UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Sustainability of Systematic Literature Reviews

Vinicius dos Santos

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________

Vinicius dos Santos

Sustainability of Systematic Literature Reviews

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate of Science – Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Elisa Yumi Nakagawa

USP – São Carlos
September 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Santos, Vinicius dos
785s Sustainability of systematic literature reviews / Vinicius dos Santos; orientadora Elisa Yumi Nakagawa. -- São Carlos, 2023.
198 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Systematic Literature Review. 2. Sustainability. 3. Software Engineering. I. Nakagawa, Elisa Yumi, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Vinicius dos Santos

Sustentabilidade de Revisões Sistemáticas da Literatura

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Elisa Yumi Nakagawa

USP – São Carlos
Setembro de 2023

I dedicate this work to my fiancée and my loved ones who never stopped believing in me.

Acknowledgements

Firstly, I would like to express my sincere gratitude to those who have been by my side throughout this journey. Without the unwavering support of my fiancée Vanessa, this work would not have been possible. She stood by me through countless revisions, presentations, and difficult moments, providing me with the strength and encouragement to succeed, there are no words to express my thankfulness. To my family, my father Pedro, my mother Maria, and my sister Ligia, I am indebted for their love and support, which gave me the security to thrive in a new city and experience that has profoundly changed me as a person. I also extend my gratitude to my second family, my in-laws Sueli and Nelson, who provided a listening ear and full support in this entire journey.

My advisor Elisa Yumi Nakagawa deserves special recognition for her invaluable guidance and support throughout my journey. Her advice has made me a better person, and I am fortunate to work with a this great professional and a human being who prioritizes the well-being of her students. I also extend my deepest gratitude to teachers Érica and Kátia for their contributions and support, which gave me the confidence to pursue this path. My heartfelt appreciation also goes to my research group and partners, particularly Anderson, who directly and indirectly supported me throughout this work, and friends Eduardo, Luiz, and Letícia, who helped me in several steps giving me the needed support to move on.

I would like to express my gratitude to FAPESP for their financial and intellectual support, which made my full dedication to this research project possible. I also thank the reviewers for their contributions to the continuous improvement and monitoring of the project. Lastly, I extend my thanks to the professors who participated in my doctorate, including those who participated in the qualification of this project, and all those who directly contributed to the production of this work, such as professors Rafael Capilla and Colin Venters.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - PROEX-11308091/D. This study was also supported by FAPESP (grants: 2015/24144-7, 2019/23663- 1), and CNPq (grant: 312634/2018-8).

The future depends on what you do today

(Mahatma Gandhi)

Abstract

SANTOS, V. **Sustainability of Systematic Literature Reviews**. 2023. 198 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Context: Systematic Literature Review (SLR) has been increasingly adopted in Software Engineering (SE) because it is considered the most reliable way to gather evidence from the literature. At the same time, SLR currently consumes much time and effort to be conducted, and its results have not impacted the software industry as expected. Despite several solutions proposed to solve the problems of SLR, the research remains very fragmented failing to provide a bigger picture of the problems and associated solutions. **Objective:** This study introduces the concept of sustainability in SLRs to the SE community as a holistic view of the current research scenario putting together a solid knowledge base that includes definitions, guidelines, and the main leveraging points of sustainability in current SLR conducted in SE. **Method:** After conducting a tertiary study investigating the main characteristics of sustainability of software systems, we applied the meta-ethnographic method to transpose sustainability characteristics into SLR context by identifying, adapting, synthesizing, and validating those which could aid researchers in conducting more useful studies with less effort. Finally, we extracted a sample of high-quality studies conducted in the SE area, next, we applied a sustainability view to identify the flaws that lead to unsustainable SLR. **Results:** Our study provides an overview of the role of sustainability in the SLR context and provides important assets for SE community to identify what are sustainable SLR, such as 15 characteristics of sustainable SLR and 15 sensitive points of SLR process that are considered critical factors to achieve sustainability. In addition, we propose 16 guidelines to support researchers to improve sustainability in their SLR and identified 13 leveraging points found in relevant studies in SE area that together represent actions that should be adopted immediately to enhance the sustainability of SLR. **Conclusion:** Our work represents a starting point for a mindset shift in the SE community about the importance of creating more sustainable reviews. Furthermore, we believe that sustainability is the missing key concept that could connect many advances achieved over the years of research and remove barriers that hinder SLR from providing useful information in a timely manner for academia and industry. Henceforth, we believe that researchers should observe sustainability as a first-class citizen while conducting/updating SLR to make them compatible with the sustainability view.

Keywords: Systematic Literature Review, Systematic Review, Software Engineering, Sustainability.

Resumo

SANTOS, V. **Sustentabilidade de Revisões Sistemáticas da Literatura**. 2023. 198 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Context: A Revisão Sistemática da Literatura (RSL) tem sido cada vez mais adotada na Engenharia de Software (ES) por ser considerada a forma mais confiável de reunir evidências da literatura. Ao mesmo tempo, uma RSL atualmente consome muito tempo e esforço para ser conduzida, e seus resultados não impactam a indústria de software como esperado. Apesar de várias soluções propostas para resolver problemas pontuais relacionados à RSL, a pesquisa permanece fragmentada e as alternativas falham em estabelecer conexões entre os problemas atuais de RSL em ES. **Objetivo:** Este estudo apresenta o conceito de sustentabilidade em RSL para a comunidade ES como uma visão holística do cenário atual de pesquisa, reunindo uma sólida base de conhecimento que inclui definições, diretrizes e os principais pontos de alavancagem da sustentabilidade das RSL conduzidas em ES. **Método:** Depois de realizar um estudo terciário investigando as principais características da sustentabilidade dos sistemas de software, foi aplicado o método meta-etnográfico para transpor as características da sustentabilidade para o contexto de RSL, identificando, adaptando, sintetizando e validando aquelas que poderiam auxiliar os pesquisadores na condução de estudos mais úteis com menos esforço. Por fim, foi extraída uma amostra de estudos realizados na área de ES, e em seguida, foi aplicada uma visão de sustentabilidade para identificar as falhas que tornam as RSLs insustentáveis. **Resultados:** O estudo apresentado fornece uma visão geral do papel da sustentabilidade no contexto de RSL e fornece recursos importantes para a comunidade de ES identificar o que são RSLs sustentáveis, bem como 15 características de RSLs sustentáveis, 15 pontos sensíveis do processo RSL que são considerados fatores críticos para alcançar a sustentabilidade. Além disso, foram propostas 16 diretrizes para apoiar os pesquisadores a melhorar a sustentabilidade em suas RSLs e 13 pontos de alavancagem encontrados em estudos relevantes na área de ES que, juntos, representam ações que devem ser adotadas imediatamente para aumentar a sustentabilidade da RSL. **Conclusão:** O trabalho apresentado representa um ponto de partida para uma mudança de mentalidade na comunidade de ES sobre a importância de criar RSLs mais sustentáveis. Para isso é necessário que os pesquisadores observem a sustentabilidade como um cidadão de primeira classe ao conduzir/atualizar uma RSL. Acredita-se que a sustentabilidade é o conceito-chave que faltava para conectar muitos avanços conquistados ao longo dos anos de pesquisa e remover barreiras que impedem a RSL de fornecer informações úteis em tempo hábil para a academia e a indústria.

Palavras-chave: Revisão Sistemática da Literatura, Revisão Sistemática, Engenharia de Software, Sustentabilidade.

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

1.1 Context

Systematic Literature Review (SLR) was originally proposed in medicine and became a key component of Evidence-Based Medicine (EBM) (SACKETT, 1997), aiming to support researchers to collect, synthesize, and interpret information from literature in a reliable way (KITCHENHAM; CHARTERS, 2007). The adaptation of the SLR method to address different fields, such as social policy and education, was due to the successful experience in the medicine area (KITCHENHAM *et al.*, 2015). Inspired by the success of EBM, researchers from Software Engineering (SE) area proposed Evidence-Based Software Engineering (EBSE) aiming to aid SE practitioners in providing a solid basis for a decision-making process that relies on scientific evidence (KITCHENHAM *et al.*, 2004a). To aid EBSE, researchers defined an SLR method to synthesize knowledge of a topic of interest and gather information from different sources to answer specific research questions associated with SE in a reliable way (KITCHENHAM; CHARTERS, 2007).

The guidelines proposed by Kitchenham and Charters (2007) became a cornerstone of most of the SLR conducted in the further years. These guidelines defined the SLR method containing three main stages, namely, *Planning*, *Conduction*, and *Reporting*. In the planning stage, researchers identify the motivation for an SLR, define the protocol (e.g., selection criteria, data sources used, search string, quality assessment criteria, etc.), and perform a pilot test to ensure the search strategy's validity and coverage. During the conduction stage, researchers execute the aforementioned protocol and identify and select primary studies according to the defined criteria. After the selection process, researchers extract data from the studies selected and synthesize it to draw conclusions. Finally, researchers report and disseminate results to the target audience through technical reports, conference papers, journal articles, book chapters, and master's dissertations/doctoral theses.

SE literature reports several benefits of conducting SLR, for instance, reduction of bias (KITCHENHAM; CHARTERS, 2007), the possibility of drawing more general conclusions from a wider range of situations and contexts (KITCHENHAM; CHARTERS, 2007; NIAZI, 2015), production of auditable and repeatable results (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a), and, the opportunity to identify research gaps and also perspectives for future investigations (KITCHENHAM *et al.*, 2015). Furthermore, when researchers conduct SLR following recommended guidelines, especially for documentation, SLR can be updated (KITCHENHAM *et al.*, 2011) making it possible to have up-to-date information (PETERSEN; ALI, 2011) even when new evidence emerges after SLR publication.

1.2 Problem Statement

SLR conduction involves diverse challenges (IMTIAZ *et al.*, 2013; CARVER *et al.*, 2013; FELIZARDO *et al.*, 2020b) especially for novice researchers (RIAZ *et al.*, 2010a; IWAZAKI *et al.*, 2022). Several of these challenges were already discussed in the literature, for instance, the lack of rigor of researchers in following well-experimented guidelines for SLR directly impacting the SLR quality (KUHRMANN *et al.*, 2017) and poor documentation of SLR (BUDGEN *et al.*, 2018a). Moreover, there are two real-world problems that remain without a feasible solution for SLR: (i) the lack of effective impact of SLR results in decision-making in the software industry (BADAMPUDI *et al.*, 2019; CARTAXO *et al.*, 2018); and (ii) the excessive amount of resources consumption (time and effort) demanded to conduct/update SLR (FELIZARDO; CARVER, 2020). Over the years, researchers have evolved the SLR method and proposed important solutions, for instance, developing techniques for better planning (CAIRO *et al.*, 2019; FELIZARDO *et al.*, 2017a), conducting (FELIZARDO *et al.*, 2017b), and reporting SLR (CARTAXO *et al.*, 2018), as well as tools (MARSHALL; BRERETON, 2013; MARSHALL; BRERETON, 2015) and automation techniques (FELIZARDO; CARVER, 2020). Despite representing an important evolution, these solutions only address punctual problems and do not encompass the whole extent of SLR problems.

As a consequence of the gradual increase in the number of secondary studies, including SLR and systematic mapping studies (SMS), in SE (BUDGEN; BRERETON, 2022), these problems aforementioned become more complex to solve demanding immediate actions to revert this situation. Continuing to neglect these problems yields several consequences for SE community, including the progressive accumulation of SLR that are conducted with a narrow vision of the process and creating a tendency of reporting results that are disconnected from the target audience needs, similarly, conducting studies that quickly become obsolete and unusable for future researchers due to their incapability of being audited, replicated, or reused due to poor documentation. This scenario also leads researchers to build increasingly complex guidelines that result in difficulties in their application into research practice. Another consequence is that many SE researchers waste their efforts searching for specific solutions without addressing the underlying causes of the problems. Ultimately, this situation leads to the creation of new types of literature reviews to fit specific contexts (e.g., Rapid Reviews (CARTAXO *et al.*, 2018), Live Reviews (ELLIOTT *et al.*, 2014)). Hence, **the main problem addressed in this Ph.D. thesis is that solutions proposed for SLR method are fragmented and unable to tackle the current SLR problems in an integrated way.**

1.3 Objective

Fragmentation of research about SLR method inspired us to propose a solution capable of expanding the community's perception of the SLR method and the importance of integrating current solutions. In this sense, it is crucial to recognize that the review process has evolved beyond being a simple research method, to transforming into a complex mechanism that plays a pivotal role in conducting SE research on a global scale. Systematic reviews form the foundation of numerous research projects and involve diverse communities of researchers and industry professionals.

The **main goal** of this Ph.D. project is to provide a holistic view of SLR proposing an integrated solution for current problems that impact the whole SLR lifecycle. As a consequence, we expect to raise awareness within the SE community over the importance of treating current problems in a more integrated way and fostering solutions that use our new view to solve problems. To achieve this goal, our research question is: **Is there a way to tackle the fragmentation of the solutions proposed for SLR and transpose them into a more integrated solution?**

To answer our question, we explored the literature in different knowledge areas to generate alternative solutions that we could adapt to the SLR context. In spite of the complexity of the problems aforementioned, we found that Ecology area faced an analogous scenario that fragmented solutions were proposed to mitigate the consequences of human actions in the environment, including the high consumption of natural resources, pollution, climate changes, and other aspects that impact directly our society, economy, and environment (PURVIS *et al.*, 2018a). In this context, sustainability refers to a development that meets the needs of the present without compromising the ability of future generations to meet their own needs (KUHLMAN; FARRINGTON, 2010). Over the years, the discussion about the meaning of sustainability and how it could be implemented gained popularity (PURVIS *et al.*, 2018b) and spread to multiple knowledge areas like economy, social sciences, and politics to present feasible solutions for the impacts of human growth (CABEZAS; DIWEKAR, 2012).

Sustainability became a driver in the discussion about our common future (KEEBLE, 1988) and, inspired by the success in Ecology and combined with the fact of software became a cornerstone of our modern society (BOOCH, 2008), researchers introduced sustainability into the SE area (HILTY *et al.*, 2011). Sustainability in SE aims to provide solutions for software development problems and reduce the negative impacts of our actions on future generations (HILTY *et al.*, 2011). Software sustainability popularization is considered a game changer because it contributes to changing the mindset of society about the importance of better handling social, economic, technical, and individual aspects of software development to create a software product that has a minimal impact on the environment (PENZENSTADLER *et al.*, 2014). We inspired ourselves by the positive experience of both areas in adapting sustainability to observe from a wider perspective the software development process, the product itself, and the impacts of introducing new software in society. Hence, our hypothesis is that sustainability serves as the missing key concept to bridge current SLR problems and connect current and future solutions towards a unified direction.

We decomposed our main question into four more specific questions to guide us throughout this research work presented below:

RQ_1 - Which is the state of the art of sustainability in the SE area?

RQ_2 - How could sustainability be applied to the SLR context?

RQ_3 - Which are the characteristics of sustainable SLR?

RQ_4 - How to leverage sustainability of SLR?

To address these research questions, we followed the research method detailed in the following section.

1.4 Research Method

We designed our research method in two main phases: sustainability characterization and evaluation (see Figure 1). Sustainability characterization comprises two main activities: A1 - *Characterization of sustainable software systems* and A2 - *Identification of sustainable SLR characteristics*. Sustainability evaluation phase also comprises two activities: A3 - *Evaluation of sustainable SLR characteristics* that appraised our results to reduce bias by double-checking with knowledgeable researchers, and A4 - *Identification of leveraging points of sustainability in SLR*, where we applied sustainability view to identifying practical opportunities to leverage SLR conduction in SE.

Figure 1 illustrates the inter-dependencies among activities A1 to A4 and artifacts flow across activities, furthermore, we present where the results derived from these activities are documented (from Chapters 2 to 5). Each chapter outlines a set of sub-questions that provide further specificity and structure to the research process. Table 1 presents a mapping of the research questions and sub-questions, associated with their activities, and the corresponding chapter where each question is addressed.

To comprehend in-depth what are sustainable software systems (A1), we formulated our first specific research question: *What is the state of the art of sustainability in SE area?* More specifically, our interest is focused on SE over Ecology for two main reasons: (i) the adaptation of sustainability for SE is relatively new and produced much knowledge on how

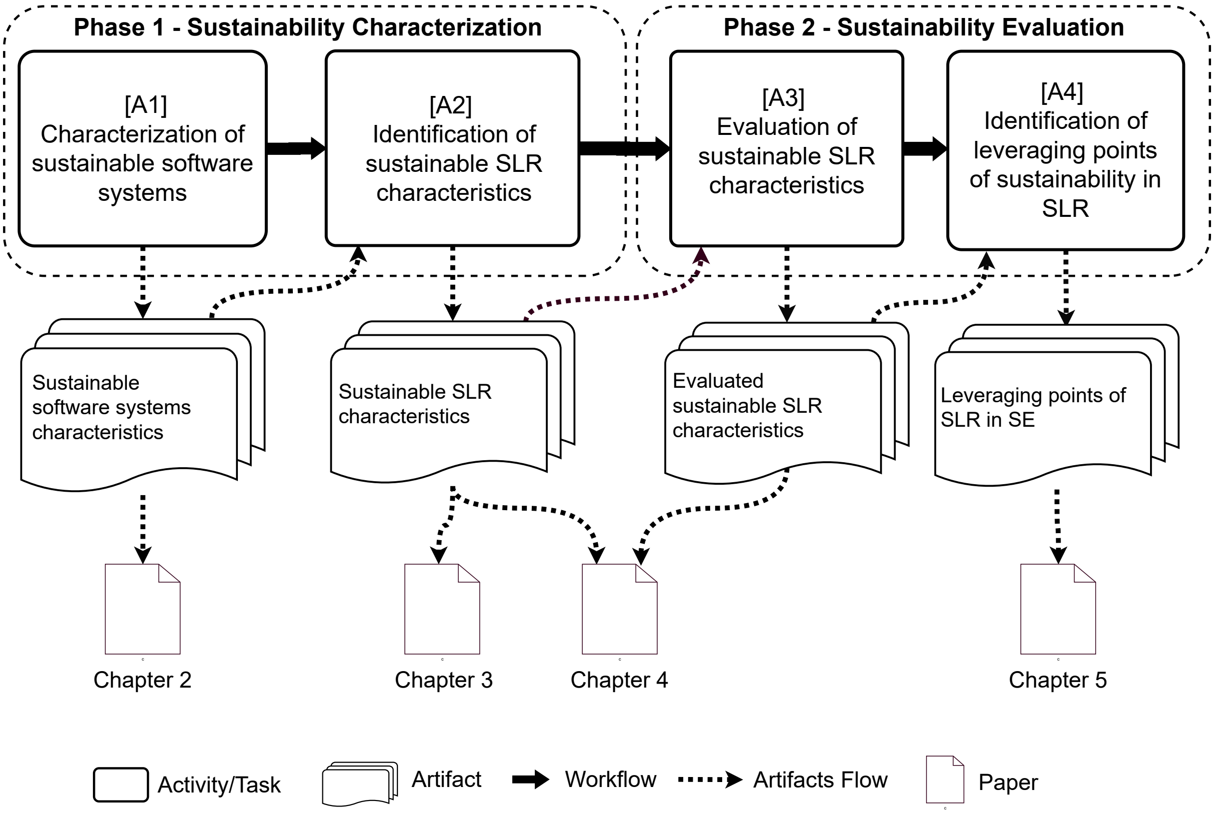

Figure 1 – Ph.D. project research method

sustainability could be adapted into different contexts, and (ii) SLR contains characteristics that could be somehow considered analogous to software development (e.g., a well-defined process, the need for knowledge management, good documentation, and collaboration with social groups), consequently, it makes SE a better starting point than Ecology. Nevertheless, whenever necessary we consulted the Ecology literature to ensure that we are interpreting correctly the concepts.

