenham *et al.* 2010). To have a unified source, a centralized SLR index in SE similar to the Cochrane Collaboration[3] initiative is recommended (Staples and Niazi, 2007).

Quality: According to (Cruzes and Dybå, 2011), the quality of SLRs conducted can be positively influenced if the challenges at the time of synthesizing the research around SE are better understood. In addition, despite the focus being placed on SLRs, limited attention is given to this item. It requires becoming a central aspect of the SLR so as to increase its importance and utility both in the research and practice of the discipline. A simplification of the original criterion raised by Kitchenham to evaluate the quality of each work is suggested by Staples and Niazi (2007). In the future, instruments should be developed to support the implementation and control of a SLR, similar to the PRISMA[4] proposal (Moher *et al.* 2010).

Protocol and stages: Considering the original protocol for SLR, Staples and Niazi (2007) talk about the lack of clarity in directives for synthesizing data. This, despite the fact that they agree with the importance of running a pilot project, and criticize Kitchenham for not clarifying when to stop the pilot or when a pilot project must be run. Improvements regarding how to conduct a SLR and a set of learning strategies are collected by Brereton *et al.* (2007).

Templates: Recommendations about using templates to conduct SLRs and to define an ontology describing the knowledge of experimental studies are suggested (Biolchini *et al.* 2006). An application of this template can be seen in Biolchini *et al.* (2005). About using guidelines to report results in EBSE, including SLR, see Jedlitschka and Pfahl (2005).

Tool support: To conduct SLRs requires considerable effort and these are time consuming. On the other hand, many stages and tasks are carried out manually, which means having tools to support this process is very important. In recent years there have been various proposals with software tools supporting different tasks for conducting SLRs (Bowes *et al.* 2012, Felizardo *et al.* 2011).

Threats to validity

We are aware there are some threats that may affect validity of the findings and results. Among these are:

(i) Possible bias in selecting works. We use data sources that are highly recognizable within the SE community (IEEE Xplore, ACM Digital Library, Science Direct). We do not consider other relevant sources, basically due to aspects of scope and time. While a total of twelve works found seems to be very low, in the future we hope to validate the results obtained by expanding the sources considered and improving the RQs.

(ii) Limitations of the search engines used to conduct the searches in electronic data sources (Dybå 2008). We tried to mitigate these threats by means of an individual selection and a joint validation of the works, thus avoiding individual bias. In order to avoid works being left out of the study, the idea was to review all the versions of a work, whether these were journals, conference proceedings or technical reports.

(iii) Limitations of the search string. The used search string was not validated by domain experts and neither was a criterion used to build this string, such as PICOC (Petticrew and Roberts 2008). This weakness undermines the generalization factor of this study and must be considered in future works.

Conclusions

The work presented covers the protocol adaptations of the SLR as a research methodology in SE. We provide a chronological study that includes its current status. In addition, the answers and evidence for the RQ have been reviewed. The collected evidence may be of interest to practitioners and new researchers who wish to approach the SLR as a relevant source of information, as well as to researchers planning to conduct additional studies on SLR and SE.

Although there are other works that present both a set of observations and criticisms made in the SE SLR, as far as we know there is no evidence of works that specifically report results of the protocol adaptations to conduct SLR as a methodology applied research in SE. This work can therefore be seen as a complement to those that are reviewing the evolution of SLR in SE. We understand that more tertiary studies are required in this area for it to delve into greater detail.

[3] http://www.cochrane.org/cochrane-reviews

[4] http://www.prisma-statement.org/index.htm

As future work, we plan to add and refine the RQs and data sources in order to test the robustness of the ideas put forward here. Also, a quality evaluation of collected data must be done to test the strength of evidence.

Acknowledgements

This work was supported by the Research and Postgraduate Office at Universidad de La Frontera, Research Project DIUFRO DI14-0065.

References

Basili, V. & Shull, F., Lanubile, F. (1999). Building knowledge through families of experiments. IEEE Transactions on Software Engineering, 25(4), 456-473. DOI: 10.1109/32.799939

Biolchini, J., Mian, P.G., Candida, A. &Natali, C. (2005). Systematic Review in Software Engineering. System Engineering and Computer Science Department COPPE/ UFRJ, Technical Report ES, 679(05).

Biolchini, J. C., P. G. Mian, Gomes, P, Candida, A, Uchóa & T. Horta, G. (2006). Scientific research ontology to support systematic review in software engineering. Advanced Engineering Informatics. 21(2), 133-151. DOI: 10.1016/j.aei.2006.11.006

Bowes, D., Hall, T., & Beecham, S. (2012, September). SLuRp a tool to help large complex systematic literature reviews deliver valid and rigorous results. Paper presented at Proceedings of the 2nd international workshop on evidential assessment of software technologies. 33-36. ACM. DOI: 10.1145/2372233.2372243

Brereton, P., Kitchenham, B.A., Budgen, D., Turner & M.,Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. Journal of Systems and Software, Elsevier, 80(4), 571-583. DOI: 10.1016/j.jss.2006.07.009

Budgen, D. & Kitchenham, B. (2008) Presenting software engineering results using structured abstracts: a randomised experiment. Empirical Software Engineering. 13(4), 433-458. DOI: 10.1007/s10664-008-9075-7

Chen, L., M. Ali-Babar & Nour, A. (2009). Variability Management in Software Product Lines: A Systematic Review. Paper presented at 13th International Software Product Line Conference, Carnegie Mellon University.

