

Defining protocols of Systematic Literature Reviews in Software Engineering: a survey

Katia Romero Felizardo
Department of Computer
Federal University of Technology
Cornélio Procópio/PR – Brazil
Email: katiascannavino@utfpr.edu.br

Érica Ferreira de Souza
Department of Computer
Federal University of Technology
Cornélio Procópio/PR – Brazil
Email: ericasouza@utfpr.edu.br

Ricardo Almeida Falbo
Depart. of Computer Science
Federal University of Espírito Santo
Vitória/ES – Brazil
Email: falbo@inf.ufes.br

Nandamudi Lankalapalli Vijaykumar
Lab. of Comp. and Applied Mathematics
National Institute for Space Research
São José dos Campos/SP – Brazil
Email: vijay.nl@inpe.br

Emilia Mendes
Blekinge Institute of Technology
University of Oulu
Depart. of Software Engineering
Karlskrona – Sweden
Email: emilia.mendes@bth.se

Elisa Yumi Nakagawa
Department of Computer Systems
University of São Paulo, USP
São Carlos/SP – Brazil
Email: elisa@icmc.usp.br

Abstract—Context: Despite being defined during the first phase of the Systematic Literature Review (SLR) process, the protocol is usually refined when other phases are performed. Several researchers have reported their experiences in applying SLRs in Software Engineering (SE) however, there is still a lack of studies discussing the iterative nature of the protocol definition, especially how it should be perceived by researchers conducting SLRs. **Objective:** The main goal of this study is to perform a survey aiming to identify: (i) the perception of SE researchers related to protocol definition; (ii) the activities of the review process that typically lead to protocol refinements; and (iii) which protocol items are refined in those activities. **Method:** A survey was performed with 53 SE researchers. **Results:** Our results show that: (i) *protocol definition* and *pilot test* are the two activities that most lead to further protocol refinements; (ii) data extraction form is the most modified item. Besides that, this study confirmed the iterative nature of the protocol definition. **Conclusions:** An iterative pilot test can facilitate refinements in the protocol.

Keywords-Systematic Literature Review; Systematic Mapping Study; Secondary Study; Protocol Definition; Survey.

I. INTRODUCTION

Protocols are essential to ensure high-quality Systematic Literature Reviews (SLRs). However, defining a protocol is a complex task. It comprises many aspects, such as: (i) formulating the Research Questions (RQs); (ii) defining a search strategy and the sources to find primary studies; (iii) defining inclusion and exclusion criteria to be used to select studies; (iv) defining the process to be used to extract, synthesize, describe, and categorize the selected studies; and (v) organizing the workload among researchers to extract data, do quality assessment, etc.

Several studies have reported experiences and lessons learned from researchers conducting SLRs [1], [2], [3].

However, these studies do not focus on the iterative nature involved in defining the review protocol.

Protocols are typically documents that are iteratively built, and they are essential to reduce research bias. For example, the use of an inclusion criterion avoids the selection of studies due to any individual preferences. Therefore, elaborating and refining protocols is an important step in increasing the review process quality. Kitchenham and Charters [4] recognize that the review process is iterative and typically requires revisions. In particular, conducting subsequent phases can result in protocol refinements. As pointed out by Staples and Niazi [5], refinements to the protocol are inevitable. For example, the construction of the search string is not a linear process; indeed, it is a time-consuming and error-prone activity [6], and it implies a continuous refinement process. Another example is that new candidate keywords typically emerge during the *selection* activity, and it is essential to refine the search string considering new terms and their synonyms.

The main contribution of this paper is to clearly identify activities that drive most of the protocol changes during conduction of SLRs, providing researchers with results that can better inform the planning and execution of SLRs in SE. This paper does provide an insight to the actual incremental process of conducting an SLR, that may serve as a practical help for researchers conducting their reviews, specially for novice researchers, conducting the first SLR. Some of the questions we aim to answer in this paper are: (i) Which protocol refinements in SLRs have been conducted by researchers with different experience? (ii) What are the activities of the SLR process that typically lead to protocol refinements? and (iii) What are the protocol parts that are refined as a result of those activities?

