

# The Use of Grey Literature Review as Evidence for Software Engineering

Fernando K. Kamei<sup>1,2</sup>  
Advisor: Sérgio Soares<sup>1</sup>  
Co-advisor: Gustavo Pinto<sup>3</sup>

<sup>1</sup>Center of Informatics – Federal University of Pernambuco (UFPE)  
Recife – PE – Brazil

<sup>2</sup>Department of Informatic – Federal Institute of Alagoas (IFAL)  
Maceió – AL – Brazil

<sup>3</sup>Faculty of Computing – Federal University of Pará (UFPA)  
Belém – PA – Brazil

fernando.kenji@ifal.edu.br, scbs@cin.ufpe.br, gpinto@ufpa.br

Level: PhD

Program: Computer Science of the Center of Informatics (CIn-UFPE)

Entry year in program: 03/2018

Expected date for final presentation: 03/2022

Date of approval of the thesis: Not qualified yet

Event: SBES

**Abstract. Context:** *In the last years, the Grey Literature (GL) had a notable increase in interest in diverse areas of knowledge. In Software Engineering (SE), with the advance of social media, the SE practitioners are heavy consumers of GL, opposite to traditional research papers, such as the Systematic Literature Review (SLR).*

**Problem:** *Despite the increase of numbers of SLRs, some researchers claim the lack of connection to the practice of SE, due to the lack of evidence derived from SE practitioners.*

**Goal:** *Introduce methods and techniques to support a Grey Literature Review (GLR) in SE, aiming to approximate the literature review to the practice of SE.*

**Method:** *First, we will investigate how secondary studies use GL; Second, we will employ a survey with researchers of SLRs and SE practitioners; Third, we will analyze the use of GL according to the context; Fourth, we will review the GL source in SE; Finally, we will propose a guideline to help SE researchers and practitioners to conduct the GLR.*

**Preliminary findings:** *We have initiated the tertiary study conducting the manual (4,875 studies) and automatic (6,279 studies) searches. We eliminated those that were not an article, and also those were not written in English, remaining 8,019 potential studies. Currently, we are applying the inclusion and exclusion criteria.*

**Keywords:** *Grey Literature; Systematic Literature Review; Grey Literature Review.*

## 1. Problem Characterization

The Grey Literature (GL) is a source of evidence that has not gone through quality control mechanisms (peer review) for a publication. Several areas of knowledge have investigated the use of GL, e.g., Medicine [1] and Management [2]. GL has some benefits, e.g., provide data not found within commercially published literature, reducing publication bias, and facilitating a more balanced view of the evidence [1].

In Software Engineering (SE), as more and more content of GL is published, SE researchers are paying attention to the potential importance of GL, which their content can make up for the deficiency of practitioners evidence in traditional literature. [3]. The lack of evidence of SE practitioners in traditional literature makes a poor connection between Systematic Literature Reviews (SLR) and the practice of SE. Some studies discussed this problem [4, 5, 6].

According to Williams [3], one of the reasons for those problems is that practitioners rarely publish and use academic forums which focused on traditional literature. Instead of, they use social media and other tools to share and acquire knowledge [7]. This shows the prominent of the Social Media field as GL evidence for SE.

Aiming to fill this gap of lack of connection of SLR to the practice, Garousi et al. [5] proposed a method called as Multivocal Literature Review (MLR) with the aim of include the evidence of GL derived from SE practitioners on the conduct of SLR. However, this approach did not solve the well-known problems to conduct SLR, e.g., time-consuming, high effort, and costly [8], but increase them when adding GL content.

For this reason, we advocate that the use of Grey Literature Review (GLR) is more feasible and prominent to close the gap between the SE literature review and a lack of connection to SE practice, due to the fact that SE practitioners create and shared content on some GL source. However, some challenges are also present on the search for GL, e.g., time-consuming to find, multiple search engines or multiples sites may be required, there may be considerable volume or results found [5, 1], and the quality of information [9]. Still, to the best of our knowledge, there are few specific guidance available for performing Grey Literature Review in SE, with methods and techniques to support the researchers with best practices to conduct a GLR.

In that context, the overarching goal of this research is to investigate, propose, and evaluate strategies to support researchers to with best practices on how to conduct a Grey Literature Review focusing on: (1) help on how to retrieve an evidence from GL; (2) on how to answer a specific kind of research question. Our guideline will provide benefits to both researchers and SE practitioners to help mitigate the gap of Evidence-Based Software Engineering [10] and better connect with the practice of SE.