Our objective was to explore the definitions and main characteristics of sustainable software that could be useful for us. To achieve this, we conducted a tertiary study according to Kitchenham and Charters (2007) that consists of a systematic review focused on retrieving data from other secondary studies, such as SLR and SMS. In Chapter 2, we answered four specific questions: RQ1.1 - *What are sustainable software systems?*, RQ1.2 - *Which are the resources taken into account when considering sustainability in software systems?*, RQ1.3 - *Which are the critical factors for sustainable software systems?*, RQ1.4 - *Which are the means to leverage the development of sustainable software systems?*. These questions aimed to analyze the available literature to identify the primary characteristics that contribute to software sustainability, examine the critical factors involved in constructing sustainable software, and explore the prevailing approaches to achieving sustainability in this domain.

After a careful analysis of the state of the art of sustainability in SE, our next step was to use the knowledge acquired as a basis to propose an adaptation for sustainability for Systematic Literature Reviews. In the second activity (A2), we focused on comprehending the role of

Table 1 – Research questions addressed in this Ph.D. Project

RQ_{ID}	Question	Addressed in	Presented in
RQ_1	Which is the state of the art of sustainability in SE area?	A1	Chapter 2
$RQ_{1.1}$	What are sustainable software systems?	A1	Section 2.3.1
$RQ_{1.2}$	Which are the resources taken into account when considering sustainability in software systems?	A1	Section 2.3.2
$RQ_{1.3}$	Which are the critical factors for sustainable software systems?	A1	Section 2.3.3
$RQ_{1.4}$	Which are the means to leverage the development of sustainable software systems?	A1	Section 2.3.4
RQ_2	How could sustainability be applied to the SLR context?	A2	Chapter 3
$RQ_{2.1}$	Which are the current barriers of SLR?	A2	Section 3.3
$RQ_{2.2}$	What is sustainability of SLR?	A2	Section 3.4
RQ_3	Which are the characteristics of sustainable SLR?	A3	Chapter 4
$RQ_{3.1}$	Which are the core characteristics of sustainable SLR?	A3	Section 4.4.2
$RQ_{3.2}$	Which are the critical factors for sustainable SLR?	A3	Section 4.4.3
$RQ_{3.3}$	Which guidelines can be applied to conduct more sustainable SLR?	A3	Section 4.4.4
RQ_4	How to leverage sustainability of SLR?	A4	Chapter 5
$RQ_{4.1}$	What is the state of practice of SLR in SE?	A4	Section 5.4.1
$RQ_{4.2}$	Which impacts do the existing practices adopted by researchers have on the social, economic, and technical sustainability of SLR?	A4	Section 5.4.1

sustainability in the SLR context, for this, we defined our second research question is: RQ2 - *How can sustainability be applied in the context of SLR?*. To answer this question, we revisited barriers for the SLR method reported in six noteworthy studies published in the last decade (from 2012 until 2022); next, we used sustainability background to propose three dimensions: Social (related to human factors and social interactions), Economic (related to resource consumption during conduction/update); and, Technical (related to technologies and support tools). Finally, we categorized the barriers within these dimensions highlighting their possible interconnections. Our intention within this chapter is to bootstrap a discussion about (i) which could be a reasonable definition for the sustainability of SLR, (ii) which topics each dimension should address, (iii) which interconnections exist between dimensions and barriers reported in the literature, (iv) which are the practical consequences of arranging dimensions in different ways.

The previous work introduced a novel idea that sparked a discussion that required further exploration. While the work identified important aspects of sustainability of SLR, there was still a need for a deeper understanding of what made an SLR sustainable. Hence, we addressed the following research question: RQ3 - *What are the characteristics of sustainable SLR?* To investigate this, we applied the meta-ethnographic method (NOBLIT; HARE, 1988), which focuses on qualitative data analysis to gather information from the literature and establish connections through relevant keywords and expressions, resulting in higher-level interpretations that surpass individual studies. By employing this method, we put together the best evidence available about sustainability in SE and SLR to compare, translate, and synthesize findings to develop a cohesive and comprehensive understanding of the topic under investigation. Therefore, our objective was to dig deep into the sustainability of SLR addressing the following topics: RQ3.1 - *What are the core characteristics of sustainable SLRs?*, RQ3.2 - *What are the critical factors contributing to sustainable SLRs?*, RQ3.3 - *What guidelines can be applied to conduct more sustainable SLRs?*

Our final step was to assess the extent to which the conducted reviews align with the sustainable SLR. We provided practical insights for researchers on how they can enhance the sustainability of their studies. Hence, our final specific research question was: RQ4 - *How to leverage sustainability of SLR?* Considering the vast number of published reviews and the substantial heterogeneity among these studies, we employed a systematic search approach and sampling to minimize bias and draw our conclusions. This approach involved selecting secondary studies from two prominent journals in the field of SE (Information and Software Technology and Journal of Systems and Software) and extracting a sample of the ten most cited studies from the last decade. Subsequently, we utilized the previously established characteristics as benchmarks for our analysis, focusing on identifying areas for improvement in SLR. More specifically, our interest was to answer two research questions: RQ4.1 - *What is the state of practice of sustainability in SLR conducted in SE?*, RQ4.2 - *Which impacts do the existing practices adopted by researchers have on the social, economic, and technical sustainability of SLR?* Our objective was to present empirical evidence highlighting the sustainability pitfalls where current reviews flaw in practice. Based on these findings, we offer practical recommendations for researchers who want to improve their SLR' sustainability.

1.5 Contributions

This is the first work that addresses sustainability in SLR, hence, our primary **contribution** is the introduction of the sustainability concept as a driver for discussing SLR issues. In this sense, we provide a new perspective for SE community on the SLR problems and an alternative to treat fragmentation of research endeavors by proposing more integrated thinking. For this, our work defines what means to be sustainable in SLR by pinning which characteristics, critical factors, and guidelines that are relevant for achieving sustainability.

This work contributes to sustainability in SE providing a comprehensive view of the existing work in software sustainability. Despite other secondary studies being conducted before, our analysis goes beyond, connecting findings and unifying evidence that can guide future research in SE and also subside future adaptations of sustainability in other areas.

We contribute to EBSE by introducing the concept of dimensions allowing researchers to observe the SLR problems from the social, economic, and technical perspectives, henceforth, we argue that these issues are interconnected and cannot be solved in isolation, emphasizing the importance of adopting an integrated approach. In this sense, we applied this vision to map out harmful behaviors found in current reviews that pose significant threats to study sustainability. By highlighting these issues, we underscore the need for immediate corrections to ensure the integrity and quality of future research in the SE area. Additionally, a valuable contribution lies in our successful use of meta-ethnography to adapt sustainability characteristics. This serves as an exemplary demonstration of how effective solutions from other domains can be applied to the EBSE context.

We intend that this work helps to expand the comprehension of the SE community about the SLR method about the aspects throughout its life cycle that makes it more sustainable. This work contributes to fostering a profound shift in mindset within the SE community. For this, it is essential that further solutions not be developed in an ad-hoc manner but instead guided by the principles of sustainability to yield comprehensive and integrated approaches. As a concrete illustration of this idea, we have included a maturity model for SLR (MM4SLR) in the appendices of this work. MM4SLR sets forth a series of practices that can significantly enhance the quality of SLRs. These practices are strategically organized into different levels of maturity, in doing so, the MM4SLR takes into account social, economic, and technical aspects, and it outlines a pragmatic and gradual pathway to ensure the adoption of practices that can improve the sustainability of SLR.

From a long-term perspective, we believe that sustainability can also act as a catalyst for SE research, enabling researchers to produce more valuable SLR results for their target audience while saving time and effort, also minimizing barriers that hamper the replication and adaptation of existing SLRs. Our work serves as a foundation for discussing sustainability within the SLR context and sets the stage for the community to propose solutions that address current barriers without compromising any of the benefits provided by this method. While initiating this dialogue, we pave the way for collaborative efforts to overcome challenges and facilitate the advancement of sustainable SLR practices.

1.6 Overview of this Thesis

The thesis comprises a total of seven chapters, each serving a specific purpose. Chapters 2 to 5 delve into the research questions mentioned earlier, while Chapter 6 provides our conclusions and outlines future directions for further investigation. Chapters 2 to 5 are build upon scientific journal and conference papers in which, this Ph.D. candidate served as the lead author and collaborated closely with the research group in designing and conducting the studies.

Chapter 2 is based on a paper submitted to the ACM Computing Surveys (CSUR) and reports a tertiary study on the current understanding of sustainability of software systems.

Chapter 3 is based on a peer-reviewed paper published in the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). This chapter reports the main barriers to conducting systematic reviews based on empirical studies and experience reports in SE and introduces the concept of Sustainability for SLR (Sus4SLR).

Chapter 4 is based on a paper submitted to the Information and Software Technology (IST) Journal. This chapter adapts the characteristics of sustainable software into the SLR context using meta-ethnography to put together multiple visions to outline the sustainable SLR characteristics.

Chapter 5 is based on a paper that will be submitted to a high-impact journal soon. This chapter identifies high-quality SLR published in the past decade (2012-2022) leveraging points in their method to make the studies more sustainable.

Chapter 6 presents our conclusions, highlighting the main contributions of our work and future directions for research.

Appendices A, B, and C provide respectively supplementary materials for Chapters 2, 3, and 4, offering the pre-processed data generated during the conducted studies. This data serves as a valuable knowledge base and a comprehensive understanding of our thought process.

Appendix D illustrates a possible usage of sustainability in future research, i.e., it presents a new approach that was inspired by the sustainability concept that is a lightweight maturity model named “MM4SLR” which provides a pragmatic pathway for researchers to evolve their SLR throughout maturity levels.

CHAPTER

2

Background

DOS SANTOS, V.; IWAZAKI, A.Y.; VENTERS, C.C; FELIZARDO, K.R; SOUZA, E.F.; NAKAGAWA, E.Y.; Sustainable Software Systems: Core Characteristics and Trends. **ACM Computing Surveys** (CSUR), 2023.

Status of this publication: this paper was **submitted** ACM computing surveys Journal and is under peer review.

Abstract

Context: Sustainability is a multi-disciplinary research topic that has drawn attention from various knowledge areas, such as ecology, engineering, healthcare, and the social sciences. The field of computing has also focused on exploring sustainability and how somehow to create sustainable software systems.

Problem: However, there is no consensus of what sustainable software systems are and what their current achievements are in supporting a more sustainable world.

Objective: The main contribution of this paper is to understand the core characteristics of sustainable software systems and the means to achieve them.

Method and Results: We conducted a tertiary study on sustainability in software systems and identified 242 studies, from which 24 were relevant for deep analysis. Our results identify 16 different perspectives of sustainable software systems and present evidence that sustainable software is currently concerned with the preserve of four resources (electric energy, human resources, monetary cost, and production time). As means to achieve sustainability in software systems, we outline 68 critical factors and 98 possible means (guidelines, tools, methods, metrics, models, and frameworks). We interpreted, selected, and grouped this multiplicity of initiatives to present the current research endeavors in this area and highlight current gaps that still need to be further investigated.

Conclusion: We conclude that sustainability must be treated as a core concern in software development together with both a better alignment between academia and industry and a change in the mindset of practitioners concerning sustainable system development.

2.1 Introduction

Software systems have underpinned all aspects of societal life, from commerce, communication, and education to energy, entertainment, finance, governance, and defense. Modern society's reliance on software systems has resulted in technologies being considered as the backbone of virtually all the solutions designed, including those to support sustainability (DEEK *et al.*, 2005). By 2040, the information and communications technology (ICT) sector is expected to account for 14% of the world's carbon footprint (BELKHIR; ELMELIGI, 2018). As a result, sustainability has emerged as a growing area of interest in software engineering (VENTERS *et al.*, 2017).

The concept of sustainability has drawn considerable attention from various fields, including social sciences, engineering, and computing, and more recently, it has been discussed in software engineering (HILTY *et al.*, 2011). In this context, sustainability of software systems proposes actions to better manage their construction and discuss how software can interact with sustainability issues, including energy consumption and resources management. A few clues over the definition of *sustainable software systems* were provided after some years of research, for instance, considering their capability of enduring over time and having potential for long-term maintenance (PENZENSTADLER; FLEISCHMANN, 2011; PENZENSTADLER *et al.*, 2012). At the same time, adapting the complex concept of sustainability itself to software systems is a difficult task, and several adjustments were needed to match the peculiarities of the software. For instance, while defining the sustainability dimensions, different views emerged. Penzenstadler and Femmer (2013) proposed five dimensions: social, environmental, economic, individual, and technical, which were later reaffirmed by other studies (BECKER *et al.*, 2015b; OJAMERUAYE *et al.*, 2016a). Meanwhile, in the GREENS workshop held at ICSE'13, two groups discussed divergent sets of dimensions (LAGO *et al.*, 2014; CALERO; PIATTINI, 2017): one acknowledging social, environmental, economic, and technical sustainability, and the second group proposed a model encompassing four perspectives, namely business, technical, environmental, and social. Subsequently, Lago *et al.* (2015a) incorporated four dimensions (social, environmental, technical, and economic) into a comprehensive framework tailored for software-quality sustainability requirements.

In software engineering, the *social dimension* ensures the same access to social resources (or personal assets, such as education, skills, and experience) for the current and future generations (LAGO *et al.*, 2015a; OMANN; SPANGENBERG, 2002), i.e., sustainability must support groups of people to exchange information in a reciprocal and fair manner (OMANN; SPAN-

GENBERG, 2002). *Economic dimension* preserves the economic value and capital (LAGO *et al.*, 2015a; BECKER *et al.*, 2015b; RAZAVIAN *et al.*, 2014a). This dimension is concerned with the stakeholders' long-term investments, high return on investment, and risks associated with economic aspects while preserving the environmental and natural resources (HILTY *et al.*, 2011). In a complementary way, the *environmental dimension* focuses on protecting the natural resources (e.g., water, air, land, minerals, etc.) consumed by activities or products, while improving the people's well-being (LAGO *et al.*, 2015a). The *individual dimension* refers to the well-being of humans as individuals, including mental and physical well-being, education, self-respect, skills, mobility, etc. (BECKER *et al.*, 2015b; NAZIR *et al.*, 2020b). Finally, *technical dimension* was defined in the computing area and addresses the long-term use of software-intensive systems and their appropriate evolution in a changing execution environment (LAGO *et al.*, 2015a).

Considering that software systems are the driving force of our modern society and it has direct impact in almost all sectors, the number of publications and researchers interested in aligning software development to sustainability goals is increasing (CALERO *et al.*, 2020). Sustainability in software development processes and software products became a common concern for many practitioners because their entire life cycle consumes various resources (e.g., human efforts (RASHID; KHAN, 2018), monetary costs (WOLFRAM *et al.*, 2017), computer processing (RASHID; KHAN, 2017; RASHID; KHAN, 2018), electricity (RASHID; KHAN, 2018)). To preserve these resources, sustainability is considered a feasible solution that can lead the software industry to be aware of the impact of introducing new software on society. Over the years, several approaches were proposed by researchers addressing software sustainability problems, for instance, the creation of systems that consume less energy (MARIMUTHU; CHANDRASEKARAN, 2017; AHMAD *et al.*, 2014) or reduction of environmental impact (MOURÃO *et al.*, 2018; ANWAR; PFAHL, 2017). Thus, it is fair to recognize the progress of software engineering community in creating awareness about the importance of sustainability and adapting current methods to fit sustainability goals and make them more eco-friendly.

Despite all existing studies, classifying software as sustainable is a difficult task (VENTERS *et al.*, 2014) because many perspectives must be considered in this analysis (e.g., energy efficiency, social and individual aspects (NAZIR *et al.*, 2020b), economy (BECKER *et al.*, 2015b), etc.). In addition, software products have indirect effects on the physical world (KERN *et al.*, 2018) that must also be considered while evaluating the actual impacts of software on society. In this context, a real-world problem is that the information about sustainable software is fragmented in multiple venues and papers, making it difficult for practitioners and researchers of different areas (including computer science and ecology) to comprehend clearly about what means dealing with sustainable software systems. For this, it is essential to understand *what* sustainable systems are, *which* their characteristics are, *how* to develop these systems, and *which* the actual impacts of developing such systems on society are.

Motivated by this scenario, this chapter presents a holistic view of the core characteristics of sustainable software systems and the means to achieve them. To do that, we systematically inquired the literature, and a total of 301 studies were found and 24 were selected to be deeply analyzed and served as the main sources for our findings. A total of 16 different perspectives of sustainable software systems were found and analyzed to present a more integrated perspective of what is exactly a sustainable system. We also analyzed which are the main resources consumed during software construction to highlight the software development impacts on society, economy, and environment. Besides, we identified 68 critical factors for building sustainable systems, and 98 development approaches, including guidelines, methods, tools, metrics, models, and frameworks that could support the development and maintenance of these systems.

This study intends to make the software engineering community (practitioners and researchers) aware of the importance and necessity of considering sustainability when developing software systems; hence, this work can have implications for researchers and practitioners:

- **Implications for researchers**: By being the first overall view of sustainable software systems and the several elements surrounding them (such as practices, resources, critical factors), this work also makes it possible to foresee many research opportunities still open.

- **Implications for practitioners**: By presenting a compilation of information about sustainable software in a unique source, practitioners can more easily improve their comprehension about it. Hence, this work can also make practitioners aware of the need to change their mindset and practices incorporating sustainability in software projects when intending to achieve success concerning lower maintenance costs and efforts and long-term existence of software systems, in particular, those large and integrated found in critical application domains, as health, transportation, and communication.

The remainder of the article is organized as follows: Section 2.2 details the research method used in this work; Section 2.3 is devoted to the holistic view of sustainable software systems; Section 2.4 summarizes the main findings, future perspectives, and the threats to validity of our study as well; finally, Section 2.5 provides the final remarks.

2.2 Research Method

For systematizing our research method, we conducted a tertiary study (KITCHENHAM *et al.*, 2015) that refers to a review considering secondary studies (i.e., systematic literature review (SLR) and systematic mapping (SM)) related to a given topic. In turn, secondary studies make it possible to build meaningful summaries of evidence (from primary studies) in a given topic by using a systematic and transparent process (i.e., auditable and repeatable) (KITCHENHAM *et al.*, 2015; MENDES *et al.*, 2020a), aggregating scientific value to literature reviews.

Tertiary studies have already been conducted in software engineering (HODA *et al.*, 2017; KITCHENHAM *et al.*, 2010) following a similar rigorous process like in secondary studies, while they answer broader research questions. They are mainly conducted when various secondary studies are already available on a given topic, as the case of sustainable software systems (GARCIA-MIRELES, 2017; PENZENSTADLER *et al.*, 2012; PENZENSTADLER *et al.*, 2014; RASHID; KHAN, 2018). Hence, the conduction of a tertiary study on sustainable software is timely, putting together possibly all secondary studies and systematically synthesizing the most relevant information. To the best of our knowledge, there are no tertiary studies to address broader research inquiries on sustainable software.

In this chapter, we answer the *RQ1 - What is the state of the art of sustainability in Software Engineering domains?*. For this, the three main phases of the process for tertiary studies (namely planning, conduction, and results synthesis (KITCHENHAM *et al.*, 2004a)), as shown in Figure 2 and summarized in the next sections.