The remainder of the paper is organized as follows. Section II briefly presents the background related to the iterative nature of SLRs. Section III presents the survey and its results, focusing on the protocol items that are refined while conducting SLRs. Limitations of this work are also discussed. Finally, Section IV presents our final considerations.

II. BACKGROUND

Although there are publications defining how to perform SLRs, such as [1], [4], there is still a need for studies addressing different aspects of the adoption of SLRs in SE [7]. In particular, the highly iterative nature of the review process is an aspect that has drawn attention to different researchers in recent years. Several authors describe an iterative way to conduct the review process, since many activities are initiated during the planning phase, but are refined later. For example, Zhang and Muhammad [8] mentioned that both study selection and data extraction are done in an iterative way, especially when novices are involved in the process. Hassler et al. [9] mentioned that the SLR process is more iterative than the original definition indicates, with interactions among all phases of the SLR process. Staples and Niazi [5], Mian et al. [10], Biolchini et al. [11], and Riaz et al. [12] claimed the same experiences. In general, they concluded that the protocol goes through many changes.

It is clear from the above mentioned studies that SLR protocols undergo changes as the SLR process takes place. However, what and when is not clear. Moreover, we believe it would be very helpful to know the relationship between SLRs process activities and protocol changes. In other words, which activities drive most of the changes in SLR protocols?

III. SURVEY ON PROTOCOL REFINEMENTS DURING SECONDARY STUDIES

We followed the six phases proposed by Kitchenham and Pflieger [13] to conduct surveys:

– **Phase 1:** Setting the objectives – The questions that this survey intended to answer were set and previously presented in Section I.

– **Phase 2:** Designing the survey – As occurs with most of the surveys in SE [14], our survey is also a cross-sectional study. Participants were asked about their past experiences on defining protocols for SLRs.

– **Phase 3:** Developing the survey instrument (i.e., the questionnaire) – The survey comprises a questionnaire with nine questions. The first three questions refer to the participants' profile. The remaining questions refer to the extent to which the protocol items are refined during the activities of the review process. The remaining questions are: (i) Considering the item “*motivation for conducting the review*”, what is its refinement degree in the subsequent activities of the review process? (ii) Considering the item “*RQs*”, what

is its refinement degree in the subsequent activities of the review process? (iii) Considering the item “*search strategy*”, what is its refinement degree in the subsequent activities of the review process? (iv) Considering the item “*strategy for identifying the primary studies*”, what is its refinement degree in the subsequent activities of the review process? (v) Considering the item “*quality assessment*”, what is its refinement degree in the subsequent activities of the review process? and (vi) Considering the item “*data extraction form*”, what is its refinement degree in the subsequent activities of the review process?

The possible choices to these questions vary from “no refinement” to “very high degree of refinement”, in a scale based on the Likert Scale method, which is a metric used in questionnaires such as attitude surveys. The final median score represents overall level of accomplishment or attitude toward the subject matter. Moreover, in each question, there was a space for the respondents to make free comments.

– **Phase 4:** Evaluating the survey instrument – We conducted a pre-testing, when we applied the survey to a smaller sample, intending to identify any problems with the questionnaire, as well as with the response choices. The questionnaire was also evaluated by three experienced researchers in SLRs. They all had participated in at least five SLRs. After this evaluation, questions were improved to make them clearer.

– **Phase 5:** Obtaining valid data – Researchers with experience in SLRs were invited to answer the survey in the following ways: (i) a request for participation was sent to the list of e-mails of members of the Brazilian Computer Society (SBC¹), a scientific and educational organization dedicated to the advancement of Computer Science in Brazil; (ii) a request was sent to LinkedIn² focusing researchers, as well as PhD and Master's students in SLRs; and (iii) a direct request by e-mail was sent to researchers known by the authors. At the end, 53 SE researchers answered our survey³.

– **Phase 6:** Analyzing the data – The answers were stored directly after they had been submitted by the participants. Moreover, after a thorough analysis, we partitioned the participants into three sub-groups to better analyze whether the level of experience changes the way that researchers perceive the refinements. The groups were: Group 1 (G1): participants that had conducted one and two SLRs; Group 2 (G2) – participants that had conducted three and four SLRs; and Group 3 (G3) – participants that had conducted five or more reviews. G1 was composed of 28 participants, G2 by 13, and G3 by 12. This partitioning was useful for evaluating the responses by level of participants' experience

¹<http://www.sbc.org.br/en/> – SBC has approximately 4.000 members, including university professors, students, and professionals in Computer Science. It has also about 230 universities, research centers, companies, and business associations as members or institutional subscribers.