Clark, T., Sammut, P. & Willans, J. (2004). Applied metamodelling: a foundation for language driven development. Ceteva.

Cruzes, D. S. & T. Dybå. (2011). Research synthesis in software engineering: A tertiary study. Information and Software Technology. 53(5), 440-455. DOI: 10.1016/j.infsof.2011.01.004

Da-Silva, F., Santos, A., Soares, S., Monteiro, C. & Farias, F. (2011). Six years of systematic literature reviews in software engineering: An updated tertiary study. Information and Software Technology, 53(9), 899--913. DOI: 10.1016/j. infsof.2011.04.004

Dybå, T. (2008). Strength of Evidence in Systematic Reviews in Software Engineering. Paper presented at Second International Symposium on Empirical Software Engineering and Measurement. ESEM, 7465, 178–187. DOI: 10.1145/1414004.1414034

Dybå, T., Kitchenham, B.A. & Jorgensen, M., Evidence-based software engineering for practitioners. Software, IEEE, IEEE, 22(1), 58-65.

Felizardo, K. R., Salleh, N., Martins, R. M., Mendes, E., MacDonell, S. G., & Maldonado, J. C. (2011, September). Using visual text mining to support the study selection activity in systematic literature reviews. Paper presented at Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on. 77-86. IEEE. DOI: 10.1109/esem.2011.16

Gwet, K. (2002). Interrater reliability: dependency on trait prevalence and marginal homogeneity. Statistical methods for interrater reliability assessment, 2, 1-9.

Jedlitschka, A. & Pfahl, D. (2005). Reporting Guidelines for Controlled Experiments in Software Engineering. Paper presented at International Symposium on Empirical Software Engineering, IEEE.

Kitchenham, B. (2004). Procedures for performing systematic reviews, Department of Computer Science, Keele University. DOI: 10.1016/j.infsof.2008.09.009

Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey & J.,Linkman, S.G. (2009). Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology, 51(1), 7-15.

Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Jørgensen, M., Mendes & E.,Visaggio, G. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering, School of Computer Science and Mathematics Keele University and Department of Computer Science University of Durham.

Kitchenham, B. & Pretorius, R. (2010). Systematic literature reviews in software engineering – A tertiary study. Information and Software Technology. 52(8), 792-805. DOI: 10.1016/j.infsof.2010.03.006

Kitchenham, B., Sjøberg, D., Dybå, T., Pfahl, D., Brereton, P., Budgen, D. & Höst, M.,Runeson, P. (2012). Three empirical studies on the agreement of reviewers about the quality of software engineering experiments. Information and Software Technology, 54(8), 804-819. DOI: /10.1016/j.infsof.2011.11.008

Kitchenham, B. & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. Information and Software Technology, 55(12), 2049–2075. DOI: 10.1016/j.infsof.2013.07.010

Marshall, C., & Brereton, P. (2013, October). Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study. Paper presented at Empirical Software Engineering and Measurement, ACM/IEEE International Symposium on. 296-299. IEEE. DOI: 10.1109/esem.2013.32

Moher, D. & Liberati, A. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. The PRISMA Group International Journal of Surgery. 8(5), 336–341. DOI: 10.1016/j.ijsu.2010.02.007

Nabi, F. & Mullins, R. (2011). Moving from Traditional Software Engineering to Componentware. Journal of Software Engineering and Applications, 4, 283-292. DOI: 10.4236/jsea.2011.45031

Petticrew, R. (2008). Systematic reviews in the social sciences. A practical guide: John Wiley \& Sons.

Rodriguez, D. (2005). Empirical software engineering research: epistemological and ontological foundations, Paper presented at First Workshop on Ontology, Conceptualizations and Epistemology for Software and Systems Engineering (ONTOSE).

Sackett, D.L., Rosenberg, W., Gray, J., Haynes, R.B. & Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. British Medical Journal (BMJ), 312(7023), 71-72. DOI: 10.1136/bmj.312.7023.71

Sepúlveda, S. & Cravero, A. (2013). Protocol to conduct Systematic Literature Reviews in Software Engineering: a chronological point of view of the changes made. Dpto. Ciencias de Computación e Informática (DCI), Centro de Estudios en Ingeniería de Software (CEIS), Technical Report TR-DCI-01-13.

Shull, F., Carver, J. & Travassos, G. (2001). An empirical methodology for introducing software processes, Paper presented at 8th European Software Engineering Conference (ESEC) and 9th ACM SIGSOFT Foundations of Software Engineering (FSE-9), Vienna, Austria, 288-296.

Staples, M. & M. Niazi. (2007). Experiences using systematic review guidelines. Journal of Systems and Software. 80(9), 1425-1437. DOI: 10.1016/j.jss.2006.09.046

Wohlin, C., Höst, M. & Henningsson, K. (2006). Empirical re-search methods in Web and software Engineering. Web Engineering, Springer, 409-429.

Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell & B.,Wesslén, A. (2000). Experimentation in software engineering: an introduction. Kluwer Academic Publisher.

Zhang, H. & Ali-Babar, M. (2013). Systematic reviews in software engineering: An empirical investigation. Information and Software Technology, 55(7), 1341-1354. DOI: 10.1016/j.infsof.2012.09.008