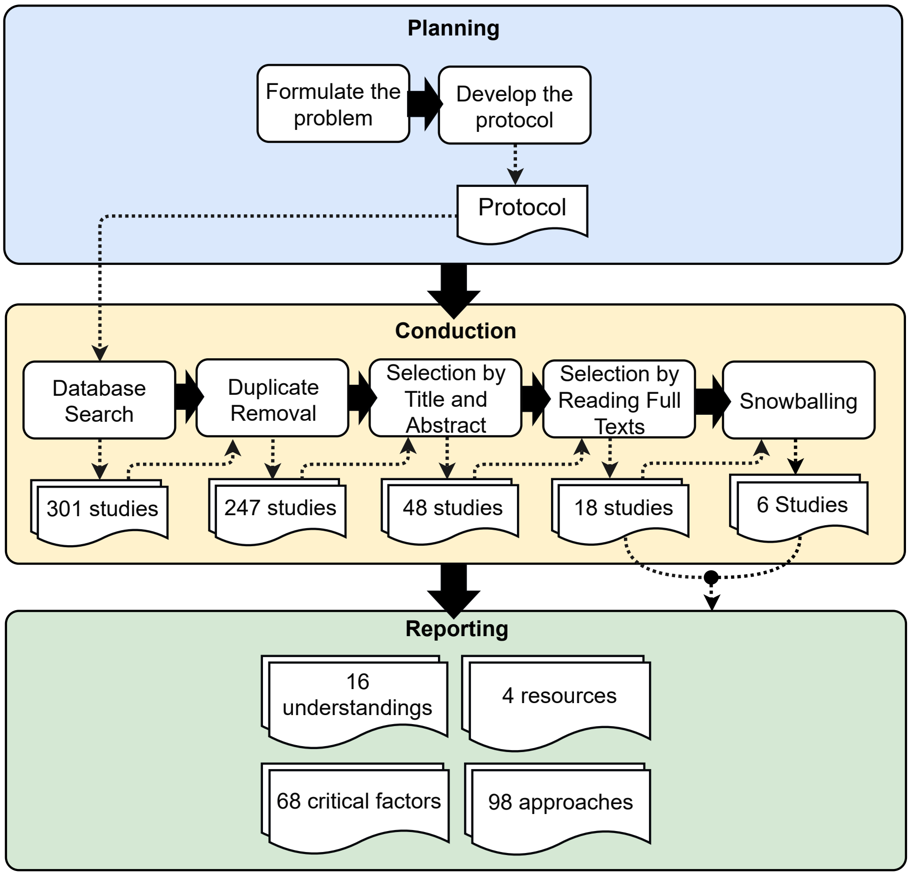

Figure 2 – Research method used to the conduction of the tertiary study

2.2.1 Planning

More specifically, we defined four research questions (RQ), and their rationales were established to identify evidence and depict the holistic view of sustainable software systems:

- *RQ1.1:* What are sustainable software systems?
  Rationale: Various understandings of sustainable software systems exist, and the identification of their commonalities and differences and the characteristics of such software could pave the way to establish a consensual definition of sustainable software systems.

- *RQ1.2:* Which are the resources taken into account when considering sustainability in software systems?
  Rationale: Software life cycle (development, evolution, and usage) consumes multiple resources, for instance, energy, computational resources (e.g., infrastructure hardware, internet connectivity, storage), and human resources (e.g., professionals' knowledge, effort, and time). While preserving resources is the main pillar of sustainability, it is not clear *which* are the resources taken into consideration when dealing with sustainable software systems and which are the real-world impacts of preserving these resources.

- *RQ1.3:* Which are the critical factors for sustainable software systems?
  Rationale: Sustainable software systems and their development and evolution rely on various critical factors (e.g., creation of energy-efficient code, usage of green infrastructure, and dissemination of knowledge about sustainability goals), which may impact positively (or even negatively) the success of those systems. Hence, collecting, analyzing, and synthesizing all these factors could leverage the building of sustainable systems.

- *RQ1.4:* Which are the means to leverage the development of sustainable software systems?

  Rationale: The means (i.e., approaches in the context of software engineering area, including guidelines, tools, methods, metrics, models, and frameworks) that are adopted to develop software systems can directly impact their sustainability. This research question examines the existing approaches and find commonalities among them.

Regarding the search strategy, a search string was defined by considering the keywords "software engineering", "sustainability", and "systematic literature review", which were connected by the AND logical operator. Variations and synonyms of these keywords were connected by the OR operator. The terms in the search string were selected to cover a larger number of studies, as shown below:

("software engineering") AND ("sustainability" OR "sustainab" OR "green*" OR "ecolog*") AND ("systematic review" OR "literature review" OR "systematic mapping" OR "mapping study" OR "systematic map" OR "meta-analysis" OR "survey" OR "literature analysis")*

Four publication databases, namely Scopus[1], IEEE Xplore[2], ACM Digital Library[3], and Science Direct[4], retrieved studies. Such databases are considered effective, representative, and renowned for reviews in the software engineering area (DYBA *et al.*, 2005; NAPOLEÃO *et al.*, 2021). Furthermore, selection criteria were defined towards including studies that provided evidence for answering our four RQ and also excluding not relevant ones. The two Inclusion Criteria (IC) and five Exclusion Criteria (EC) are:

- IC_1: Study is a secondary study (i.e., SLR or SM).
- IC_2: Study addresses sustainability in software engineering.

- EC_1: Study is not a secondary study.
- EC_2: Study does not address sustainability in software engineering.
- EC_3: Study is written in a language other than English.
- EC_4: Study is a shorter version of another study already included.
- EC_5: Full text of the study is not available.

2.2.2 Conduction

An automatic search and snowballing technique were used to find relevant studies until August/2021. We re-executed our search process in April/2023 and no additional studies were identified. Figure 2 illustrates the process to identify studies[5]. A total of 301 studies were retrieved from databases; Scopus, IEEE Xplore, ACM Digital Library, and Science Direct returned respectively 236, 40, 21, and 4 studies. After removing duplicated, 247 studies remained. The title and abstract of each study were then read, and the application of the selection criteria resulted in 48 studies. Next, the full text of each study was read, and the application of the selection criteria resulted in 18 studies. One-depth backward snowballing (WOHLIN, 2014a) was applied to the studies selected in the previous step and identified six relevant studies, totaling 24 secondary studies.

Table 2 lists the selected studies, their ID, title, type of study (SLR or SM), number of primary studies addressed, and respective references. Overall, we observe these studies synthesized relevant information on sustainable software systems that was gathered from several primary studies.

[1] https://www.scopus.com
[2] https://ieeexplore.ieee.org
[3] https://dl.acm.org
[4] https://www.sciencedirect.com
[5] Details about the selection process and quality assessment are available in supplementary material: <https://doi.org/10.5281/zenodo.8086774>

Table 2 – Studies selected in the tertiary study

ID	Title	Type	# of PS	Pub. Year	Ref.
S1	Agile practices for global software development vendors in the development of green and sustainable software	SLR	53	2018	(RASHID; KHAN, 2018)
S2	Current challenges and conceptual model of green and sustainable software engineering	SM	97	2016	(KOMEIL *et al.*, 2016)
S3	Using agile methods for the development of green and sustainable software: Success factors for GSD vendors	SLR	80	2017	(RASHID; KHAN, 2017)
S4	Developing green and sustainable software: Success factors for vendors	SLR	74	2016	(SALAM; KHAN, 2016)
S5	Environmental sustainability in software process improvement: A systematic mapping study	SM	7	2017	(GARCIA-MIRELES, 2017)
S6	Green and sustainable software engineering - A systematic mapping study	SM	75	2018	(MOURÃO *et al.*, 2018)
S7	Practices for addressing environmental sustainability through requirements processes	SM	16	2018	(GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018)
S8	Practices of energy consumption for sustainable software engineering	SLR	23	2018	(MOISES *et al.*, 2018)
S9	Software engineering aspects of green and sustainable software: A systematic mapping study	SM	82	2017	(MARIMUTHU; CHANDRASEKARAN, 2017)
S10	Sustainability in software engineering	SM	168	2017	(WOLFRAM *et al.*, 2017)
S11	Sustainability in software engineering - A systematic mapping	SM	36	2017	(BERNTSEN *et al.*, 2017)
S12	Sustainability in software engineering: A systematic literature review	SLR	96	2012	(PENZENSTADLER *et al.*, 2012)
S13	Systematic literature review protocol for green software multi-sourcing with preliminary results	SM	74	2015	(SALAM; KHAN, 2015)
S14	Systematic mapping study on software engineering for sustainability (SE4S)	SM	83	2014	(PENZENSTADLER *et al.*, 2014)
S15	Towards greener software engineering using software analytics: A systematic mapping	SM	50	2017	(ANWAR; PFAHL, 2017)
S16	A systematic literature review on sustainability studies in software engineering	SLR	175	2014	(AHMAD *et al.*, 2014)
S17	A systematic literature review on green software metrics	SLR	23	2013	(BOZZELLI *et al.*, 2013)
S18	A systematic literature review for software sustainability measures	SLR	16	2013	(CALERO *et al.*, 2013)
S19	Interactions between environmental sustainability goals and software product quality: a mapping study	SM	66	2018	(GARCIA-MIRELES *et al.*, 2018)
S20	Motivators in green IT-outsourcing from vendor's perspective: A systematic literature review	SLR	82	2015	(KHAN *et al.*, 2015)
S21	Sustainability evaluation of software architectures: A systematic review	SLR	*	2011	(KOZIOLEK, 2011)
S22	Sustainable software engineering: A perspective of individual sustainability	SM	*	2020	(NAZIR *et al.*, 2020b)
S23	Situational factors for modern code review to support software engineers' sustainability	SLR	158	2020	(NAZIR *et al.*, 2020a)
S24	Green measurement metrics towards a sustainable software: A systematic literature review	SLR	14	2016	(DEBBARMA; CHANDRASEKARAN, 2016)

Legend: PS - Primary Studies | * - Not provided by the secondary study.

It is worth highlighting that we also conducted a quality assessment of the secondary studies aiming at verifying their relevance to be considered in our tertiary study. To do this, we adapted the quality criteria proposed by Zhou *et al.* (2015) and considered four main aspects: report, rigor, credibility, and relevance. In short, regarding the **report**, the secondary studies were evaluated to check if they clearly report the aims, problem, motivations, research questions, inclusion/exclusion criteria, data extraction/synthesis procedures, and whether they provided references (i.e., evidence) to ensure trustworthy results. In addition, studies were evaluated regarding their **rigour** in following systematically the review process, as well as the **credibility** in the results by verifying whether these studies provided sufficient data to support the findings, analyze their limitations/threats to validity, and assess the capability of their replicability (or auditability). Concerning the **relevance**, we verified whether studies presented relevant conclusions and implications to ensure their value for researchers or practitioners. Finally, after careful analysis, all studies were considered suitable to be included in our study we used a data extraction form to collect qualitative data, such as the understanding given to sustainable software systems, resources used, and critical factors, from the 24 secondary studies. When a given secondary study did not provide enough information, the primary studies reported in that study were systematically checked. Following this, we analyzed the qualitative data to answer the RQ; to do that, we performed a thematic synthesis (CRUZES; DYBA, 2011b).

2.3 Results

Figure 3 presents an overview of the 24 studies included in our tertiary study. Studies were published between 2011 and 2020, with a concentration between 2016 and 2018. This trend indicates the research topic of sustainability in SE is a new topic. Besides, the number of studies published in journals and conferences is quite similar, indicating that, in general, studies are somehow mature and suitable to be published in journals. Following, Sections 2.3.1 to 2.3.4 answer our four RQs.

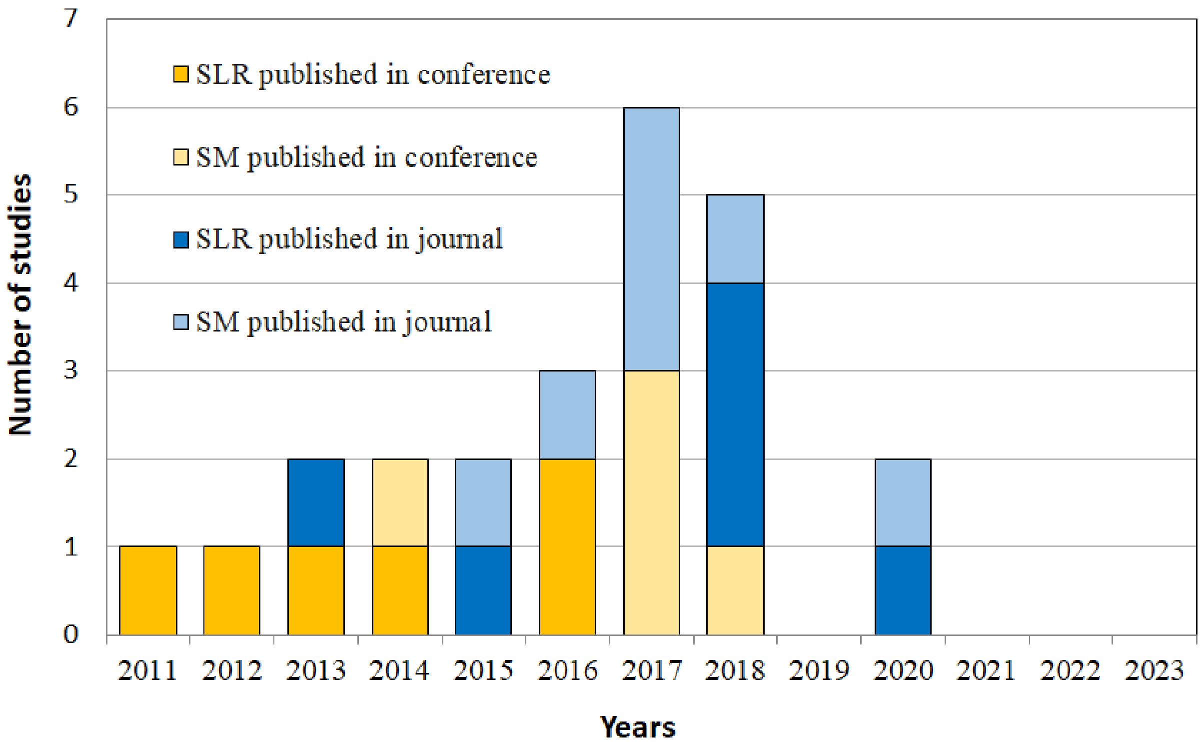

Figure 3 – Distribution of secondary studies along the years

2.3.1 Understandings of Sustainable Software Systems

This section answers RQ1 (What are sustainable software systems?). After examining the 24 studies, a majority (18 of 24) somehow presented understandings or definitions of sustainable software systems (details available in supplementary material). We observed that some studies present a common understanding, i.e., the need for producing software with minimal negative impacts on the economy, society, human beings, and environment. We also collected relevant particularities to create a holistic view of sustainability in software systems.

We also deeply analyzed each understanding and found the characteristics of sustainable software systems. Table 3 summarizes these characteristics and the studies where they were found. Overall, some characteristics are more recurrent than others, for instance, the *alignment to the economic, environmental, social, and individual sustainability* was found in 16 studies while *easy to learn software* is a characteristic found in just one study (S18). To better comprehend these characteristics, we categorized them into three perspectives: (i) those associated with the development process of sustainable software; (ii) those associated with the sustainable software itself; and (iii) those associated with the impacts caused by sustainable software.

Regarding the first perspective, i.e., those related to sustainable software development, they are completely aligned to the trends of current software development. Activities like *documentation* (S1, S11, S22, S23), *maintenance* (S12, S21, S22, S23), and *assessment* (S1, S11, S22, S23) should be performed in a continuous way (S1, S11, S12, S21, S22, S23). Hence, agile methods or iterative processes are examples of methodologies widely mentioned in studies and adopted to develop sustainable software (S1, S2, S3, S4, S5, S10, S11, S13, S20). Concerning the second perspective, we observe the software could become more sustainable when it considers sustainability as a quality attribute and implements requirements to comply with sustainability goals. Some quality attributes are directly connected to sustainability, namely reliability (i.e., *Reliable software* in S6 and S15), learnability (*Easy to learn software* in S18), accessibility (*Accessible software* in S18), reusability (*Reusable software* in S11, S12, and S18), and modifiability (*Modifiable software* in S1, S12, and S18). Other studies argue that considering environmental aspects is essential, like *energy-efficiency software* (S1, S9, S16, and S19) and *responsible use of resources* (S1, S2, S12, and S20), while designing, maintaining, or using software systems. Regarding the third perspective, we observe sustainable software (i.e., both the software itself and the processes to develop and evolve it) should be aligned to sustainability goals concerning economic, environmental, social/individual, and technical dimensions. Software engineers should be aware of the software impacts in a wide range of scenarios, including the direct impacts resulting from its development and indirect impacts from its interactions with society. Furthermore, we can observe that most of studies use external impacts of software to characterize sustainable software, this is an indicator that the current understanding of sustainability is tightly connected with the external impacts.

Table 3 also presents evidence that characteristics can also be connected by sustainability

Table 3 – Characteristics of sustainable software systems

Aspect	Characteristic	Similar Characteristics	Studies
Sustainable Development Process	Less development effort	• Essential software deliverable production	S1
	Continuous maintenance	• Conservation of software viability during maintenance or replacement • Cost-efficient maintenance	S12, S21, S22, S23
	Continuous documentation	• Non-extensive software documentation	S1, S11, S22, S23
	Sustainability-driven software development	• Alignment to sustainability goals	S8, S19
	Continuous evolution	*	S21
	Less e-waste (hardware and software discarded during software life cycle)	*	S4, S5
	Architecture quality improvement	*	S21
	Continuous assessment	*	S1, S11, S22, S23
	Software performance improvement	*	S20
Sustainable Software	Energy-efficient software	• Less energy consumption • Manage energy-intensive applications • Promotes energy efficiency	S1, S9, S16, S19
	Responsible use of resources (e.g., electric energy, computational resources)	• Efficient resource consumption • Profitable software	S1, S2, S12, S20
	Long-living software	• Satisfaction of current and long-term software goals	S1, S9, S6 S15, S16, S21, S22, S23
	Modifiable software	• Adaptable software • Polymorphic design	S1, S12, S18
	Reusable software	• Software continuously reused for improvements in software product	S11, S12, S18
	Easy to learn software (user-friendly software whose use is easily learned by users)	*	S18
	Accessible software	*	S18
	Reliable software	*	S6, S15
External Impacts	Alignment to the economic, environmental, social, and individual sustainability	• Minimal negative impacts on economy, society, human beings, and environment • Less impact on future generations capacity to fulfill their needs • Less environmental impact • Negligible effect on the economy, society, individuals, and environment • Minimal environmental impact • Positive impact on social and economic sustainability	S1, S2, S3, S4, S5, S6, S7, S9, S11, S12, S15,S14, S16, S18, S19, S20
	Positively impacts sustainable software production	• Positively effects the sustainable development • Software that leverages sustainability during its usage	S1, S3, S4, S5, S7, S11, S14, S18, S19, S20
	Software that leverages sustainability during its usage	-	S12

*** No similar characteristics found**

dimensions, since each characteristic could affect multiple dimensions at the same time with more or less impact. For instance, the characteristic *Energy-efficient software* is directly related to the economic, environmental, and technical dimensions; at the same time, it is not clearly stated it is about the individual/social dimension. Besides, a general observation is that sustainable software development, evolution, and execution are concerned with reducing the negative impact on all five dimensions. For instance, sustainable software can deeply impact environmental sustainability since our society has used software for diverse purposes (e.g., large-scale industrial software systems running on data centers or software used in smartphones for everyday tasks). In this scenario, even punctual optimizations in software widely used like web browsers or new methods to better manage the energy consumption of smartphones or data centers can considerably reduce the overall energy consumption (SHINGARI *et al.*, 2018).