²<https://goo.gl/6oatqj>

³<https://goo.gl/HfesVl>

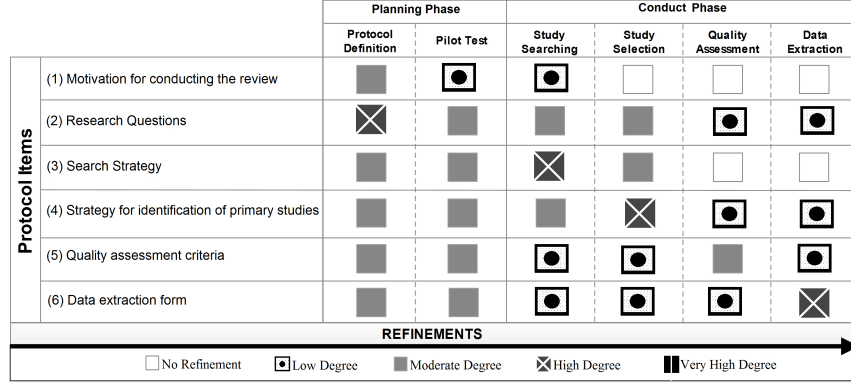


Figure 1. Refinements of protocol items in SLRs.

in conducting SLRs. A total of 19 participants (35.8%) wrote free comments and all these comments also were analyzed.

A. Survey data analysis

As mentioned before, 53 participants completed the questionnaire, 36 were researchers and 17 were graduate students – Master or PhD degrees. A total of 174 SLRs were conducted by the 53 participants. A total of 52.8% of the participants took part in one or two studies, 24.5% took part in three or four studies, and 22.6% took part in five or more studies. Figure 1 shows a survey results overview. In this figure, the five different format of the squares represent the degree of refinement to which a protocol item is subjected as a result of executing the corresponding review process activity. The refinement degree scale varies from “no refinement” (represented by a white square) to “very high degree of refinement” (represented by a Black square with a dash in the middle).

As expected, the survey outcomes reveal that most of the changes occur during the *protocol definition* and *pilot test* activities. The pilot test verifies the feasibility of implementing the SLR, and any further refinements that are needed. Our finding is consistent with that of Imtiaz et al. [3] and Brereton et al. [15] that agree that the most difficult phase of the review process is the planning. Most of protocol items are moderately refined during the planning phase, except *RQs* and *motivation for conducting the review* items. *RQs* are the most refined (high degree of refinement) protocol item during this phase. *Motivation for conducting the review*, in turn, has a low degree of refinement during *pilot test*.

During the conduct phase, refinements continue to occur, but to a lesser extent. The *strategy for identification of primary studies* is the most refined (high degree of refinement) protocol item during the *study selection* activity. Similarly, *data extraction form* is the most refined (high degree of refinement) protocol item during the *data extraction* activity. *Quality assessment criteria* is moderately refined during *quality assessment*. The *motivation for conducting the review* protocol item is not refined after the *study searching* activity.

In the following six subsections, we discuss in detail the survey results concerning to what extent and how the main protocol items are refined during the review process.

1) **Motivation for conducting the review:** The first item to be defined in a review protocol should be *motivation for conducting the review*. The motivation should address the importance of conducting the SLR, including the review goals, and a discussion of the importance of the topic being reviewed.

– General survey results: As Figure 1 shows, the *motivation for conducting the review* tends to be relatively well-defined in the planning phase. It is possible to observe that it is slightly refined during the later phases of the review process. In fact, there should be no major changes in the motivation, since the goal of the SLR should be clear and discernible since the beginning of the review process. As pointed out by the participants in the survey, minor changes typically occur along the review process to improve the scope and to detail the research goals.