## 2. Theoretical Foundation and Related Work

This section presents our core concepts. First about the preeminent area of Evidence-Based Software Engineering and Systematic Reviews, with their concepts and some research gaps. Then, the concepts of Grey Literature and their usage in SE are present, and some related studies.

## 2.1. Evidence-Based Software Engineering

Kitchenham et al. [10] proposed the term Evidence-Based Software Engineering (EBSE), aiming “*to provide the means by which current best evidence from research can be integrated with practical experience*”, in other words, to identify and appraise all relevant evidence to the problem or technology under consideration. The Evidence-Based Medicine (EBM) [11] inspired the idea of EBSE. The most common use of SLR is both in terms of Systematic Reviews (SR) and Mapping Studies (MS).

Since the introduction of EBSE, some tertiary studies showed an increase in numbers of SLR published [4]. Kitchenham et al. [12] found 20 screened secondary studies published from January 2004 to June 2007. In a similar effort, Da Silva et al. [4] found a total of 120 SLRs published from January 2004 to December 2009. Lastly, Cruzes and Dybå [13] has found 169 SLRs published from January 2004 to July 2010. This shows the emerging character of the field.

However, the increasing interest in SLRs revealed important issues. For instance, Da Silva et al. [4] claimed that few SLRs are providing guidelines to practitioners. Garousi et al. [5] observed that most of the results provided by SLRs focused on academic needs, meaning that maybe the voice of the practitioners are not being heard. Finally, Car-taxo et al. [6] perceived that SLRs did not focus on questions that interest practitioners.

## 2.2. Grey Literature in SE

The most widely used and accepted definition of the term Grey Literature is the so-called Luxembourg definition [9], which states: “*that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body*”.

Kitchenham et al. [14] conducted the first SLR in SE, including the content of GL. The findings of the research showed the importance to include GL, especially to the case of Research Questions (RQ) that need practical and technical answers. Years later, Yasin and Hasnain [15] also confirmed that GL answers more technicals RQs, and found that the evidence of grey literature in SLRs is around 9% of total evidence found.

Garousi et al. [9] then proposed a method called Multivocal Literature Review (MLR), which includes GL as a data source to conduct SLRs. They classified the GL in three different tiers: 1st tier is High outlet control/High credibility, 2nd tier is Moderate outlet control/Moderate credibility, and 3rd tier is Low outlet control/Low credibility. We found three studies conducted using the MLR method [16, 17, 18].

Raulamo-Jurvanen et al. [19] performed the only Grey Literature Review we have known in SE, aiming to analyze how software practitioners address the practical problem of choosing the right test automation tool. The data derived from experiences and opinions were present in most of the findings.

All those studies show the importance of experience and opinions of SE practitioners as a source of evidence, that is embedded on the social and communication channels. Storey et al. [7] listed and classified these channels according to: code hosting sites (e.g., GitHub and BitBucket), face-to-face interactions (e.g., Skype), Q&A sites (e.g., Stack Overflow and Quora), search engine (e.g., Google and DuckDuckGo), microblogs (e.g.,

Twitter), private chats (e.g., IM, Skype chat, and Google chat), feeds and blogs, private discussions (e.g., email), public chats (e.g., IRC), discussion groups (e.g., mailing lists, Google groups, and forums), aggregators (e.g., Reddit), project coordination tools (e.g., Trello), Social networks sites (e.g., Facebook groups and Google Plus), rich content (e.g., screencasts, podcasts, and videos). All these channels can be included on the 2nd and 3rd tier of GL for SE. According to Williams [3], the increase of numbers of research which investigate some social media channel shows the prominent of this field.

Our research is related to EBSE, mainly in how to conduct an effective SLR, MLR, or GL in SE. The previous studies are related to our research, some directly related to grey literature, and others to EBSE. However, there is a lack of support to conduct an SLR using only the GL as a source of evidence in SE research. Even though the method MLR proposed to continue the search in traditional literature and adding the content of GL source, the problem claimed by Carver [8] not only will continue but will increase them.

### 3. Methods

The goal of this research is to bring the EBSE closer to the SE practice, with a proposal of a method that could help researchers to conduct SLR using only Grey Literature as a source of evidence. In this section, the research questions and the proposed approaches to achieve them are present. We also present the planned schedule of our research until thesis defense in Table 1.