Through our analysis, we observed sustainable software does not only refer to the software itself but also the entire scope that involves its development, evolution, and use. Hence, to define sustainable software, besides the software itself, it is necessary to consider the entire processes responsible for developing and evolving it and the impacts that its execution causes where it operates. Based on all its characteristics, we propose a possible definition of sustainable software as follows:

> *Sustainable software is a class of software whose development and evolution processes are driven by economic, environmental, individual/social, and technical sustainability concerns. It must comply with quality attributes throughout its life cycle, always using sustainability principles as a baseline to define its requirements. It must also positively impact the context where it operates while assuring minimal adverse effects on the environment, economy, society, individuals, and future human generations.*

This definition highlights the triple objective of sustainable software: (i) The software development must adhere to sustainability principles ensuring that sustainability is ingrained in every aspect of the software's creation; (ii) The software design must meet stringent quality requirements, with sustainability serving as guidance for defining these requirements; (iii) The software must have a positive impact on the context where it operates while mitigating potential negative effects. In essence, this definition underscores the importance of integrating sustainability into software development, emphasizing the alignment of software with sustainability principles, incorporating sustainability into requirements, and pursuing positive impacts of software.

The proposed definition also aligns seamlessly with the concept of *sustainability **in/by** software* (CALERO *et al.*, 2022). In particular, *sustainability **in** software* focuses on integrating sustainability into the software development process to develop software that adopt practices, such as optimizing software to become energy efficient, reducing resource consumption, minimizing waste generation, and prioritizing software durability and maintainability (CALERO *et al.*, 2015). *Sustainability **by** software* refers to the notion of utilizing software to promote sustainable practices in various domains. It leverages software to address environmental, social, and economic challenges and achieve sustainable outcomes, for example, optimizing building energy consumption, enabling efficient resource management, supporting renewable energy integration, facilitating eco-friendly transportation solutions, or enhancing sustainable supply chain management (CALERO *et al.*, 2015). Hence, our definition of sustainable software addresses both perspectives: sustainability in/by software.

Our definition is also aligned with the work of (BERKHOUT; HERTIN, 2001) that investigated the impacts of software and classified them into first-, second-, and third-order impacts. First-order impacts result from software production, including the use of resources, power consumption during software development, hardware production, and disposal of electronic equipment waste. Second-order impacts are associated with software usage, like resource conservation by optimizing processes (e.g., machine learning techniques to reduce resource consumption (SHINGARI *et al.*, 2018)) or by substituting material products with their immaterial counterparts (e.g., through reduction of paper usage). Third-order impacts refer to long-term effects on the environment that result from software usage (e.g., the change in the human lifestyle that promotes faster economic growth). Moreover, (GÖHRING, 2004) also described the same impacts but extended them to other dimensions besides the environmental issues.

In summary, the concept of "sustainable software" requires a definition that addresses the extensions of the impacts caused by bringing sustainability into software development. Our definition aims to cover different perspectives (also elicited from the literature) to support the software industry to move towards producing software using good practices and responsible use of resources and, as a consequence, reducing negative impacts.

2.3.2 Resources Taken Into Account in Sustainable Software Systems Development

This section answers RQ2 which specifically focuses on resources consumed when addressing sustainable software systems. To do this, we examined the 24 studies looking for mentions of resources considered a concern in software development. In supplementary material, we summarized the passages from each study and listed the resources addressed. After deeply analyzing these passages, we categorized the resources into three perspectives, as shown in Table 4. We can observe that the development, evolution, and execution of software systems can consume different types of resources simultaneously, for example, electric energy, human resources, or monetary investments to afford, for instance, cloud services. Moreover, software systems can directly impact the consumption of resources, e.g., smart cities or smart buildings can save energy by optimizing facilities and eliminating factors that promote energy waste.

Table 4 – Resources consumed during software life cycle (development, evolution, and execution) and resources impacted by sustainable software production

Perspective	Resources	Studies
Resources consumed during the **development** and **evolution** of sustainable software	Energy Consumption	S1, S3, S4, S9, S10, S12, S15, S16, S19
	Monetary Cost	S1, S3, S10, S12, S13, S17, S21
	Human Resources	S2, S3, S10, S12, S14, S16, S17
	Production Time	S1, S3, S13
Resources consumed during the **execution** of sustainable software	Energy Consumption	S8, S9, S10, S12, S19
	Monetary Cost	S8, S17, S21
	Human Resources	S17
Resource impacted by the sustainable software	Energy Consumption	S4, S5, S7, S9, S10, S11, S14, S17

In more detail, most studies (19 of 24) mention electric energy as a resource to be considered when developing, evolving, and executing sustainable software. Considering that most software systems have nowadays large-scale usage (e.g., mobile apps) or have required high-performance computers (e.g., data centers), practitioners must be aware of the impact caused by each decision in software projects, so requiring a sustainability-driven decision-making process aiming to produce energy-efficient software (ANWAR; PFAHL, 2017). When software systems aim to reduce carbon emissions in the atmosphere or energy consumption, like in electric cars and smart buildings, these systems can impact multiple dimensions of sustainability, including economic, social, and environmental.

Monetary cost is another resource mentioned (in 12 of 24) and is closely related to economic sustainability (KOMEIL *et al.*, 2016; CORBETT, 2010), which aims to maintain the capital and the added value (BECKER *et al.*, 2015b). Software projects naturally consume this resource for several issues, for example, in creating and maintaining development environments (software and hardware) or hiring skilled developers (S1 and S3). In this context, maintaining capital means that developers are aware of sustainability goals and consider this aspect to implement software that reduces the amount of computational resources (e.g., memory, CPU) (S1, S3). Thereby, sustainability-driven decisions can also reduce external impacts caused by software (e.g., e-waste created by the acquisition/substitution of hardware).

Building software is essentially an intellectual task and directly depends on human resources. Table 4 lists seven studies that mention this resource. Social and individual sustainability are tightly related to human aspects of software development: social sustainability aims to maintain social capital and preserve social groups in their entirety (KOMEIL *et al.*, 2016), while individual one focuses on preserving the human as individuals (e.g., health, education, skills, knowledge, leadership, and access to services) (NAZIR *et al.*, 2020b). Both dimensions must be carefully considered during software development as they are critical factors for humans.

Some studies (4 of 24) point out production time as another resource. For example, S1 and S3 mention that sustainable software must be more efficiently developed, to speed up its development. To do this, agile methods have been adopted, shifting the focus to working software instead of heavy documentation. By reducing the production time, another resource (monetary cost) could also be reduced.

We can also associate those resources with the definitions/understandings of sustainable software previously examined in RQ1. For instance, *electric energy* is mentioned in S9, S16, and S19 when the authors argue that software must *be energy efficient*, *consume less energy*, and *promote energy efficiency*. Besides, S2, S12, S16, and S21 refer to *monetary cost* when they mention that sustainable software *handles cost* or *can be cost-efficiently maintained and evolved over their entire life cycle*. *Human resource* is mentioned in S3 that points out that sustainable software *minimizes the negative impacts on human work*. Particularly, S12 and S18 mention that software should be *easily learnable and accessible, reusable, modifiable, and adaptable*, indicating that software should not be developed from scratch, consequently saving human resources, monetary cost, production time, and electric energy. *Production time* is also considered as a resource when S12 and S18 mention that *a sustainable software is a system that manages time and resources used*. Finally, it is necessary to consider that software projects have suffered from time-to-market pressure; consequently, production time has become an essential resource to be considered.

2.3.3 Critical Factors for Sustainable Software Systems

This section answers RQ3, which addresses those factors to be considered to leverage sustainable software systems. We synthesized 68 factors mentioned as the most critical in sustainable software, as listed in Table 5. Our main goal here is not to discuss the details of each factor but to weave connections to comprehend how these factors can impact sustainability. For this, we classified these factors according to the aspect they could impact directly: development process, resource usage, human aspects, and the software itself.

39 factors are associated with the development process, and many of which were already investigated in software engineering, even before sustainability emerged in the area. They can be considered **good practices** for the current software development and evolution, for example, requirement engineering, software continuous validation and integration, software testing, and software maintenance improvement. **Agile** methods also emerged before sustainability and have changed the practitioners' mindset to focus on working software, on individuals, and on collaboration among stakeholders, and have been aligned to industry needs for fast response to changes and fast delivery. Some agile principles are aligned with sustainability goals (as observed in Table 5), for example, reduction of unnecessary documentation (S1 and S3) and usage of iterative development (S5 and S13), which allow producing software and optimizing the resource consumption over its whole life cycle. Moreover, the software industry is complex, dynamic, and influenced by **business** decisions that can cause direct impacts on software sustainability. For example, excessive time-to-market pressure (product delivery pressure) (S13) and project management (S1, S3, S4, and S13) are critical factors that could directly impact many sustainability aspects.

According to the studies, we observed that adopting only traditional software engineering good practices seems not enough to achieve sustainability. Hence, software engineers and developers' teams should adapt their workflow to include **sustainability practices**. Five studies (S4, S8, S13, S16, and S22) present different practices. For instance, S16 suggests clearly defining sustainability goals and prioritizing actions that contribute to reducing resource consumption. Similarly, S22 points out the need to follow sustainable software guidelines and practices. Moreover, four studies (S4, S8, S13, and S19) recommend using specific **technologies**, such as requirements filtering tool (green evaluator) (S4 and S13), virtualized systems (S8), and cloud for software distribution (S4 and S13).

Human aspects are also considered a critical factor and refer to **social** factors (which preserve the community around the software) or **individual** factors (which preserve the human interests). Four studies (S1, S3, S22, and S23) present social factors that are tightly related to communication/collaboration (S1 and S3) and usage of reliable technologies for improving

Table 5 – Critical success/risk factors

Category	Sub-category	Critical Factor	References
Sustainable Development Process	Software Engineering good practices	• Usage of requirement engineering	S2, S4, S13
		• Improvement of software design	S4, S13
		• Usage of software continuous validation and integration	S1, S3
		• Improvement of software maintenance	S4, S13
		• Provide support for real-time and large systems	S3, S22
		• Improvement of software testing	S4, S13, S22, S23
		• Improvement of software delivery	S3
		• Improvement of project attributes and release management	S23
		• Usage of estimation strategies	S4, S13
		• Creation of software that support reusability	S3, S4, S13, S19
		• Creation of energy-efficient coding	S8
		• Improvement of graphical user interface	S4, S13
		• Improvement of processes and code	S3
		• Usage of minimal re-engineering	S3
		• Usage of refactoring techniques	S19
	Agile	• Improvement customer presence and knowledge	S3
		• Source code attributes and documentation changes improvement	S23
		• Reduction of unnecessary documentation	S1, S3
		• Creation of polymorphic designs	S1, S3
		• Usage of agile strategies	S4, S13
		• Usage of iterative development	S3
	Business	• Improvement of product life cycle and project management	S1, S3, S4, S13
		• Improvement of acquisition and use fallout assessment	S2
		• Creation of long-term planning	S3
		• Reduction of variable and unidentified situations	S22
		• Reduction of time-to-market pressure	S13
	Sustainable practices	• Usage of green infrastructure	S4, S13
		• Adaptation of sustainability practice	S22
		• Improvement of green application development environment	S4, S13
		• Reduction of hardware usage	S8
		• Definition sustainability goals	S16
		• Usage of sustainable software guidelines and practices	S22
	Technology adoption	• Usage of requirements filtering tool (green evaluator)	S4, S13
		• Usage of recommender systems and carbon footprints calculators	S19
		• Usage of virtualized systems	S8
		• Usage of design patterns	S19
		• Usage of program analysis techniques	S19
		• Usage of cloud for software distribution	S4, S13
		• Usage of mobile devices performance, compressing, memoization techniques, ads filter, energy-aware libraries	S19
Human aspects	Social	• Usage of technologies for communication	S22
		• Improvement of communication/collaboration	S1, S3, S23
		• Reduction of time pressure	S22
		• Reduction of management overhead	S3
		• Improvement of higher management support	S22
		• Dissemination of organization policies, practices, standards, and attributes	S23
	Individual	• Reduction of variation in productivity	S22
		• Improvement of motivation for sustainable development	S16
		• Improvement of satisfaction with work environment	S22
		• Improvement of competencies, skills, experience, productivity, and ethical behavior	S22
		• Dissemination of knowledge about sustainability	S22
			Continues...

Table 6 – Continuation of critical success/risk factors

Continuation of Table 5			
Category	**Sub-category**	**Critical Factor**	**References**
Resources Usage		• Improvement of efficient resources utilization	S1,S3,S4,S13,S23
		• Improvement of knowledge management	S22, S23
		• Usage of energy-saving libraries	S19
		• Reduction of monetary cost	S3, S16
		• Reduction of carbon emission	S4, S13
		• Reduction of e-waste generation	S3, S4, S13
		• Reduction of paper usage in communication	S4, S13
		• Reduction of energy consumption	S2, S4, S13, S19
Sustainable Software Features		• Flexibility	S3, S4, S13
		• Less security applications and encryption usage	S19
		• Long-life support (legacy systems support)	S4, S13
		• Effective usage of programming languages run-time scope of android-based systems, sorting algorithms, data structures	S19
		• Software quality enhancement	S3
		• Reduction of the usage of high-quality video level, and usage of video/program compression strategies	S19
		• Improvement of technology maturity and accessibility	S23
		• Reduction of the usage of non-sustainable libraries and frameworks, web server, web application features, cache	S19
		• Reduction of software defects	S23
		• Improvement adware detection in mobile systems	S19

the team's communication, preserving social aspects, and stating clear policies, practices, and standards for sustainability (S22). Two studies (S16 and S22) present individual factors intend to preserve the interests of developers, i.e., competencies, skills, experience, productivity, and ethical behavior, and create a healthy work environment and, consequently, reduction of variation in productivity.

We observed the efficient management of **resources usage** is a critical factor, while RQ2 identified which these resources are. A total of nine studies (S1, S2, S3, S4, S13, A16, S19, S22, and S23) presented eight different factors. All of them are somehow related to the resources identified in RQ2. In summary, these factors refer to the reduction of resources usage while improve the efficiency of their usage. For instance, S3, S4, S13, S16, and S19 argue reduction of monetary cost, carbon emission, e-waste, paper usage, and energy consumption are necessary.

We also noted that specific **software features** could impact how much that software is sustainable. Five studies (S3, S4, S13, S19, and S23) discuss such features, as listed in Table 5. For example, S19 argues that security and encryption features can consume extra amount of computing resources, possibly harming sustainable software. From another perspective, adware detection may have a positive impact because it eliminates additional resources consumed by malicious software. Still, according to S19, the energy-efficient data structures or sorting algorithms in Android-based systems are essential for ensuring battery autonomy. Hence, it is necessary to reduce the use of non-sustainable libraries, frameworks, services, and others, and leverage the adoption of sustainable features, e.g., good compression strategies for high-quality video.

Analyzing the critical factors aforementioned, it is clear that sustainable software development must be seen from an integrated view, considering different perspectives often dependent on them. Such development should consider the software process itself, the human aspects of software development, resources used, and even features implemented in the systems. We believe that software engineers and developers could produce more sustainable software systems by taking into account these factors.

2.3.4 Means to Leverage the Development of Sustainable Software Systems

This section answers RQ4, which explores the means to leverage the development of sustainable software systems. We identified 98 different approaches by examining each primary study referenced in all 24 secondary studies. In supplementary material, we present details of such approaches, i.e., their ID (i.e., A1 to A98), studies where they appear (S1 to S24), type of approach (i.e., methods, models, tools, metrics, guidelines, and frameworks), and the reference to the primary study that provides such approach. Figure 4 shows a categorization of these approaches according to their types: (i) **Guidelines** (they refer to good practices commonly accepted and suggested for software development and maintenance (KEHOE; JARVIS, 1996)); (ii) **Tools** (automated tools that support software development, evolution, and execution); (iii) **Methods** (systematic plan with a comprehensive description of its steps to achieve a specific goal); (iv) **Metrics** (means to measure aspects of software systems); (v) **Models** (abstraction of a system to better understand, communicate, or design such system (DORI, 2011)); (vi) **Frameworks** (comprehensive and logical structure, including definitions, guidelines, tools, metrics, and models and the interaction among them, to support a given activity). Observing the number of approaches in Figure 4, we can say researchers are compromised to investigate different means to achieve sustainable software development.

Regarding **guidelines**, 11 studies (S1, S2, S3, S7, S8, S11, S13, S16, S19, S21, and S22) provided a total of 258 different guidelines (sometimes referred as to lessons learned and best practices), after performing a careful meta-analysis considering several primary studies. Examples of guidelines are short sentences, such as "create mechanisms to reduce CPU energy consumption" and "Identify green deployment requirements for the system under development". Hence, to understand these guidelines better, we eliminated duplicates and applied thematic analysis to group those similar into 12 categories (detailed in supplementary material). In short, all guidelines support sustainable software development considering all sustainability dimensions. 36 guidelines address social and individual dimensions; for example, the importance of team communication (S1 and S22), preservation of the motivation of developers (S1 and S8), and means to perform knowledge management (S1, S7, S8, S21, and S22). For economic and technical dimensions, the guidelines highlight the importance of aligning the software development process with sustainability. For instance, 48 guidelines refer to the improvement of planning and design of software (S1, S7, S8, S13, S16, S19, S21, and S22) and 59 address the usage of metrics to evaluate software sustainability (S1, S2, S8, S19, and S21). The environmental dimension is mostly mentioned in the studies and can be associated with 42 guidelines that

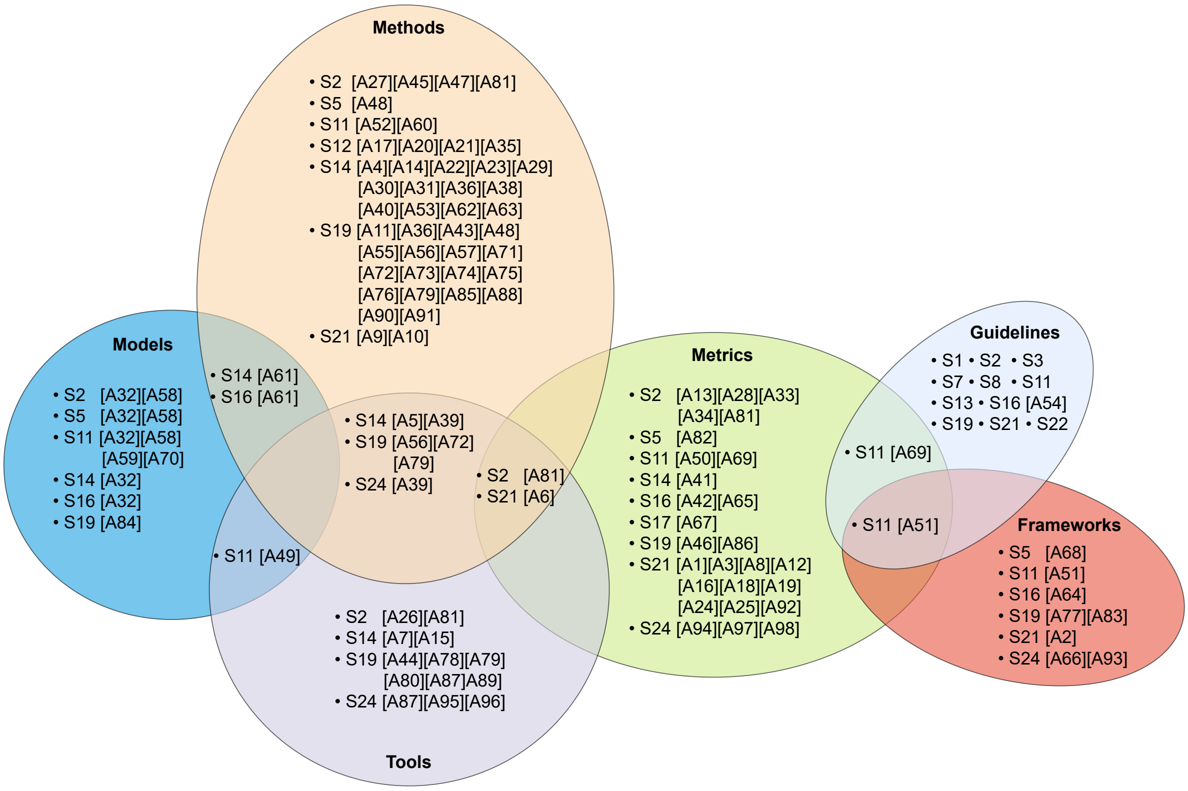

Figure 4 – Approaches to leverage the sustainable software development

provide directions to reduce energy consumption (S1, S2, S8, and S19) or optimize hardware utilization and avoid e-waste (S8).