– Survey results by level of experience: Figure 2 shows the participants’ opinions divided into the three aforementioned groups (G1, G2, and G3). With an objective to facilitate comparisons, we added a line in this figure (labeled as “All”) presenting the overall opinion of the 53 respondents. As Figure 2 shows, on the one hand, regardless of the experience level in conducting SLRs, participants of the three groups agree that refinements to the motivation for conducting the review are concentrated in the planning phase and they decrease with the review process. On the other hand, more experienced participants (in particular from G3) perceive that the motivation for the review can suffer small refinements during the activities ahead.

2) **Research questions:** RQs define what reviewers want to know about a given research topic. They are the heart of an SLR, since they guide it.

– General survey results: As expected (see Figure 1),

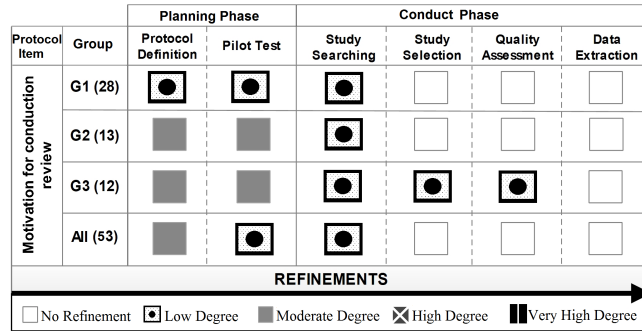


Figure 2. Refinement degree in the *motivation for conducting the review* item – Level of experience versus degree of refinement.

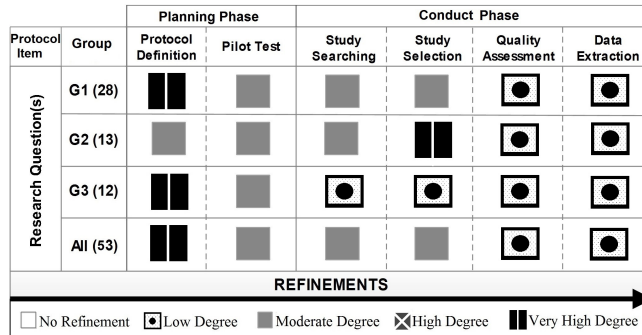


Figure 3. Refinement degree in *research question(s)* item – Level of experience versus degree of refinement.

RQs are strongly refined (high and moderate degree of refinement) during the planning phase. Moreover, RQs continue to be refined in a greater extent during the conduct phase, mainly in the *searching* and *selection* activities. One participant said that in *pilot test* and also *study searching*, it is possible to find some studies pointing out an interesting aspect that should be investigated, leading to new RQs. A second participant said that during *study selection*, researchers should check if the selected studies are able to answer the RQs. Finally, some participants said that they could adapt the RQs, after realizing what can and cannot be answered with the obtained data.

– Survey results by level of experience: Figure 3 shows the three groups agreeing that refinements in the RQs are mainly concentrated in the planning phase, but that they are also frequent in the conduct phase. Considering the central tendency (median) of the answers from the three groups, the participants agree on the same degree of refinement of the RQs during *pilot test*, *quality assessment*, and *data extraction* activities. There is no consensus among groups with respect to the degree of refinement of the RQs during *study searching* and *study selection*.

3) **Search strategy:** A search strategy looks for as many primary studies related to the RQs as possible.

– General survey results: As Figure 1 shows, refinements in the search strategy are greater during the *protocol def-*

inition, *pilot test* and *study searching* and *study selection*, since these activities put the search strategy into practice. In particular, during the *pilot test*, the search string is submitted to part of the selected databases (typically one of them), and the results are checked against a set of relevant studies (the control group), typically suggested by experts. If the studies of the control group are not retrieved, the string needs to be calibrated (adjusted) or new databases need to be added to guarantee that at least these studies are retrieved. In this sense, the coverage of the control group is decisive for the definition of the search strategy.

One participant said that during the *pilot test*, new terms can be added or removed in the search string to avoid noise in the outcomes. There are many other potential causes for refinements in the search strategy. Corroborating our results, a participant said that protocol reviewers should identify the lack of any relevant term or database. Changes in the RQs can also lead to changes in the search strategy. Refinements still keep continuing during *study searching* and *study selection*, as presented in Figure 1.