**RQ1:** *How do secondary studies use grey literature?*

*Rationale:* This question intends to identify how evidence derived from Grey Literature is used in secondary studies of SE, aiming to understand the relation of that evidence to answer specific kind of research question. Still, understand if the use of grey literature as a data source of evidence has increased in the synthesis of secondary studies over the years.

*Approach:* We are planning to conduct this approach in three parts: 1) We will retrieve a list of secondary studies found in previous tertiary studies in SE [12, 20, 4, 13], whose investigated 169 SLRs between 2004 and 2010; 2) We will conduct a Tertiary Study to find all Secondary Studies in SE published between 2011 and 2018; 3) With those list of secondary studies, we plan to investigate and integrate all available evidence about the use of GL.

**RQ2:** *What are the researchers' motivations for relying or not on grey literature?*

*Rationale:* This question intends to understand what motivates and demotivates researchers to feed on GL, identifying the potential issues and problems of their use, and the reason to did not include GL.

*Approach:* We are planning to conduct a web-survey and interviews to collect the researcher's opinions. We will divide the researchers into two groups: the first group with authors of most cited SLRs that used evidence from GL, and the second group with authors of most cited SLR that did not use evidence retrieved from GL.

**RQ3:** *What are the practitioners' motivations for relying or not on grey literature?*

*Rationale:* This question intends to understand what motivates SE practitioners to

feed on GL, identifying the potential issues and problems of their use.

*Approach:* We are planning to conduct a web-survey and interviews to collect the practitioner's opinions. We will select the practitioners from the most cited different groups of the GL data source retrieved from our Tertiary Study.

**RQ4:** *What is an evidence in a grey literature?*

*Rationale:* This question intends to present an overview of grey literature sources used for SE practitioners, aiming to understand how it works, what is considered evidence, and how we can infer this evidence, aiming to help SE researchers and practitioners to get and analyze grey literature content.

*Approach:* We are planning to conduct a general review of empirical studies that used a group of GL. Moreover, employ a web-survey with open-ended questions with researchers and SE practitioners to understand. In both to improve the understand and perception of evidence derived from GL.

With those research questions, we intend to propose a guideline with methods and practice to conduct a Grey Literature Review in SE. We will evaluate and validate the guideline in the practice of Software Engineering through web-surveys and interviews with researchers experts on SLR.

#### **4. Preliminary Results**

Our research is on the 2nd year, and we have started the execution of our tertiary study (Phase 1). First, we concluded the planning and elaboration of the protocol. Second, we performed our search for peer-reviewed articles published between 2011 and 2017. We combined automatic (five digital libraries) and manual (five journals and three conferences) searches to increase the coverage.

We performed the automatic search using the same search string of Da Silva [4]. This search retrieved 6,279 documents. The manual search retrieved 4,875 documents. We recorded all title, abstract, year, and authors of published articles in each source. We put the results of the two searches together, and we eliminated irrelevant (not an article) and duplicated studies, remaining thousand separator potential studies.

We have finished the step to *apply inclusion and exclusion criteria*, where at least two researchers appraised each study. *At the moment* we are solving all conflicts over the opinion of a certain paper, where we are stress again to discuss if the study should be included or not. However, if they do not agree, the opinion of a third researcher should be considered. We found a substantial degree of concordance of each pair using the Kappa coefficient.

For the next steps, we are planning to initiate the procedure of data extraction from remaining studies, considered potential relevant studies. Lastly, conduct an appraisal of the quality of studies. The Table 1 presents the planned schedule of our research until thesis defense.

#### **5. Limitations**

We are aware that our research has some limitations. We have some limitations to conduct our tertiary study to answer RQ1: the search string, search sources, and selection

**Table 1. Research Schedule**

Phase/Period	Jan-Dec 2019	Jan-Jun 2020	Jul-Dec 2020	Jan-Jun 2021	Jul-Dec 2021	Jan-Mar 2022
<b>P1</b>						
<b>P2</b>						
<b>P3</b>						
<b>P4</b>						
<b>P5</b>						
<b>P6</b>						
<b>P7</b>						
<b>Thesis</b>						

procedures. These are known bias in the area [21]. We will mitigate them conducting a broad search using automatic and manual search, and always conducting each step of selection procedures in pairs. Aiming to avoid the bias of the individual interpretation of the procedures to answer RQ2 and RQ3, we will use the cross-case analysis through multiple data source collects.