Methods are the most common type of contributions totaling 47 methods. These methods propose solutions for several existing problems in software development but focus on sustainable software, for instance, assessing software architecture from a sustainability perspective (A9 and A10) or even modeling software sustainability requirements (A22, A36, A45, and A47). These methods are not only limited to the first-order effects but also propose solutions that consider third-order effects, for instance, methods to reduce the software impact on the environment, like the improvement of traffic and transportation in cities (A4 and A5) and improvement of manufacturing systems (A23 and A52).

Concerning the **models**, six studies (S2, S5, S11, S14, S16, and S19) provide seven different models (A32, A58, A49, A59, A61, A70, and A84). We observe these models were proposed after researchers achieved some maturity with sustainability in software. Results include models to provide a generic view of sustainable software (A70) or more specific models that show how sustainability can be applied in software development from its conception (A84) to its development (A58, A59, and A61) and deactivation (A32).

We also found eight different **frameworks** that are sometimes aggregative and discuss how models, tools, and metrics can be used in sustainable software development (e.g., A51 and A83). Nevertheless, some frameworks are focused on applying sustainability in specific software engineering tasks, for instance, design of software requirements that comply with sustainability goals (A66 and A77) or measurement of software sustainability (A2, A64, and A68).

Metrics for software sustainability are tightly associated with the resources depicted in RQ2. These metrics provide means to quantify aspects that are often too abstract and hard to comprehend. The main benefit of using metrics is to provide concrete means to appraise the extension of software impacts and better manage the actions to leverage sustainability. Our results showed that the most common metrics are closer to technical aspects of software development, for instance, analysis of source code or architecture to measure characteristics, like modularization (A1, A3, A6, A8, A12, A16, A18, A24, A25, and A92), quality (A13), structural complexity (A19), or stability (A25). Another way to measure sustainability refers to energy consumption (A33, A46, A50, A67, A86, A94, and A97) and quantifies the carbon footprint of software (A34 and A67). To support measurement, several **tools** were proposed to analyze software (A72) and to monitor sustainability indicators tracking energy-intensive software routines (A56, A72, A81, and A96)

Finally, the sustainability of software systems can be considered very recent compared to the ecology area. However, the number of approaches found gives us evidence that it is a trending topic. Most approaches were proposed recently to deal with the different perspectives of sustainable software systems.

2.4 Discussions

2.4.1 Main Findings and Perspectives for Future Work

The amount of more than 1,500 primary studies (from 24 secondary studies and concentrated in the early literature) proves that sustainability of software systems is a growing interest field. We observe an important movement of software engineering community starting a conversation with other science fields (e.g., ecology, economy, and sociology) to create software systems aligned with the sustainability goals concerning economic, environmental, social, individual, and technical dimensions. At the same time, the **definitions** and **understandings** of sustainable software systems are indeed still very diverse, fragmented, and multidisciplinary including those science fields and, as a consequence, making difficult to reach a consensus or a common understanding of sustainable software systems. In this scenario, from the compilation of all studies examined in this work, we proposed a possible **definition of sustainable software systems** (previously presented in Section 2.3.1) from a wide view that encompasses the software itself, its development process, and the impact it causes, all of them aligned to the five sustainability dimensions (economic, environmental, social, and individual).

Most **characteristics of sustainable software systems** refer to those with effects of **first order** (i.e., those associated with the development and maintenance of software, e.g., human resources and time consumed by software projects) and **second order** (those related to software execution, e.g., electric energy consumed). However, those with **third-order effects** are less recurrent, naturally more complex, and often described in a generic way without pragmatic ways to comprehend and measure them, e.g., the impact of that software on human life. Solutions to preserve natural resources (e.g., creation of smart cities (KOLESNICHENKO *et al.*, 2021; SAKUMA *et al.*, 2021) and intelligent environment monitoring through data science (ALLEN *et al.*, 2021)) are increasingly found, similarly to the increasing number of studies; however, it is not clear if first-, second-, and third-order effects and the sustainability goals as a whole have been adequately considered in these studies. A deep and holistic analysis of these three orders indicates that the field of sustainability in software systems needs to evolve and consolidate. Therefore, for future work, it is essential to invest the effort to understand better the first- and second-order effects and mainly the third-order ones and the trade-offs among them, to achieve the required sustainable software systems. In other words, the research agenda must include the investigation of real software impacts on each sustainability dimension.

This work revealed almost all studies examined are aware that producing, maintaining, and executing software systems naturally consume different **resources**. These resources mainly refer to electric energy, production time, monetary cost, and human resources, and reducing their consumption is a goal from the sustainable software perspective. Many initiatives focus on analyzing and improving software systems to consume less power but sometimes ignore other aspects of sustainability. Hence, future work should investigate the connection between software and the environment that is surely not limited to energy consumption.

Closely interconnected with understanding of sustainable software and resources consumed throughout this software life cycle, there exists a diversity of **critical factors** that should be taken into account to realize the sustainability in software systems. We observe agile methods have been highlighted to develop sustainable software and it seems to be a trend to solve critical factors of the sustainable development process. These factors address different perspectives from technical to human, but are treated often separately in each study. Accordingly, the software engineering community should examine all these factors together and do a trade-off among them.

Analyzing the **means to leverage sustainability of software systems**, there is a multitude of approaches considering heterogeneous kinds of solutions; still, they are also fragmented, many of them without being extensively experimented with. Hence, no concrete trend can be defined. Therefore, a comparative analysis aiming at discovering those with the potential to deal better with sustainable software is welcome.

Our work makes us able to conclude that the next step to advance the area of software engineering is to incorporate sustainability as a "first-class citizen", i.e., *sustainability-driven software engineering*, not only improving the inside activities but also aggregating definitively the environment and the external and real impacts of software systems to the society into its practices.

2.4.2 Threats to Validity

Some potential threats that could have affected the validity of this tertiary study are:

Internal validity: During data extraction, some secondary studies omitted important details about the conduction, data extraction, and results summarization; hence, we had to interpret some pieces of information. To minimize this threat, we accessed the primary studies to gather more details and performed a double-check to improve our study reliability. In cases where information was not available, we used consensus meetings to reduce wrong interpretations of data.

External validity: A possible threat was to perform a wrong selection or miss important studies. To mitigate it, we systematically followed the review protocol (according to the guidelines of Kitchenham and Charters (2007)), which was previously validated with research group members specialists in secondary studies. In addition, we are confident that studies are representative to depict the state of the art of sustainability in software engineering. Possible distortions regarding the study's selection can have a minimal impact on our results.

Construct validity: A possible threat was to miss some studies because they do not use the same terms contained in our search string. To mitigate it, we performed several tests to calibrate our string and ensure the terms had good coverage. We also performed the backward snowball technique (JALALI; WOHLIN, 2012) by visiting each reference cited in the secondary studies.

Conclusion validity: A potential threat was the reliability of data extraction and summarization, since some data needed to be interpreted. To minimize it, we performed several brainstorming sessions to define the protocol elements (e.g., research questions and data extraction form). All authors of this work also reviewed the data extracted carefully and solved any doubts or conflicts in consensus meetings.

2.5 Final Remarks

Sustainability inaugurated a new era for software engineering, which now needs to rethink how to deal with software systems. All heterogeneous and important contributions already made need to be combined in a coordinated way under the umbrella of sustainability-driven software engineering. Software engineering has to mature much more to adequately encompass sustainability in all its dimensions and address the software impacts and ways to mitigate them. This paper contributes to this context by providing an overview of the achievements of the research community, putting together recent literature, and bringing to light important issues about the impact of software systems on society, the environment, and the economy. We believe in a further movement of software engineering to change how software systems are developed, evolved, and executed to primarily reduce the several types of resources that software consumes in their entire lifetime. More importantly, software engineering will be required to understand better the impacts (positive and negative) of software on other fields, such as ecology, economy, and society, and establish actions for reaching the required sustainable software systems.

CHAPTER

3

Towards Sustainability of Systematic Literature Reviews

DOS SANTOS, V.; IWAZAKI, A.Y.; FELIZARDO, K.R; SOUZA, E.F.; NAKAGAWA, E.Y.; Towards Sustainability of Systematic Literature Reviews. **Proceedings of the 15th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement**, 11 out. 2021. DOI: 10.1145/3475716.3484192

Status of this publication: this paper was **published** into 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement[1] and is available online in the following electronic address: <https://dl.acm.org/doi/10.1145/3475716.3484192>.

Abstract

Background: The software engineering community has increasingly conducted systematic literature reviews (SLR) as a means to summarize evidence from different studies and bring to light the state of the art of a given research topic. While SLR provide many benefits, they also present several problems with punctual solutions for some of them. However, two main problems still remain: the high time-/effort-consumption nature of SLR and the lack of an effective impact of SLR results in the industry, as initially expected for SLR.

Aims: The main goal of this paper is to introduce a new view — which we name Sustainability of SLR — on how to deal with SLR aiming at reducing those problems.

Method: We analyzed six reference studies published in the last decade to identify, group, and analyze the SLR problems and their interconnections. Based on such analysis, we proposed the view of Sustainability of SLR that intends to address these problems.

[1] https://conf.researchr.org/home/esem-2021

Results: The proposed view encompasses three dimensions (social, economic, and technical) that could become SLR more sustainable in the sense that the four major problems and 31 barriers (i.e., possible causes for those problems) that we identified could be mitigated.

Conclusions: The view of Sustainability of SLR intends to change the researchers' mindset to mitigate the inherent SLR problems and, as a consequence, achieve sustainable SLR, i.e., those that consume less time/effort to be conducted and updated with useful results for the industry.

3.1 Introduction

The community of software engineering (SE) has increasingly adopted Systematic Literature Review (SLR) in recent years (MENDES *et al.*, 2020a) as a reliable method to summarize evidence from a number of studies and find out the state of the art in a given research topic (FELIZARDO *et al.*, 2017; KITCHENHAM *et al.*, 2015). SLR has presented many important benefits, including the possibility of dealing with information from different studies in an unbiased manner (KITCHENHAM *et al.*, 2015; NIAZI, 2015), producing auditable and repeatable results (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a), and identifying research gaps and also perspectives for future investigations (KITCHENHAM *et al.*, 2015).

At the same time, SLR has also suffered from diverse problems. Studies reported some of them, e.g., poor documentation of SLR conducted (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019) and lack of rigor in following well-experimented guidelines for SLR directly impacting the SLR quality (KUHRMANN *et al.*, 2017). Important solutions have been proposed for specific problems, such as techniques for better planning SLR (CAIRO *et al.*, 2019; FELIZARDO *et al.*, 2017a), conducting SLR (FELIZARDO *et al.*, 2017b), reporting (CARTAXO *et al.*, 2018), and adopting supporting tools (MARSHALL; BRERETON, 2013). However, two main problems still remain with no suitable solutions: (i) the SLR conduction is still a very time- and effort-consuming task (FELIZARDO; CARVER, 2020), and (ii) the lack of effective impact of SLR results in the industry (BADAMPUDI *et al.*, 2019). This current scenario of SLR leads the SE community to question if it is worth continuing to invest in SLR conduction. Hence, the research question addressed in this work is: Is it possible to reduce the high consumption of resources (time and effort) while assuring to achieve of a more accessible and useful SLR for the industry? Another question that intrigues us is: Should the SLR researchers change their mindset on how to better deal with SLR conduction and dissemination?

The main goal of this work is to introduce a new view — which we name Sustainability of SLR — to better deal with SLR and reduce those two main problems (high resources consumption and lack of impact in industry) and, as a consequence, achieve sustainable SLR (i.e., SLR that consume fewer resources to be conducted and updated with useful results for the industry). To achieve our goal and answer the research question, we carefully examined six reference studies published in the last decade (AMPATZOGLOU *et al.*, 2019; BUDGEN *et al.*, 2018a;

IMTIAZ *et al.*, 2013; KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b; ZHOU *et al.*, 2016) to identify, group, and systematically analyze the main problems in both SLR conduction and update. We found four groups of problems and 31 barriers (i.e., possible causes of the problems). Following this, we proposed the view of the Sustainability of SLR, which connects these problems among them as well as their causes and encompasses three perspectives (social, economic, and technical) that could somehow deal with the SLR problems. The social perspective addresses human aspects, such as reviewers' communication and stakeholders' participation during the SLR conduction. The economic one is related to the resources (effort and time) to conduct and update an SLR. The technical perspective is related to the supporting tools and technologies used to conduct and update SLR.

It is worth highlighting that this perspective view does not intend to solve all current problems of SLR; instead, the main contribution of this work is to raise the awareness of researchers to change the mindset when proposing new solutions for SLR problems through a more holistic view.

The remainder of this chapter is organized as follows: Section 3.2 presents the research method. Section 3.3 presents the SLR problems identified in the literature and their possible causes; Section 3.4 presents the view of Sustainability of SLR; Section 3.5 discusses some challenges for achieving the Sustainability of SLR as well as the threats to validity; and Section 3.6 presents the final remarks.

3.2 Research Method

This chapter answers *RQ2 - How sustainability could be applied to the SLR context?*. Therefore, to propose the view of the Sustainability of SLR, it was necessary first to survey the literature and understand the current problems of both SLR conduction and update. For this, we selected six reference studies published over the last decade (2010 - 2020) (AMPATZOGLOU *et al.*, 2019; BUDGEN *et al.*, 2018a; ZHOU *et al.*, 2016; RIAZ *et al.*, 2010b; KITCHENHAM; BRERETON, 2013; IMTIAZ *et al.*, 2013) based on our group experience in conducting SLR. We also used experts' opinions to validate our choice and mainly avoid bias.

The six studies selected are tertiary studies (SLR of secondary studies) and experience reports from researchers' perspectives to identify the main problems of conducting and updating SLR. Riaz *et al.* (2010b) presented an experience report involving three Ph.D. students that identified the main difficulties of novices while they conduct SLR and compared the results with SLR expert experience. Imtiaz *et al.* (2013) conducted a tertiary study that evaluated 116 secondary studies published from 2005 to 2011, summarized the experience of researchers, and highlighted the challenges in conducting SLR. Kitchenham and Brereton (2013) conducted a tertiary study that assessed 68 secondary studies published from 2005 to 2012 to summarize the researchers' opinions about their experiences of conducting SLR and discuss techniques that could be used to improve the SLR process. Budgen *et al.* (2018a) conducted another tertiary study assessing 178 studies published from 2010 to 2015 summarizing lessons learned, researchers' main problems, and focusing on how good SLR could be reported.

We can also interpret threats to validity as problems in SLR that could not be solved providing clues about the researchers' difficulties during the SLR conduction. Hence, two tertiary studies (AMPATZOGLOU *et al.*, 2019; ZHOU *et al.*, 2016) identified and discussed the most common threats to validity and strategies to mitigate them. Zhou *et al.* (2016) assessed 316 studies published from 2004 to 2015 and Ampatzoglou *et al.* (2019) analyzed 165 studies published from 2007 to 2016.

Based on the six studies aforementioned, we performed the four main steps, as shown in Figure 5, which summarized the research method adopted.

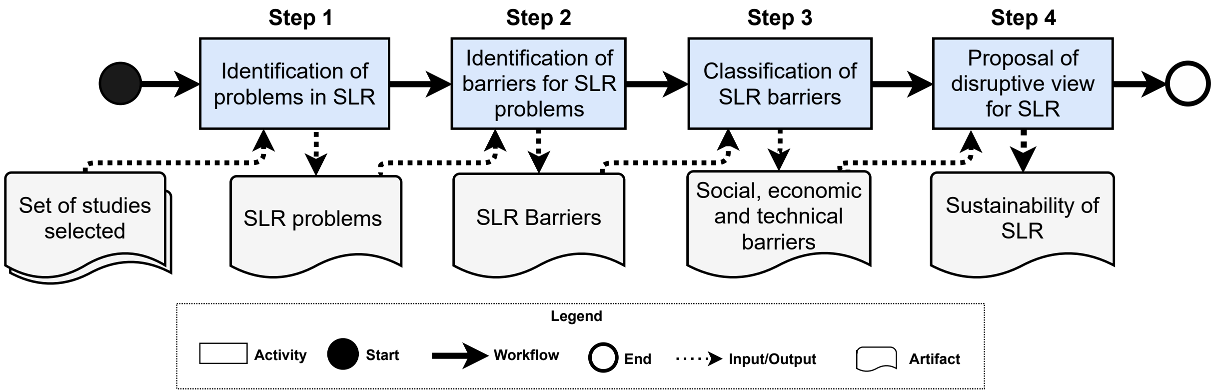

Figure 5 – Research method used to propose Sustainability for SLR

In **Step 1**, we systematically identified the main problems that occur during the SLR conduction and reported in the six studies. We also deeply checked the studies cited by these six studies. In **Step 2**, to understand the source of those problems, we carefully identified the current barriers, i.e., the possible causes for these problems, that could be hampering authors to conduct good and effective SLR. In **Step 3**, we deeply analyzed and summarized those barriers removing duplication and classifying them[2] according to three dimensions (or perspectives): social, economic, and technical. This classification was based on well-known sustainability dimensions already explored in other areas (BECKER *et al.*, 2015c; HILL; BOWEN, 1997; KEEBLE, 1988; PURVIS *et al.*, 2018a; SPANGENBERG, 2002; SOINI; BIRKELAND, 2014), as described below:

- *Social barriers* are mainly associated with human aspects of SLR (e.g., communication, culture) and how SLR results can impact the society (academia and industry);
- *Economic barriers* are directly associated with the high consumption of time and effort while planning, conducting, or updating SLR; and

[2] Replication package: <https://github.com/CSM-Research/06-LR-TSInSLR>

- *Technical barriers* address problems with tools and technologies used to support SLR conduction.

It is also worth highlighting that Steps 1, 2, and 3 were performed considering our long-term experience researching, conducting, and updating SLR. Besides, we based on evidence from the literature to make decisions. Finally, in **Step 4**, we analyzed the SLR problems and proposed a new view that connects these problems and their interactions as well as a possible means to deal with these problems. For this, we inspired ourselves on the views of sustainability from other areas, in particular, the ecology area (BROWN *et al.*, 1987) and SE area (BECKER *et al.*, 2015c), to derive the view of Sustainability of SLR.