– Survey results by level of experience: As shown in Figure 4 participants of groups G1 and G2 agree that refinements in the *search strategy* continue to be performed during the conduct phase (i.e., during the *studies searching* and *studies selection*). This result is in accordance with the overall results (Figure 1). Participants of G3 perceive these refinements at a lower intensity level.

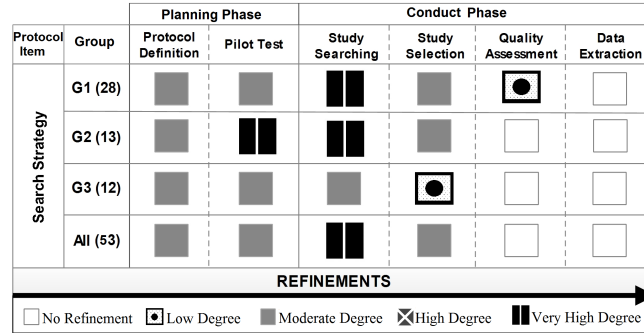


Figure 4. Refinement degree in *search strategy* item – Level of experience versus degree of refinement.

One plausible explanation for this difference of perception could be the high level of experience of researchers of G3 in both the topic being investigated and the review methodology, and this experience can lead to a low degree of refinement of the *search strategy*. Considering the central tendency (median) of answers of the three groups, participants share the opinion that the degree of refinement of *search strategy* is moderate during the *protocol definition* and that there is no refinement in the *data extraction* activity.

4) **Strategy for identification of primary studies:** The set of inclusion and exclusion criteria are written during the planning phase. However, the *study selection* activity is done in an iterative way [7] and this may even lead to refinements of the inclusion and exclusion criteria.

– General survey results: It is possible to observe in Figure 1 that the degree of refinement in the *strategy for identification of primary studies* item considerably increases during the *study selection*, since the clarity and reliability of the selection criteria are tested in this activity. Confirming this result, one participant stated that during the *study selection*, the inclusion/exclusion criteria were reformulated for clarity purposes.

– Survey results by level of experience: Regardless of the participants’ experience, they all agree that the degree of refinement of the *strategy for identification for primary studies* item during the activities that compose the planning phase (i.e., protocol definition and pilot study) is moderate (see Figure 5, columns 3 and 4). It is intensified in *studies selection* during the conduct phase, since this activity puts the selection criteria and procedures into practice. Considering the central tendency (median) of answers of the three groups, participants share the opinion that the degree of refinement of the *strategy for identifying for primary studies* is low during the *quality assessment* and *data extraction*.

5) **Quality assessment criteria:** Hassler et al. [9] discuss that one of the most challenging issues in SLR is the quality assessment that, although difficult, it is feasible to apply it

assessment in SLRs.

– General survey results: It is possible to observe in Figure 1 that the *quality assessment criteria* item has a reduced refinement throughout the activities of the conduct phase, except for the *quality assessment* activity, since these criteria are strongly used during this activity. The use of checklists is likely to improve the quality of SLRs. On this issue, one participant mentioned that there is no consensus among researchers regarding what quality criteria (checklists) should be used in SE. During the *quality assessment* activity, feedback from reviewers can help to refine the items of the checklist.

– Survey results by level of experience: As shown Figure 6, the degree of refinement of *quality assessment* range from no refinement to moderate. There is no high or very high degree of refinement in activities of the SLR process. The three groups agree that the degree of refinement of *quality assessment* is low during the *studies selection* and it is moderate during the *quality assessment*. The degree of refinement increases one degree after the *studies selection*. For example, considering the median, participants of G1 classified the degree of refinement as low during the *studies selection* and as moderate during the *quality assessment* (low to moderate – one degree more), since this activity puts checklists into use.

6) **Data extraction form:** Similarly to the study selection, data extraction is also performed in an iterative way [7], leading to refinements of the *data extraction form*, which is designed to collect all information needed to address the RQs.