## 6. Concluding Remarks

This paper presents our research proposal, aiming to improve the EBSE to focus on the practitioners need. Preliminary research and results also revealed that the evidence of practitioners is still limited in traditional SLRs.

Also, we believe that the existing proposals for conducting SLR are essential to academic research. However, it is necessary an approach focused only on practitioners, being closest to their voice in earlier stages, and including only grey literature as a source of evidence sounds more reliable to an environment that practitioners emerge.

## References

- [1] Arsenio Paez. Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3):233–240, aug 2017.
- [2] Richard J. Adams, Palie Smart, and Anne Sigismund Huff. Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews*, 19(4):432–454, apr 2016.
- [3] Ashley Williams. Using reasoning markers to select the more rigorous software practitioners online content when searching for grey literature. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018 (EASE '18)*. ACM Press, jun 2018.
- [4] Fabio Q. B. da Silva, André L. M. Santos, Sérgio Soares, A. César C. França, Cleiton V. F. Monteiro, and Felipe Farias Maciel. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Information and Software Technology*, 53(9):899–913, sep 2011.

- [5] Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. The need for multivocal literature reviews in software engineering. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE '16)*. ACM Press, jun 2016.
- [6] Bruno Cartaxo, Gustavo Pinto, Danilo Ribeiro, Fernando Kamei, Ronnie E. S. Santos, Fabio Q. B. da Silva, and Sérgio Soares. Using q&a websites as a method for assessing systematic reviews. In *Proceedings of the 14th International Conference on Mining Software Repositories (MSR '17)*. IEEE, may 2017.
- [7] Margaret-Anne Storey, Alexey Zagalsky, Fernando Figueira Filho, Leif Singer, and Daniel M. German. How social and communication channels shape and challenge a participatory culture in software development. *IEEE Transactions on Software Engineering*, 43(2):185–204, feb 2017.
- [8] Jeffrey C. Carver, Edgar Hassler, Elis Hernandez, and Nicholas A. Kraft. Identifying barriers to the systematic literature review process. In *Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '13)*. IEEE, oct 2013.
- [9] Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, 106:101–121, feb 2019.
- [10] Barbara A. Kitchenham, Tore Dybå, and Magne Jørgensen. Evidence-based software engineering. In *Proceedings of 26th International Conference on Software Engineering (ICSE '04)*. IEEE Comput. Soc, may 2004.
- [11] Gordon Guyatt. Evidence-based medicine. *JAMA*, 268(17):2420, nov 1992.
- [12] Barbara A. Kitchenham, O. Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering - a systematic literature review. *Information and Software Technology*, 51(1):7–15, jan 2009.
- [13] Daniela S. Cruzes and Tore Dybå. Research synthesis in software engineering: A tertiary study. *Information and Software Technology*, 53(5):440–455, may 2011.
- [14] Barbara A. Kitchenham, Pearl Brereton, Mark Turner, Mahmood Niazi, Stephen Linkman, Rialette Pretorius, and David Budgen. The impact of limited search procedures for systematic literature reviews - a participant-observer case study. In *Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM '09)*. IEEE, oct 2009.
- [15] Affan Yasin and Muhammad Ijlal Hasnain. On the quality of grey literature and its use in information synthesis during systematic literature reviews. Master's thesis, Blekinge Institute of Technology, Karlskrona, Sweden, 2012.
- [16] Vahid Garousi and Mika V. Mäntylä. When and what to automate in software testing? a multi-vocal literature review. *Information and Software Technology*, 76:92–117, aug 2016.

- [17] Håvard Myrbakken and Ricardo Colomo-Palacios. Devsecops: A multivocal literature review. In *Communications in Computer and Information Science*, pages 17–29. Springer International Publishing, oct 2017.
- [18] Vahid Garousi, Michael Felderer, and Tuna Hacaloğlu. Software test maturity assessment and test process improvement: A multivocal literature review. *Information and Software Technology*, 85:16–42, may 2017.
- [19] Päivi Raulamo-Jurvanen, Mika V. Mäntylä, and Vahid Garousi. Choosing the right test automation tool. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE '17)*. ACM Press, jun 2017.
- [20] Barbara A. Kitchenham, Riallette Pretorius, David Budgen, O. Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. Systematic literature reviews in software engineering - a tertiary study. *Information and Software Technology*, 52(8):792–805, aug 2010.
- [21] Barbara A. Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. techreport, Keele University and Durham University Joint Report, jul 2007.