3.3 SLR Problems and Barriers

After conducting Steps 1 to 3 of our research method, Table 7 summarizes the four main problems were identified: (i) lack of industry's interest in SLR results, (ii) SLR conduction is still very, time and effort consuming (iii) poor documentation of SLR, (iv) lack of SLR's quality. For each large problem, we identified the barriers (i.e., possible causes for these problems and listed in column 4 of Table 7) and classified them in social, economic, or technical dimensions (shown in column 3). A more detailed version containing the raw data processing is presented in Table 23 in supplementary materials.

For the first problem "*Lack of industry's interest in SLR results*", four social barriers were identified and one can be also interpreted as economic. Studies reported that SLR are conducted using inappropriate research questions (AMPATZOGLOU *et al.*, 2019; IMTIAZ *et al.*, 2013; KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b; ZHOU *et al.*, 2016) and highlighted that the cultural differences among researchers threats the validity of results (e.g., preferences for the studies of some researchers' nationality) (ZHOU *et al.*, 2016; RIAZ *et al.*, 2010b). In addition, the lack of a clear synthesis of SRL results and lack of practical recommendations for SE practitioners impact somehow the generalizability of the results and hamper the use of these results in the industry and academia (BUDGEN *et al.*, 2018a; AMPATZOGLOU *et al.*, 2019).

For the second problem "*SLR conduction/update is still very time- and effort-consuming*", seven barriers were found and all of them are economic barriers, while five of them can also be interpreted as technical barriers. A possible cause for this problem is the lack of standardization of terms to be used in the search string (as the case of SE area (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019; KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b; IMTIAZ *et al.*, 2013)) and as consequence, the search might return a large number of primary studies (including many irrelevant ones) or miss some relevant studies (AMPATZOGLOU *et al.*, 2019). Another cause is the inefficiency of electronic databases (IMTIAZ *et al.*, 2013; RIAZ *et al.*, 2010b; KITCHENHAM; BRERETON, 2013) and, despite the improvements in many aspects over the years (e.g., usability, search engine, coverage, interface, etc), several studies still mention

Table 7 – Problems of SLR and possible causes

Prob.	ID	Dim.	Barriers (Possible causes)	Reference
Lack of industry's interest in SLR results	1	Eco/Soc	Data synthesis, interpretation, and presentation of results are damaged when researchers have different interpretations of data extracted or when are conducted by a single researcher.	(AMPATZOGLOU *et al.*, 2019) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	2	Soc	Cultural differences may impact the validity results	(RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	3	Soc	Inappropriate research questions are defined	(AMPATZOGLOU *et al.*, 2019) (IMTIAZ *et al.*, 2013) (KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	4	Soc	SLR does not present generalizable results which can be used in industry or other areas	(AMPATZOGLOU *et al.*, 2019) (BUDGEN *et al.*, 2018a)
SLR conduction/update is still very time- and effort-consuming	5	Eco	Lack of standardization and incorrect/incomplete keywords hinder the search string construction	(AMPATZOGLOU *et al.*, 2019) (IMTIAZ *et al.*, 2013) (KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	6	Eco	Inadequate number of studies are evaluated hindering the validity of results	(RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	7	Eco/Tech	Use of not credible or very specific/broad digital databases may return many irrelevant studies or miss relevant studies	(AMPATZOGLOU *et al.*, 2019) (IMTIAZ *et al.*, 2013) (KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	8	Eco/Tech	Primary studies may be duplicated in different databases	(ZHOU *et al.*, 2016)
	9	Eco/Tech	Inaccessibility of resources (papers/databases)	(ZHOU *et al.*, 2016)
	10	Eco/Tech	Database limitations and inefficiencies (e.g., interface, search string syntax)	(IMTIAZ *et al.*, 2013) (KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b)
	11	Eco/Tech	Need tool to support SLR	(IMTIAZ *et al.*, 2013)(KITCHENHAM; BRERETON, 2013)

Legend: Dim – Dimensions; Eco – Economic; Soc – Social; Tech – Technical **Continues...**

problems indicating that the choice of databases is still a challenge (AMPATZOGLOU *et al.*, 2019). Five studies mention that non-credible or very specific or broad databases can return many irrelevant studies or can even miss relevant studies, increasing the effort for conducting SLR (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019; KITCHENHAM; BRERETON, 2013; IMTIAZ *et al.*, 2013; RIAZ *et al.*, 2010b). In this context, many supporting tools were developed to mitigate such problem (MARSHALL; BRERETON, 2015), especially aiming at reducing the additional effort caused by database inefficiency or integration problems (e.g., studies indexed by multiple databases). However, many of these tools solve only punctual problems. Hence, the lack of reliable and integrated tools is still considered an important factor for excessive time and effort consumption.

For the third problem "*Poor documentation of SLR*", nine economic barriers were found and two of them can also be interpreted as social barriers. Results indicate a lack of

Table 8 – Continuation of problems of SLR and possible causes

Prob.	ID	Dim.	Barriers (Possible causes)	Reference
Poor documentation of SLR	12	Eco	Lack of information about the SLR conduction process (e.g, initial date, duplicated studies).	(BUDGEN *et al.*, 2018a) (KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b)
	13	Eco	Lack information about the use of inclusion/exclusion criteria.	(BUDGEN *et al.*, 2018a) (KITCHENHAM; BRERETON, 2013)
	14	Eco	Insufficient details/report about the quality evaluation	(BUDGEN *et al.*, 2018a)
	15	Eco	Additional search is not being reported correctly	(BUDGEN *et al.*, 2018a)
	16	Eco	Lack of details about the primary studies selected (e.g. context, participants, source material)	(BUDGEN *et al.*, 2018a)
	17	Eco	Lack of information about the exclusion of studies impacting SLR data extraction and synthesis	(BUDGEN *et al.*, 2018a)
	18	Eco	Studies do not specify important SLR details, causing problems to repeatability and replicability	(AMPATZOGLOU *et al.*, 2019) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	19	Eco/Soc	The synthesis process is not clear	(BUDGEN *et al.*, 2018a) (RIAZ *et al.*, 2010b)
	20	Eco/Soc	Lack information about the reviewers' participation during the SLR conduction process	(BUDGEN *et al.*, 2018a) (IMTIAZ *et al.*, 2013) (RIAZ *et al.*, 2010b)
Lack of quality of SLR	21	Eco	Many studies do not perform the quality evaluation or do not use it correctly in selection process.	(BUDGEN *et al.*, 2018a) (RIAZ *et al.*, 2010b)
	22	Eco	Few studies had a good search coverage.	(BUDGEN *et al.*, 2018a)
	23	Eco	Inappropriate search methods leading to problems in SLR coverage	(IMTIAZ *et al.*, 2013) (ZHOU *et al.*, 2016)
	24	Eco	Selection process can be difficult when inclusion/exclusion criteria are generic or inappropriate.	(AMPATZOGLOU *et al.*, 2019) (ZHOU *et al.*, 2016)
	25	Eco	The wrong classification of the primary studies may cause the secondary study to lack robustness.	(AMPATZOGLOU *et al.*, 2019) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	26	Eco	Delimiting a time span affect the coverage of SLR	(ZHOU *et al.*, 2016)
	27	Eco	Lack of impartiality of researchers resulting in a bias in study selection	(ZHOU *et al.*, 2016)
	28	Eco	Extraction process is difficult when quality assessment is biased by reviewer subjectivity or they do not completely understand the data extraction items	(AMPATZOGLOU *et al.*, 2019) (ZHOU *et al.*, 2016)
	29	Soc	Lack of expert evaluation of the results	(KITCHENHAM; BRERETON, 2013) (RIAZ *et al.*, 2010b) (ZHOU *et al.*, 2016)
	30	Eco	Data model and data extraction forms may change during extraction	(KITCHENHAM; BRERETON, 2013)
	31	Eco	Difficulties in deciding when to stop the piloting process	(RIAZ *et al.*, 2010b)

Legend: Dim – Dimensions; Eco – Economic; Soc – Social; Tech – Technical

commitment of researchers in following the well-experimented guidelines proposed for reporting SLR (BUDGEN *et al.*, 2018a). Three studies reported that SLR do not report important details about the conduction process that allows SLR to be updated or replicated (BUDGEN *et al.*, 2018a; KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b). The poor documentation directly impacts academia and industry by creating a lack of credibility and problems in auditability and reproducibility of SLR (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019; RIAZ *et al.*, 2010b). In addition, this problem extends to the social dimension since a clear data synthesis is considered very important to disseminate useful results of SLR to industry and also academia (BUDGEN *et al.*, 2018a).

For the fourth problem "*Lack of quality of SLR*", 10 economic barriers and one social barrier were found. The economic barriers indicate that authors have difficulty in following the guidelines proposed for SLR and ensuring the quality of selection (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019), data extraction (RIAZ *et al.*, 2010b; AMPATZOGLOU *et al.*, 2019), synthesis and documentation (BUDGEN *et al.*, 2018a). The wrong classification of the primary studies or application of inappropriate search methods, for example, causes problems in the SLR coverage and robustness. In addition, the lack of quality can also be considered a social problem because it is recommended that SLR include experts to evaluate the SLR results (KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b; ZHOU *et al.*, 2016), because when they are not involved in the SLR process, results cannot be adequately synthesized, even generalized, and useful for SE practitioners.

Observing from a broader perspective, all problems and their possible causes mentioned previously are somehow interconnected among them. For example, using the economic perspective, poor documentation hampers the auditability of SLR, resulting in a lack of confidence in the SLR results (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019; RIAZ *et al.*, 2010b). It also hampers the repeatability of SLR; hence, SLR cannot be easily updated (KITCHENHAM *et al.*, 2011), and researchers need sometimes to re-conduct the SLR from "scratch" and, as a consequence, consume extra resources (time and effort). This problem is also connected with a social problem because the low quality of SLR also directly impacts confidence in the results and generates doubts in readers about biases that could be occurred during the SLR conduction. Jointly to the lack of practical recommendations for SE practitioners in such documentation (BADAMPUDI *et al.*, 2019; BUDGEN *et al.*, 2018a), in many cases, SLR is not used source of evidence for the decision-making process (CARTAXO *et al.*, 2018), which could improve the productivity of industry and academia. Aiming to reduce the time and effort spent on the SLR conduction and deliver fast results, researchers have ignored important well-experimented guidelines (BUDGEN *et al.*, 2018a), which could ensure such quality, and have created poorly documented SLR with shallow results, which are not useful in the end.

Based on the set of problems, their dimensions, and barriers, we proposed a new view that could suggest to the researchers how to better deal with these problems in an integrated way, as presented in the next section.

3.4 Sustainability of Systematic Literature Review

Before presenting the view of Sustainability of SLR, we provide a brief overview of the topic of sustainability that has been already researched in other areas and has served as a basis for our work.

3.4.1 Brief Overview of Sustainability

Sustainability started in the ecology context (PURVIS *et al.*, 2018a) and became popular in 1987 when the UN World Commission on Environment and Development published the report "Our Common Future" (Brundtland Report) defining "sustainable development" as a development that meets the needs of the present without compromising the ability of future generations to meet their own needs (KEEBLE, 1988).

Next, computer science researchers realized that sustainability could be adapted to other contexts, such as SE (HILTY *et al.*, 2011). In 2009, the International Conference on Software Engineering (ICSE) started a special track called "software engineering for the planet"[3] discussing how software can interact with sustainability, including energy consumption and the impact of software production in society (HILTY *et al.*, 2011). In the next years, the term "Sustainable Software" was interpreted in two ways (PENZENSTADLER *et al.*, 2014): (i) the software code being sustainable; and (ii) the purpose of the software being to support sustainability goals, i.e., to improve the sustainability of humanity on our planet. Both interpretations must coincide in a software system that contributes to a more sustainable life, i.e., sustainable software is energy efficient, minimizes the environmental impact of the processes it supports and has a positive impact on social sustainability and/or economics. In 2015, the *Karlskrona manifesto*[4] was signed (BECKER *et al.*, 2015c), creating awareness of the SE community on the need for sustainability and becoming a non-functional requirement for software (RAISIAN *et al.*, 2016) and aiming to support software engineers to reduce unnecessary resource consumption.

To adequately address sustainability, it is usually broken down in different dimensions (PURVIS *et al.*, 2018a). The Brundtland Report (KEEBLE, 1988) described sustainability in ecology with three dimensions: social, economical, and environmental. Although these dimensions have become mainstream throughout the literature, they are not universal (PURVIS *et al.*, 2018a). Some authors consider necessary to have additional dimensions (e.g., institutional (SPANGENBERG, 2002), cultural (SOINI; BIRKELAND, 2014), or technical (HILL; BOWEN, 1997)). In the SE community, the Karlskrona Manifesto included the individual and technical dimensions to maintain human capital and ensure the longevity of information, systems, and infrastructure and their adequate evolution with changing surrounding conditions.

In summary, software construction has a well-defined process, consumes time and effort,

[3] <https://www.cs.uoregon.edu/events/icse2009/specialSessions/#planet>
[4] <http://www.sustainabilitydesign.org/>

uses human resources, and its success depends on efficient tools, human skills, and social aspects (e.g., participation of stakeholders). In this perspective, researchers in the SE area already understood that this problem must be solved using a systemic solution (BECKER *et al.*, 2015c). Coming from this context, we have used all knowledge previously accumulated and well-experimented in the environmental and SE area to propose a view of sustainability that could be valid to SLR.

3.4.2 Sustainability of SLR and Their Dimensions

Considering that the problems and barriers found in this work are associated with economic, social, and technical perspectives, we broke down the Sustainability of SLR into these three dimensions, as described below:

- **Social dimension:** it addresses the human aspects associated with SLR, such as communication and collaboration, and ensures effective participation of stakeholders in the SLR process. This dimension focuses on ensuring that current and future stakeholders (including both, academic community and SE practitioners) have open access to the SLR results and ensuring that these results could be easily used.
- **Economic dimension:** it preserves the resources (time and effort) during the conduction, audition, and update or replication of SLR by creating high-quality SLR, i.e., this dimension ensures high-quality documentation, credible databases, efficient search engines, and rigor in following the guidelines to preserve reproducibility and auditability, minimizing unnecessary efforts.
- **Technical dimension:** it preserves the reliability of supporting tools used to conduct SLR, facilitating their use for the update or reconduction of those SLR. This dimension must assure technical ways to minimize the efforts by automating tasks and supporting their execution.

Based on these three dimensions, the view of Sustainability of SLR can be understood as *“the process and a set of actions to make it possible to preserve SLR that endure over time (i.e., longevity) with less possible time and effort consumed and an effective impact to the industry.”*

Regarding other sustainability dimensions, we believe they are not directly related to SLR. For instance, the environmental dimension handles the impacts of human activities in natural resources (e.g., water, land, air) or, in SE, it manages the excessive amount of energy consumption caused by software. We did not include such a dimension because there is no clear connection between SLR and the environment. Therefore, for the while, we believe that the three dimensions seem to be sufficient to connect SLR problems and understand their connections and possible ways to achieve sustainable SLR.

Following this, we develop a discussion about how to better represent graphically the Sustainability of SLR. For this, we revisit studies in the literature from other areas (GIDDINGS *et* *al.*, 2002; KEEBLE, 1988; PURVIS *et al.*, 2018a) and observed that the sustainability dimensions are described as interconnected among them (PURVIS *et al.*, 2018a). Therefore, an important discussion in the sustainability context is how to make a fair graphical representation including all dimensions defined and how they interact with each other. The Brundtland Report (KEEBLE, 1988) proposed a representation (i.e., a Venn diagram with interlaced circles and the intersection between them) that became the most popular one for sustainability in the ecological area. The key idea is that human society is only sustainable if it can be sustained in all dimensions.

Figure 6 shows a graphical representation of the Sustainability of SLR using the Venn diagram. Figure 6.A represents the *Ballanced View* with an intersection of all dimensions. We can observe that attributing the same weight for all dimensions and balancing them to achieve sustainability may lead to questions about its feasibility, due to the large number of factors that impact each dimension. An alternative is the *Unballanced View* that considers the technical dimension as a common concern that cannot be dissociated from social and economic perspectives, as shown in Figure 6.C. We consider that both representations (shown in Figures 6.A and 6.C) present a "Weak Sustainability" (which was previously defined in (GUTÉS, 1996)) because solving all problems associated only with, for instance, social dimension does not assure to achieve a sustainable SLR.

In the opposite perspective, Giddings *et al.* (2002) proposed a different representation that considers that dimensions must be nested and with different weights. This representation assumes that it is not possible for the society and economy to develop outside the biosphere (MORANDÍN-AHUERMA *et al.*, 2019) and this perspective warns that a finite planet cannot sustain human life with an economy that intends unlimited growth. It is necessary to acknowledge that there are fundamental biophysical limits that constrain the natural resources on the planet (NEUMAYER, 2013). This perspective arranges the sustainability dimensions providing a hierarchic view, i.e., problems of each dimension must be treated by different disciplines and each decision must consider the constraints of each dimension (BECKER *et al.*, 2015c) and this representation was later called "Strong Sustainability" (MORANDÍN-AHUERMA *et al.*, 2019).

Figure 6.B presents the *Sociocentric View* that nests the dimensions considering the social dimension is the broadest and technical one is the most specific, so attributing different weights for them. This representation assumes technical problems are essentially connected with economic problems, and the social constraints must guide the decision-making of researchers while conducting SLR. Despite this representation seems to be fair for SLR, it may be a problem to put social priorities higher than economic needs. For example, the software industry (which is tightly associated with the social dimension) needs to speed up the SLR conduction; however, it may cause a lack of rigor in documentation and reliability that could directly impact the feasibility and value of SLR and, as a consequence, it may result in unsustainable SLR. Figure 6.D presents the *Econocentric View* that considers that social needs are constrained by economic aspects. Due to the resources constraints sometimes imposed by the economic dimension, researchers are

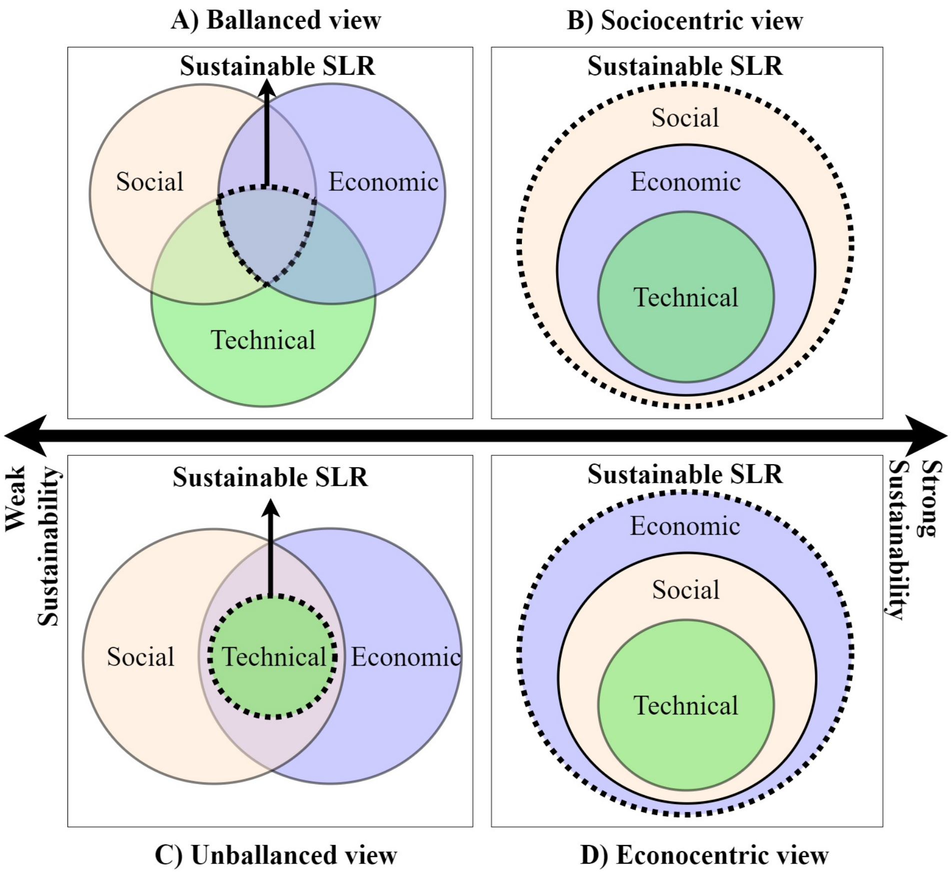

Figure 6 – Different views of Sustainability of SLR

prone to conduct SLR ignoring the social concerns. An example is not to include a given search question that could indeed answer the doubt from a wider community only to speed up the SLR process.