– General survey results: It is possible to observe in Figure 1 that the *data extraction form* has a reduced refinement throughout the activities of conduct phase, except for *data extraction* activity. This can be explained as during the reading of the full text of the studies, additional data extraction fields may be necessary and thus could be considered in a reviewed version of the data extraction form. Some participants expressed the belief that during the *data extraction* activity, ideas often emerge showing how to structure the

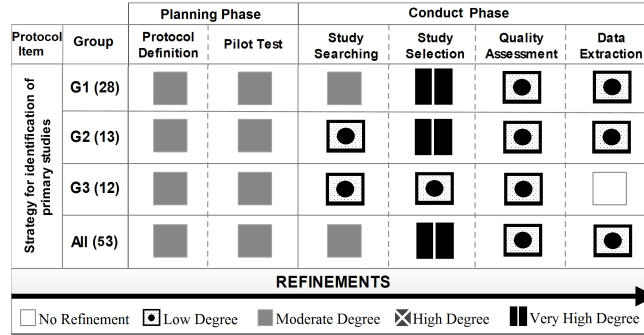


Figure 5. Refinement degree in *strategy for identifying the primary studies* item – Level of experience versus degree of refinement.

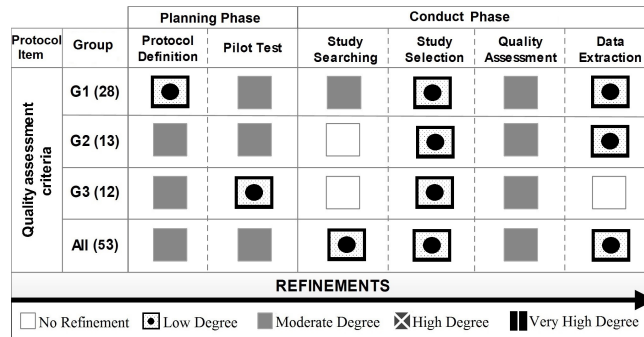


Figure 6. Refinement degree in *quality assessment* item – Level of experience versus degree of refinement.

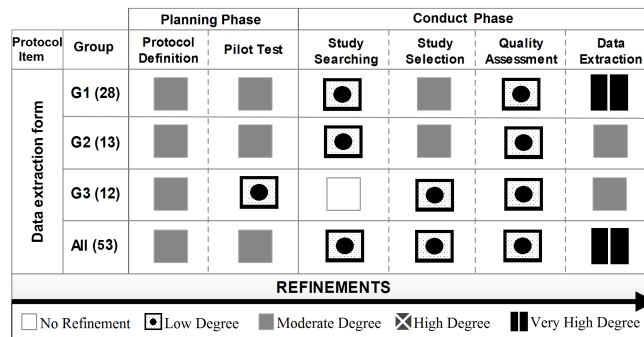


Figure 7. Refinement degree in *data extraction form* item – Level of experience versus degree of refinement.

extraction form in a better way.

– Survey results by level of experience: Corroborating with the overall results (Figure 1), the degree of refinement increases during the *data extraction* activity. Considering the median, participants of G1 and G2 argue that there is a moderate refinement during the *studies selection* (Figure 7). A possible explanation is that during the reading of the full text, there is the possibility in identifying new features to be extracted. There is a consensus that during the *protocol definition*, the degree of refinement is moderate and that during the *quality assessment*, the degree is low. One participant mentioned that only during the *studies selection* it was perceived that data extraction form did not contain

all data needed to answer the RQs.

B. Discussion

Traditionally, SLR protocol is firstly prepared and researchers take a small sample of studies (e.g. studies returned from a unique database) and use them to pilot the protocol. We can rename the pilot test as “mini SLR”. Insights through the pilot test may be used to modify protocol items. In fact, pilot test provides the opportunity to deal with problems and make refinements. Refinements are essential to ensure good protocols.

Our results show that the greatest protocol refinements occur during the *protocol definition* and *pilot test*. Based on our results we advocate that an iterative pilot test could

be beneficial to refine a protocol because in each iteration, there is a progressive gain in expertise. In summary, the SLRs process should be composed of n “mini SLR”. In our opinion, the main advantage of an iterative approach is the refinement opportunity. Completeness and repeatability are the two main qualities required from an SLR [16] and to achieve both qualities, iterations (refinements) are highly recommended [17]. Figure 8 illustrates an example of a “mini SLR” iteration.

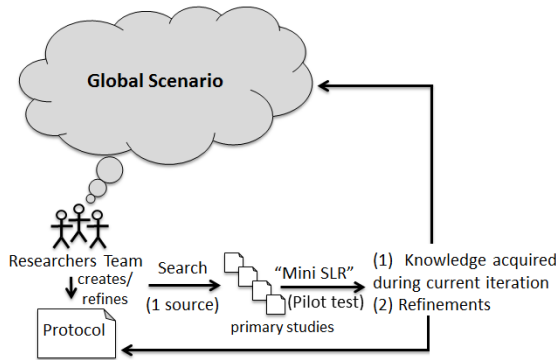


Figure 8. “Mini SLR” iteration.

The first iteration (“mini SLR – 1”) could begin with the elaboration of a preliminary protocol. This preliminary protocol should contain the protocol items defined in Kitchenham’s guidelines [4]. In order to facilitate the elaboration of the preliminary protocol, we recommend to: (1) reuse parts of protocols of published SLRs in the same domain of one’s interest; (2) consult complementary guidelines to complete protocol items; for example, the set of databases. As recommended by Kitchenham and Brereton [1] and Dybå et al. [18], the use of IEEE and ACM as well as two indexing databases (e.g., Inspec/Compendex and Scopus) is sufficient to search evidences in SLRs; and (3) start the definition of exclusion criteria using a set of general criteria, such as, “the study does not have a summary”; “the study is not a primary study”; “it is not possible to have access to the study”; among others.

As illustrated in Figure 8, in each “mini SLR” iteration, researchers refine protocol items. Based on our results, the *study selection* activity can lead to RQs refinements. Only after reading studies, it is possible to verify which RQs can and cannot be answered. Moreover, new RQs can emerge. During the *study searching* activity, researchers can note the need of new keywords or terms to the search string or even removal of some of them and also inclusion of new studies into the control group. Inclusion/exclusion of quality checklist items can be checked during the *quality assessment* activity. Similarly, inclusion/exclusion of *data extraction* items can be checked during the *data extraction* activity. Moreover, there is an amount of implicit knowledge (that we refer to as “global scenario”) acquired from primary

studies usually found in each iteration. Interesting to say that this global scenario contains knowledge from included studies, which and why studies were excluded, as well as the entire process and tasks by which the iteration was carried out. This global scenario can ground all tasks that are conducted during the iteration.

We can affirm that the data extraction form is the most modified item. If too much data is extracted, it is a waste of effort and time, on the other hand, if too little data is extracted, revisiting studies for more data is necessary, therefore, it is also waste of effort and time. The main data that should be collected include the title, authors, publication year, reviewer identification (if more than one), reference type, and aim of the studies. Other specific data can be selected by analyzing each RQ and identifying data that should be extracted from primary studies to answer these questions. It is important to mention that some refinements may require running again activities that were already completed to avoid that the answer to one question would influence the next question’s answer. For example, if the data extraction form is refined during the second iteration, and any data that have already been extracted from the first set of studies, it might require to re-extract data from these studies. In these cases, an iterative SLR process could be also beneficial, because most of the refinements are made in the first iterations and the trend is that refinements decrease as iterations increase. Consequently, less rework will be generated, since in the first iterations less searches have been carried out, less studies have been analyzed, etc.

A difficulty during the iterations is how to merge data extracted from new studies into the set of data extracted in the previous iterations. This merging must be properly done to further carry out the data synthesis and result report. One possibility to outcome this difficulty is to store previous data in a standardized format, e.g., tables in spreadsheet programs. Besides supporting data merging, tables also make logical trends easier for readers to follow and are the simplest type of data presentation. Tables also make it possible to combine key findings from different tables, enabling quantitative analysis and creating more complete and useful data synthesis.

For this study, we considered the following threats to validity: (i) the questionnaire construction: this is related to our ability to create questions that reflect the issue to be researched. To mitigate this threat, we created questions containing words that do not influence the way participants think about the issue. We also sorted the questions to avoid that the answer to one question does not influence on the next. Besides multi-choice questions, we encouraged participants to share their experiences in free comments; (ii) the participants’ responses: the survey was conducted as a post-mortem analysis, making participants reflect upon their past SLRs efforts. It is not possible to ensure how well they remember exactly when and how the review protocols were

changed. We believe that the participants responded their best their memories allowed.