In the end, the graphical representations of Sustainability of SLR intend to provide us the opportunity for better understanding the inherent difficulties to create hierarchies among the dimensions and to better balance them. Considering the current scenario of SLR, we believe that it is necessary to move backward and recognize that the problems cannot be solved without integrated thinking.

3.5 Discussion

We believe that this chapter provides a good starting point and raises awareness of the community about the current problems of SLR from a holistic panorama. Many of the barriers have been mentioned since 2010 but not solved yet, in particular, the need for adequate synthesis of SLR results to be suitable for the industry. Our work indicates there is also a lack of: (i) knowledge about the SLR conduction process; (ii) application of the good practices for SLR conduction; (iii) details and important information to reproduce and update SLR; and (iv) tools' integration. We also observed many barriers found are related to the conduction process, while few social barriers are reported indicating a lack of knowledge about social problems.

We believe this chapter is provocative in the sense that it could bring a disruptive view and, as a consequence, new and coordinated actions, and solutions to make SLR reduce the high consumption of resources while producing useful results for the industry and academia. In this context, create effective methods to include the stakeholders (i.e., members of the industry) could be one of the possible solutions for social dimension problems. Economic dimension must create feasible methods to support the production of high-quality SLR consuming fewer resources (time and effort) to the conduction and update. For this, it is necessary to overcome many challenges including the researcher's mindset about the importance of correctly documenting SLR using the best practices available for reporting and include SLR sustainability as a quality criterion. The technical dimension must support tools and technologies used to plan/conduct/report aiming to ensure its quality, accessibility, and integration. Furthermore, this dimension must handle the slow and fragmented development of tools to support SLR and propose solutions towards integration. For this, it is necessary an international collaboration to provide long-term support and integration of SLR tools. Finally, the main challenge of Sustainability of SLR is to change the researchers' mindset about the importance of considering this integrated and holistic view while conducting SLR and proposing new solutions to overcome the current problems.

Regarding the threats to the validity, we did not perform an exhaustive search for studies that could have reported other problems and barriers in SLR. It is possible some problems and barriers have been missed. Besides, our work considered a 10-year period (2010-2020); hence, it is possible that problems and barriers reported before this period have not been considered. To mitigate these two threats, we used the experts' opinions to validate the set of studies considered in our analysis. Moreover, the results presented in this work could have been influenced by our knowledge and experience in researching and conducting SLR. To mitigate this threat, we strictly based on evidence from the literature to make any decisions.

3.6 Final Remarks

The academic community has increasingly conducted SLR, but they still suffer from critical problems, and the several punctual solutions being proposed for those problems have not been widely effective. This chapter is placed in this context and intends to somehow be provocative to call the attention of this community to change its mindset regarding the conduction and use of SLR results. Derived from the sustainability concept from other areas, the Sustainability of SLR primarily intends to achieve sustainable SLR, i.e., low consumption of resources (time and effort) and with useful results for the industry. As future work, the community needs to work together to promote sustainable SLR and, as a consequence, to achieve the initial goal of SLR that is the support of the industry.

CHAPTER

4

Sustainable Systematic Literature Reviews

DOS SANTOS, V.; IWAZAKI, A.Y.; FELIZARDO, K.R; SOUZA, E.F.; NAKAGAWA, E.Y.; Sustainable Systematic Literature Reviews. **Information and Software Technology**, 2023. DOI: 10.2139/ssrn.4401544

About the status of this publication: this paper was **submitted** into Information and Software Technology (ISSN: 0950-5849) and a pre-print is available online in the following electronic address: <http://dx.doi.org/10.2139/ssrn.4401544>.

Abstract

Context: Systematic Literature Reviews (SLR) have been recognized as an important research method for summarizing evidence in Software Engineering (SE). At the same time, SLR still presents several problems, such as the high resource consumption (mainly human resources) and lack of effective impact on SE practitioners, although much research has already been done.

Objective: The main goal of this paper is to explore the concept of sustainability in the SLR area, intending to contribute to understanding better and solving such problems in an integrated way. More specifically, this paper characterizes what sustainable SLR are, their core characteristics, critical factors (i.e., sensitive points in the SLR process), and guidelines for conducting such SLR.

Method: We performed a meta-ethnographic study to find key concepts of sustainable software systems and transpose them to sustainable SLR. For this, we systematically selected 16 studies about sustainable software systems and 14 distinguished studies about SLR. Following, we extracted the main keywords and metaphors, determined how both areas are correlated, and transposed them to obtain a set of core characteristics of sustainable SLR as well as critical factors and guidelines. Additionally, we validated them with specialists using the Delphi method.

Results: We found 15 core characteristics that offer a broad view of sustainable SLR, 15 critical factors in the SLR process that should be carefully addressed when conducting and updating SLR, and also 16 guidelines to manage SLR from the sustainability perspective.

Conclusion: The concept of sustainability in SLR can contribute to solving SLR problems in a more integrated way, while this work could change the mindset of the SLR community about the need to conduct sustainable SLR.

4.1 Introduction

Evidence-based Software Engineering (EBSE) has supported researchers and practitioners with research methods that create a body of useful knowledge in Software Engineering (SE) practices (FELIZARDO *et al.*, 2020b). EBSE techniques aggregate results from empirical studies, provide recommendations for practitioners, and, consequently, support decision-making processes in the software development industry (KITCHENHAM *et al.*, 2004a). In this scenario, Systematic Literature Review (SLR) is recognized as a reliable method that can provide a solid basis for practitioners and also researchers by synthesizing results from studies and drawing a comprehensive and methodical evaluation of the literature aiming to maximize coverage and minimize bias (KITCHENHAM *et al.*, 2004a; KITCHENHAM; CHARTERS, 2007). SLR can summarize evidence from primary studies (e.g., controlled experiments, case studies, surveys) and find out the state of the art in a given research topic (FELIZARDO *et al.*, 2017b; KITCHENHAM *et al.*, 2015). Three main phases compose the SLR process (KITCHENHAM *et al.*, 2015): (i) *planning* that defines a review protocol, including research questions, inclusion and exclusion criteria, sources of studies, search string, and study selection process; (ii) *conduction* that searches for studies in data sources and selects relevant studies to extract and synthesize data; (iii) *reporting* that answers research questions and reports results to disseminate them to the interested community (KITCHENHAM; BRERETON, 2013).

The literature documents several advantages of conducting SLR: (i) benefits for researchers, including the improvement of research skills (PEJCINOVIC, 2015; BORREGO *et al.*, 2015) and the possibility of learning from studies (BABAR; ZHANG, 2009; ZHANG; MUHAMMAD, 2012); and (ii) benefits for SE community, for instance, the possibility of dealing with information from different studies in an unbiased manner (KITCHENHAM *et al.*, 2015; NIAZI, 2015), production of auditable and repeatable results (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a), and identification of research gaps and perspectives for future investigations (KITCHENHAM *et al.*, 2015). At the same time, SLR suffers from diverse problems; some are caused by human factors, such as poor documentation of SLR (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019) and lack of rigor in following well-experimented guidelines for SLR impacting directly their quality (KUHRMANN *et al.*, 2017). Others are caused by the lack of maturity of the techniques or technologies that hamper the researcher's work, such as problems in electronic databases (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019; IMTIAZ *et al.*, 2013; KITCHENHAM; BRERETON, 2013; RIAZ *et al.*, 2010b) and lack of reliable and integrated tools with long-term support (FELIZARDO; CARVER, 2020). All these problems leaded researchers to propose solutions and adaptations to the SLR method. Important solutions addressed punctual SLR problems, such as means and techniques to aid search string construction (CAIRO *et al.*, 2019), minimize efforts while selecting primary studies (WATANABE *et al.*, 2020), automate SLR activities (FELIZARDO; CARVER, 2020), improve SLR protocol definition (FELIZARDO *et al.*, 2017a), and report and disseminate SLR results better (CARTAXO *et al.*, 2018). However, these solutions are often addressed separately while two major concerns in the SE community already remain: (i) SLR conduction consumes much time and effort from researchers and practitioners (KITCHENHAM *et al.*, 2004a; RIAZ *et al.*, 2010b; HASSLER *et al.*, 2014; MOURÃO *et al.*, 2020); and (ii) lack of an effective impact of SLR results in the industry (BADAMPUDI *et al.*, 2019).

In this scenario, the concept of *sustainability* was recently introduced to the SLR context as a disruptive view on the SLR problems and associated mitigation (SANTOS *et al.*, 2021). This view brings to light an integrated way to observe SLR problems, consequently, raising awareness about the importance of changing researchers' mindsets about the impacts of their decisions while planning, conducting, reporting, or updating an SLR. In short, sustainability in SLR is defined as "*the process and a set of actions to make it possible to preserve SLR that endure over time (i.e., longevity) with less possible time and effort consumed and an effective impact to the industry*" (SANTOS *et al.*, 2021). In particular, such sustainability addresses three main dimensions (SANTOS *et al.*, 2021): social, economic, and technical. These dimensions state that the current SLR problems need to be treated using multiple perspectives and integrated solutions. It is worth mentioning the concept of sustainability was first introduced in ecology to preserve natural resources and understand the impact of human actions on society, environment, and economy (PURVIS *et al.*, 2018b). Further, the SE community (including the software architecture) introduced this concept aiming to produce more efficient, less costly, and long-life software systems (MARIMUTHU; CHANDRASEKARAN, 2017; AHMAD *et al.*, 2014), also aiming to reduce the environmental impact of software production (MOURÃO *et al.*, 2018; ANWAR; PFAHL, 2017).

In short, the sustainability in SLR brings a new view on how to holistically deal with problems in SLR, including the high resource consumption (time and effort) and lack of impact of SLR results on academia and industry, and as a consequence, leveraging the conduction of sustainable SLR. However, it is still unclear what makes SLR sustainable, leading us to a general research question that guided this work: *What does it mean for an SLR to be sustainable*?

The main contribution of this chapter is to characterize what sustainable SLR are exactly by defining the core characteristics of these SLR, the critical factors (i.e., points in the SLR process) that impact on such SLR, and guidelines that can contribute to making them sustainable.

To do this, we applied meta-ethnography, a rigorous synthesis method that systematically collects, analyzes/interprets, synthesizes, juxtaposes, and transposes/translates concepts and characteristics from an origin context to a target one (NOBLIT; HARE, 1988; FU *et al.*, 2019). We selected *sustainability in SE* as the origin context, considering it is the closest context to SLR if compared to others (such as ecology or social sciences), besides the amount of knowledge already accumulated in this context. As a result, we created a brand new interpretation of sustainable SLR, i.e., their characteristics, critical factors, and guidelines associated to them. To evaluate our findings and mitigate subjectivity in data interpretation, we conducted semi-structured interviews with nine experienced researchers in the SE area using the Delphi method (NOBLIT; HARE, 1988). As a main result, we obtained 15 characteristics, 15 critical factors, and 16 guidelines. Finally, we believe that a better understanding of sustainable SLR could support SE community to deal with the several still existing problems of SLR (which are sometimes treated separately) and could raise awareness about the importance of considering the concept of sustainability when conducting, updating, or researching on SLR.

The rest of the chapter is organized as follows: Section 4.2 presents a brief background on sustainability in both SE and SLR and related works. Section 4.3 details the steps of the meta-ethnographic study that we conducted. Section 4.4 presents our results. Section 4.5 discusses the results, the threats to validity, and future work. Finally, Section 4.6 concludes this work.

4.2 Background and Related Work

Sustainability started to be discussed in ecology and become popular in the 1980s; its main concern was to handle environmental issues (PURVIS *et al.*, 2018b). The excessive consumption of natural resources accelerated by economic growth led political scientists to question the effects of human activities and better understand the limits of economic and demographic growth (PURVIS *et al.*, 2018b). The United Nations (UN) World Commission on Environment and Development published in 1987 the report "Our Common Future" (BROWN *et al.*, 1987), which defined that sustainable development aims to ensure humans' survival without jeopardizing the current and future generations. Sustainability is considered an integrative concept (HANSMANN *et al.*, 2012) commonly described using dimensions (also known as pillars, components, aspects, or perspectives (PURVIS *et al.*, 2018b)). The most common dimensions in ecology are environmental, economic, and social (PURVIS *et al.*, 2018b; HANSMANN *et al.*, 2012), each representing a group of different topics that must be encompassed while treating sustainability. In other words, responsible development must consider the natural, human, and economic capital, i.e., planet, people, and profits (HANSMANN *et al.*, 2012).

SE has incorporated this concept since software systems have a large impact on the current society and, therefore, can play an important role in global sustainability. The special session "Software Engineering for the Planet"[1] in the International Conference on Software Engineering (ICSE 2009) also contributed to awaring the community to the sustainability in SE. In 2015, "The Karlskrona Manifesto for Sustainability Design" (BECKER *et al.*, 2015a) described the challenges in developing sustainable software and the public major open issues, goals, values, and principles of sustainability for software (VENTERS *et al.*, 2015). At the same time, the SE community has then investigated the possible effects of such sustainability on software development and society.

SE also defined different dimensions to address the complexity of sustainability in SE. Razavian (RAZAVIAN *et al.*, 2014b) and Lago (RAZAVIAN *et al.*, 2014b; LAGO *et al.*, 2015b) mentioned four dimensions (social, environmental, technical, and economic), while Karlskrona Manifesto (BECKER *et al.*, 2015a) contributed with a fifth dimension (individual). *Social dimension* focuses on ensuring that current and future generations have the same or greater access to social resources by pursuing generation equity (LAGO *et al.*, 2015b). It encompasses the direct support of social communities in any domain, as well as activities or processes that indirectly create benefits for social communities. This dimension also considers political, organizational, and regulatory requirements and evaluates the ability of the software systems to maintain social capital preserving social communities (OJAMERUAYE *et al.*, 2016b). *Environmental dimension* addresses the natural resources and how to use them adequately. This dimension concerns in improving human welfare while protecting natural resources, such as water, land, air, and minerals (BECKER *et al.*, 2015a). The main concern is to analyze some aspects (e.g., the cost of energy to run algorithms) of software development that can impact environment (OJAMERUAYE *et al.*, 2016b), raw resources, climate change, food production, water, pollution, waste, etc (BECKER *et al.*, 2015a). *Technical dimension* addresses the long-term use of software systems and their appropriate evolution considering the constant changes in execution environment (RAZAVIAN *et al.*, 2014b; LAGO *et al.*, 2015b). This dimension refers to the longevity of information, systems, and infrastructure and their adequate evolution with changing surrounding conditions, including maintenance, innovation, obsolescence, data integrity, among others (BECKER *et al.*, 2015a). According to Venters *et al.* (2015), this dimension should be considered as a measure of the system's availability, integrity, maintainability, and reliability. *Economic dimension* focuses on preserving and maintaining capital and financial values (LAGO *et al.*, 2015b; BECKER *et al.*, 2015a). Razavian *et al.* (2014b) believes that economic sustainability is about creating a positive economic value exchange between all the stakeholders participating in software systems. Hence, this dimension should consider budget constraints, development/operational costs, market information when available, and long-term business goals that can relate to one or more functional and/or non-functional requirements (OJAMERUAYE *et al.*, 2016b). The *individual dimension* addresses the well-being of humans as individuals including their mental and physical well-being, education, self-respect, skills, mobility, etc.(BECKER *et al.*, 2015a). This dimension covers a wide range of topics mostly related to individual freedom and agency, human dignity,

[1] <https://www.cs.uoregon.edu/events/icse2009/specialSessions/#planet>

and fulfillment and includes individuals' ability to thrive, exercise their rights, and develop freely (PENZENSTADLER *et al.*, 2018). Finally, in the last decade, sustainability in SE and its dimensions have created a systemic and holistic vision that covers complex aspects and problems involving software production.

The conduction and update of SLR still present several problems that were identified by Carver *et al.* (2013). One of these problems is the high consumption of time and effort while executing SLR tasks (e.g., for study selection, data extraction, and quality assessment). Currently, this problem remains valid since it was reaffirmed by other more recent studies, e.g., (WATANABE *et al.*, 2020; FELIZARDO; CARVER, 2020). Another problem that remains is the lack of connection and impact on the industry needs and practices, focusing mainly on academic problems (SANTOS; SILVA, 2013; HASSLER *et al.*, 2014). Similarly, this problem was also reaffirmed by Cartaxo *et al.* (2020) which introduces the concept of rapid reviews, a lightweight secondary study focused on delivering useful evidence to practitioners in a timely manner.

Considering the problems aforementioned, the work of Santos *et al.* (2021) introduced the concept of *sustainability in SLR* as a new vision about how to deal with SLR, in particular, regarding the main barriers of SLR and current difficulties/threats to validity. This work encompassed the *social* barriers (mainly associated with human aspects of SLR, such as communication, and culture), *economic* barriers (associated with the high consumption of time and effort while planning, conducting, or updating SLR), and *technical* barriers (related with tools and technologies used to support SLR conduction). Associated with the sustainability dimensions (i.e., social, economic, and technical), these barriers have been mentioned since 2010, and most have not been solved yet. Hence, sustainability in SLR intends to make SLR achieve its original goal (i.e., the impact on society (academia and industry)) while reducing resource consumption. In turn, sustainability intends to become SLR sustainable; however, it is essential to define exactly what means sustainable SLR.

Regarding related work, several solutions have been proposed to the problems in SLR (e.g.,(FELIZARDO; CARVER, 2020; WATANABE *et al.*, 2020; CARTAXO *et al.*, 2018; MENDES *et al.*, 2020b; WOHLIN, 2016)); however, they have often focused on punctual problems without addressing a more encompassed solution. Hence, to the best of our knowledge, there are no studies that deal with SLR problems in an integrated way.

4.3 Research Method

This chapter answers *RQ3 - Which are the characteristics of sustainable SLR?*. This section describes the steps of our research method that adopted meta-ethnography (NOBLIT; HARE, 1988; FU *et al.*, 2019; CRUZES; DYBA, 2011b) as well as steps to the study's selection and validation of findings. In turn, meta-ethnography is a synthesis method already used in several studies of SE (FU *et al.*, 2019; CRUZES; DYBA, 2011b) and widely adopted in other disciplines, such as medicine/nursing (CAMPBELL *et al.*, 2011; BRIDGES *et al.*,), education (HUF; RAGGL, 2016; KAKOS; FRITZSCHE, 2016) and social sciences (DEWAELE *et al.*, 2021; HEAD *et al.*, 2016). It involves several steps, including the review of documents, observation, and interviews, allowing researchers to systematically understand how studies from different areas are related (NOBLIT; HARE, 1988). Researchers can also select, analyze, and interpret qualitative data through a process of translation, in which the studies are coded into metaphors, whose synthesis allows them to answer specific questions or provide new insights about a given topic (NOBLIT; HARE, 1988). The final product of the meta-ethnographic process is the translation of studies that goes beyond individual accounts to produce a new interpretation of a given topic (CRUZES; DYBA, 2011b).