IV. CONCLUSIONS

The protocol is an important element of an SLR and needs to be carefully prepared and documented to achieve reliable results and further update such study.

In this scenario, the main contribution of this work was a survey answered by different researchers, and their opinions offered us an opportunity to better understand how the creation of the protocol has been performed. Results provided important evidence about which phases and activities cause more refinements in protocols. We identified that the activities that most lead to further refinements are *protocol definition* and *pilot test*. We also noted that the *data extraction form* is the most modified item. We argue that piloting the protocol is an important activity, since it gives to reviewers the opportunity to refine their protocols.

Our results confirm that the protocol is in fact iteratively defined, and this iteration is perceived in a similar way by researchers with different levels of experience in SLR. We can affirm that iterations are unavoidable; a good quality SLR could only be obtained from iterations in the process and in the protocol definition.

As future work we intend to investigate if there is an association between protocol complexity (higher numbers of procedures, criteria and reviewers, large-scope, etc) and the number of refinements.

ACKNOWLEDGMENT

The authors would like to thank the Brazilian funding agency CNPq (Process n. 401033/2016-3).

REFERENCES

- [1] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, no. 12, pp. 2049–2075, 2013.
- [2] P. Brereton, B. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007.
- [3] S. Imtiaz, M. Bano, N. Ikram, and M. Niazi, "A tertiary study: Experiences of conducting systematic literature reviews in software engineering," in *17th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. ACM, 2013, pp. 177–182.
- [4] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele University and Durham University, UK, Tech. Rep. EBSE 2007-001, 2007.
- [5] M. Staples and M. Niazi, "Experiences using systematic review guidelines," *Journal of Systems and Software*, vol. 80, no. 1, pp. 1425–1437, 2007.
- [6] H. Zhang and P. Muhammad, A.B. Tell, "Identifying relevant studies in software engineering," *Information and Software Technology*, vol. 53, no. 1, pp. 625–637, 2011.
- [7] H. Zhang and A. Muhammad, "Systematic reviews in software engineering: An empirical investigation," *Information and Software Technology*, vol. 55, no. 7, pp. 1341–1354, 2013.
- [8] —, "An empirical investigation of systematic reviews in software engineering," in *5th International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2011, pp. 1–10.
- [9] E. Hassler, J. Carver, N. Kraft, and D. Hale, "Outcomes of a community workshop to identify and rank barriers to the systematic literature review process," in *18th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2014, pp. 1–10.
- [10] P. Mian, T. Conte, A. Natali, J. Biolchini, and G. Travassos, "A systematic review process for software engineering," in *2nd Software Engineering Latin American Workshop (ES-ELAW)*, 2005, pp. 498–507.
- [11] J. Biolchini, P. Mian, A. Natali, and G. Travassos, "Systematic review in software engineering: Relevance and utility," Federal University of Rio de Janeiro (COPPE/UFRJ), Brazil, Tech. Rep., 2005.
- [12] M. Riaz, N. Sulayman, M. Salleh, and E. Mendes, "Experiences conducting systematic reviews from novices' perspective," in *14th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2010, pp. 1–10.
- [13] B. Kitchenham and S. Pfleeger, "Personal opinion surveys," in *Guide to Advanced Empirical Software Engineering*. Springer London, 2008, pp. 63–92.
- [14] B. Kitchenham, "The role of replications in empirical software engineering – a word of warning," *Empirical Software Engineering*, vol. 13, no. 1, pp. 219–221, 2008.
- [15] P. O. Brereton, "A study of computing undergraduates undertaking a systematic literature review," *IEEE Transactions on Education*, vol. 54, no. 4, pp. 558–563, April 2011.
- [16] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*, ser. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series, 2015.
- [17] B. Oates and G. Capper, "Using systematic reviews and evidence-based software engineering with masters students," in *13th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2009, pp. 79–87.
- [18] T. Dybå, T. Dingsøy, and G. Hanssen, "Applying systematic reviews to diverse study types: An experience report," in *1st International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2007, pp. 225–234.