Figure 7 depicts our complete research method[2]. In summary, the meta-ethnographic study encompassed Steps M.1 to M.7 (detailed in Sections 4.3.1 to 4.3.7, respectively). M.2 calls an auxiliary process composed of S.1 and S.2 to perform the studies selection (further detailed in Section 4.3.2). Another auxiliary process (composed of V.1 to V.3) was called by M.6 to validate our findings (detailed in Section 4.3.6) and, finally, the validated results were returned to the meta-ethnographic process (i.e., M.7). Following, these steps are described while Section 4.4 details all results.

[2] The supplementary material containing all artifacts generated throughout our research method is available at <http://dx.doi.org/10.17632/v7bbs5cx7p.1>.

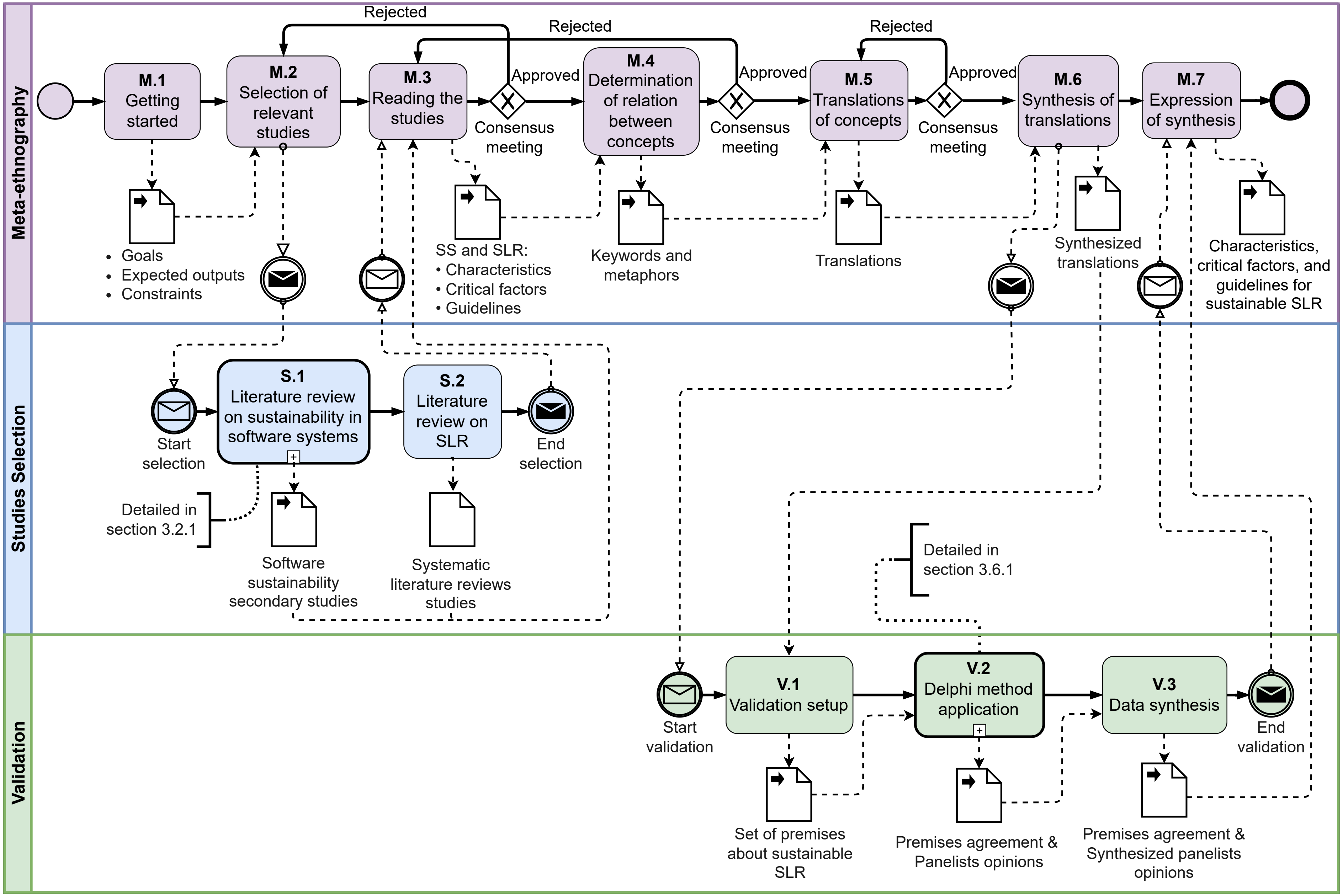

Figure 7 – Research method used in meta-ethnography, studies selection, and validation

4.3.1 M.1 - Getting Started

In this step, we conducted the planning sessions to define the main goal, expected results, and constraints/limitations of our meta-ethnography study. The *main goal* was to characterize what sustainable SLR are exactly. Our *expected results* were the key characteristics of sustainable SLR, critical factors, and guidelines/best practices to the conduction of these SLR. For this, we defined two *main constraints*: (i) we only considered SLR and excluded variations or adaptations, e.g., systematic mappings (PETERSEN; ALI, 2011) and multivocal reviews (GAROUSI *et al.*, 2019); and (ii) we considered only studies about sustainability in SE, i.e., sustainability in software systems, and excluded sustainability in other areas such as ecology (because it proposes solutions for a wide range of problems like environmental issues that do not have similarities with SLR). In turn, software construction and SLR conduction are similar regarding several points, e.g., there is a process for their construction/conduction, time and effort consumption, use of human resources, and their success depends on efficient supporting tools. Section 4.4.1 details these similarities.

4.3.2 M.2 - Selection of Relevant Studies

This step identified and selected studies from both areas (software systems and SLR). We collected studies of the state of the art of both sustainability in software systems (in Step S.1) and SLR (in Step S.2), as detailed below.

4.3.2.1 S.1 - Literature Review on Sustainability in Software Systems

To gather evidence of the current understanding of sustainability in software systems, we conducted a tertiary study. The decision of undertaking a tertiary study (over a secondary study) was due to several secondary studies already published in the area and taken in the problem definition step (S.1.1 in Figure 8). We also defined the protocol (in S.1.2) using the guidelines proposed in (KITCHENHAM *et al.*, 2015). Our protocol defined three research questions (RQ):

- *RQ3.1:* What are sustainable software systems?

  Rationale: Various understandings of sustainable software systems exist, and most of them highlight the main characteristics of these systems. This question intends to collect understandings/definitions/characteristics of sustainable software systems and pave the way to adapt them for SLR.

- *RQ3.2:* Which are the critical factors for sustainable software systems?

  Rationale: Several critical factors may impact positively or negatively the success of sustainable software systems. This question intends to identify such factors to be further juxtaposed to the critical factors of sustainable SLR.

- *RQ3.3:* Which guidelines can be applied to conduct more sustainable SLR?

  Rationale: Diverse guidelines have been proposed to leverage sustainable systems development; hence, we intend to adapt them to also leveraging sustainable SLR.

We also defined two inclusion criteria (IC) and five exclusion criteria (EC) to select studies:

1. IC_1: Study is a secondary study (i.e., SLR or SM).
2. IC_2: Study addresses sustainability in SE or provides critical factors or guidelines.

1. EC_1: Study is not a secondary study.
2. EC_2: Study does not address sustainability in SE or does not provide critical factors or guidelines.
3. EC_3: Study is written in a language other than English.
4. EC_4: Study is a shorter version of another already included.
5. EC_5: Full text of the study is not available.

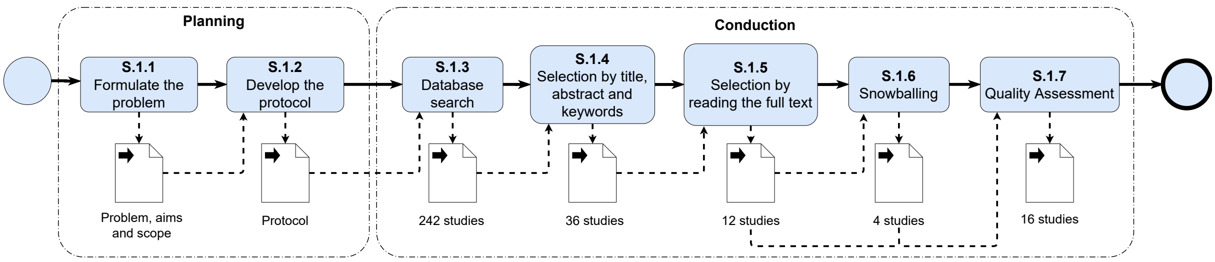

Figure 8 – Details of literature review on sustainability in software systems

In Step S.1.3, we inquired about the literature using the following search string: (*"software engineering") AND (" sustainability" OR "sustainab*" OR "green" OR "ecolog*") AND ("systematic review" OR "literature review" OR "systematic mapping" OR "systematic map" OR "meta-analysis" OR "survey" OR "literature analysis"*). The search was performed in four electronic databases — IEEE Xplore[3], ACM Digital Library[4], Science Direct[5], and Scopus[6] — and considering three metadata fields (title, abstract, and keywords).

A total of 242 studies returned during the search process. After eliminating duplicates and reading the title, abstract, and keywords, we applied the selection criteria resulting in 36

3 <https://ieeexplore.ieee.org/>
4 <https://dl.acm.org/>
5 <https://www.sciencedirect.com/>
6 <https://www.scopus.com/>

studies (in S.1.4). Next, we performed the full reading of each study and applied the selection criteria resulting in 12 studies (in S.1.5). We performed a backward snowballing (JALALI; WOHLIN, 2012) (in S.1.6) by revisiting the references from the 12 studies and identified other 15 studies. We applied the selection criteria in these studies firstly considering only the metadata and further reading the full study; finally, four studies were selected. Hence, a total of 16 studies (12 + 4) remained and are listed in Table 9.

Table 9 – Relevant studies about Sustainability in Software Systems

#	Ref.	Title	Year	Nº of primary studies	Type
SS1	(PENZENSTADLER *et al.*, 2012)	Sustainability in software engineering: A systematic literature review	2012	96	Journal
SS2	(CALERO *et al.*, 2013)	A systematic literature review for software sustainability measures	2013	16	Conference
SS3	(PENZENSTADLER *et al.*, 2014)	Systematic Mapping Study on for Sustainability (SE4S)	2014	83	Conference
SS4	(AHMAD *et al.*, 2014)	A systematic literature review on sustainability studies in software engineering	2014	175	Conference
SS5	(KHAN *et al.*, 2015)	Motivators in green IT-outsourcing from Vendor's perspective: A systematic literature review	2015	82	Journal
SS6	(SALAM; KHAN, 2016)	Developing green and sustainable software: Success factors for vendors	2016	74	Conference
SS7	(GARCIA-MIRELES, 2017)	Environmental Sustainability in Software Process Improvement: a Systematic Mapping Study	2017	7	Journal
SS8	(BERNTSEN *et al.*, 2017)	Sustainability in Software Engineering - A Systematic Mapping	2017	36	Conference
SS9	(ANWAR; PFAHL, 2017)	Towards Greener Software Engineering Using Software Analytics: A Systematic Mapping	2017	50	Conference
SS10	(MOISES *et al.*, 2018)	Practices of Energy Consumption for Sustainable Software Engineering	2018	23	Conference
SS11	(RASHID; KHAN, 2018)	Agile practices for global software development vendors in the development of green and sustainable software	2018	53	Journal
SS12	(MOURÃO *et al.*, 2018)	Green and Sustainable Software Engineering - a Systematic Mapping Study	2018	75	Conference
SS13	(GARCIA-MIRELES *et al.*, 2018)	Interactions between environmental sustainability goals and software product quality: A mapping study	2018	66	Journal
SS14	(GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018)	Practices for Addressing Environmental Sustainability through Requirements Processes	2018	16	Conference
SS15	(NAZIR *et al.*, 2020a)	Situational Factors for Modern Code Review to Support Software Engineers' Sustainability	2020	158	Journal
SS16	(NAZIR *et al.*, 2020b)	Sustainable Software Engineering:A Perspective of Individual Sustainability	2020	21	Journal

To evaluate the quality of these studies (in S.1.7), we adapted the quality appraisal instrument from Zhou et al. (ZHOU *et al.*, 2015), considering four main aspects: report, rigor, credibility, and relevance. Regarding *report*, we appraised whether studies clearly reported the aims, problem, motivations, research questions, inclusion/exclusion criteria, data extraction/synthesis procedures, and references (i.e., evidence) to ensure reliable results. Studies were also evaluated regarding their *rigor* in following systematically the review process, as well as the *credibility* in the results by verifying whether they provided sufficient data to support the findings, limitations/threats to validity, and replicability/auditability. Concerning the *relevance*, we verified whether studies presented relevant conclusions and implications to ensure their value for researchers or practitioners. After careful analysis, we considered all 16 studies suitable to proceed to the next step[7].

4.3.2.2 S.2 - Literature Review on SLR

To collect the SLR literature, we consulted SLR specialists and also considered our knowledge in the area. The study selection was conducted iteratively, i.e., we progressively identified studies that could provide evidence and could compose our final set. We first selected two well-cited, well-known, and classic studies (published as technical reports) that describe the SLR process and its characteristics (KITCHENHAM *et al.*, 2004a; KITCHENHAM; CHARTERS, 2007). After these publications, the SE community continued researching SLR in the following years, and several authors reported the lessons learned; hence, we selected two main ones (BUDGEN *et al.*, 2018a; BRERETON *et al.*, 2007). Additionally, we decided to complement our set of select studies with significant contributions to the SLR process. These studies address the SLR reliability (MACDONELL *et al.*, 2010a), repeatability (KITCHENHAM *et al.*, 2011), the most common threats to validity (AMPATZOGLOU *et al.*, 2020), tools (MARSHALL; BRERETON, 2013), improvement in the SLR update process (MENDES *et al.*, 2020b) and more recent guidelines (WOHLIN *et al.*, 2020). We also selected studies with relevant discussions that could relate to the three dimensions of sustainability in SLR. In particular, we found three studies in the social dimension that discussed strategies to manage knowledge and apply SLR results in practice (CARTAXO *et al.*, 2018; BADAMPUDI *et al.*, 2019; FABBRI *et al.*, 2013). For economic and technical dimensions, we included studies that proposed strategies to reduce the time and effort consumed in the SLR conduction/update (FELIZARDO; CARVER, 2020; MENDES *et al.*, 2020b). Hence, we included studies that discuss more recent topics that permeate the sustainability area. At the same time, these studies were evaluated regarding three aspects: (i) their relevance to SLR scenario (number of citations); (ii) their ability to capture the current state of the art of SLR method (type of study and publication year); and (iii) their relevance for sustainability context. Table 10 lists the final set of 14 studies selected in this step.

It is worth highlighting that meta-ethnography guidelines do not consider mandatory an exhaustive selection of studies (NOBLIT; HARE, 1988). Otherwise, the summarization process may become challenging when many studies are included and there is an amount of data to be processed and analyzed (FU *et al.*, 2019), consequently leading to a possible overload and quality loss. Hence, we decided that studies about SLR would not be exhaustively gathered.

[7] Details of the quality assessment are available at <http://dx.doi.org/10.17632/v7bbs5cx7p.1>

Table 10 – Relevant studies about Systematic Literature Reviews

#	Ref.	Title	Year	Type
SLR1	(KITCHENHAM *et al.*, 2004a)	Procedures for Performing Systematic Reviews	2004	Tech Report
SLR2	(KITCHENHAM; CHARTERS, 2007)	Guidelines for performing Systematic Literature Reviews in Software Engineering	2007	Tech Report
SLR3	(BRERETON *et al.*, 2007)	Lessons from applying the systematic literature review process within the software engineering domain	2007	Journal
SLR4	(MACDONELL *et al.*, 2010a)	How Reliable Are Systematic Reviews in Empirical Software Engineering?	2010	Journal
SLR5	(KITCHENHAM *et al.*, 2011)	Repeatability of systematic literature reviews	2011	Conference
SLR6	(FABBRI *et al.*, 2013)	Externalising tacit knowledge of the systematic review process	2013	Journal
SLR7	(MARSHALL; BRERETON, 2013)	Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study	2013	Conference
SLR8	(CARTAXO *et al.*, 2018)	The Role of Rapid Reviews in Supporting Decision-Making in Software Engineering Practice	2018	Conference
SLR9	(BUDGEN *et al.*, 2018a)	Reporting systematic reviews: Some lessons from a tertiary study	2018	Journal
SLR10	(BADAMPUDI *et al.*, 2019)	Contextualizing Research Evidence through Knowledge Translation in Software Engineering	2019	Conference
SLR11	(AMPATZOGLOU *et al.*, 2020)	Identifying, categorizing and mitigating threats to validity in software engineering secondary studies	2020	Journal
SLR12	(WOHLIN *et al.*, 2020)	Guidelines for the search strategy to update systematic literature reviews in software engineering	2020	Journal
SLR13	(FELIZARDO; CARVER, 2020)	Automating Systematic Literature Review	2020	Journal
SLR14	(MENDES *et al.*, 2020b)	When to update systematic literature reviews in software engineering	2020	Journal

4.3.3 M.3 - Reading the Studies

This step was responsible for systematizing the reading of studies of both areas and extracting text passages. For this, two authors of this work read individually the studies (in Tables 9 and 10) to discover/identify relevant text passages. Moreover, all authors discussed the terminology and concepts presented in those studies to rationalize the text passages associated with sustainability in software systems.

In short, studies about sustainability in software systems are quite recent and offer interesting particularities. Nine studies (AHMAD *et al.*, 2014; MOURÃO *et al.*, 2018; ANWAR; PFAHL, 2017; PENZENSTADLER *et al.*, 2012; CALERO *et al.*, 2013; PENZENSTADLER *et al.*, 2014; BERNTSEN *et al.*, 2017; MOISES *et al.*, 2018; GARCIA-MIRELES *et al.*, 2018) provide an overview of research activities, including research topics, limitations, main frameworks, methods, models, guidelines, and practices to support sustainability in software systems. Additionally, two studies Salam and Khan (2016) and Rashid and Khan (2018) focus on identifying critical factors and main practices to mitigate the risk associated with these factors. Other three studies aim to better understand sustainability by considering practical scenarios: Garcia-Mireles (2017) gather ideas about how sustainability can be integrated into software processes; Khan *et al.* (2015) collect the main factors that motivate the software industry to choose sustainable software, and García-Mireles and Villa-Martínez (2018) present the interaction between software quality and sustainability. Finally, two studies (NAZIR *et al.*, 2020a; NAZIR *et al.*, 2020b) address challenges of the individual dimension (one of the sustainability dimensions) and propose mitigation strategies and situational factors that hinder software engineers from creating sustainable software systems. After deeply analyzing each study, we extracted *71 text passages* that contained characteristics of sustainable software systems, critical factors, and guidelines to achieve sustainability in such systems. We also conducted group meetings to discuss whether such passages were relevant to the scope of this study. Appendix A, Table 24 presents the selected passages.

Concerning the studies on SLR, they offer important contributions by reporting best practices and recommendations for the SLR conduction as well as recent experiences and problems in SLR (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a; MENDES *et al.*, 2020b; BRERETON *et al.*, 2007; MACDONELL *et al.*, 2010a; AMPATZOGLOU *et al.*, 2020; WOHLIN *et al.*, 2020). Moreover, three studies address topics that are related to sustainability: the importance of managing and transferring knowledge gathered through SLR using a good documentation (FABBRI *et al.*, 2013), techniques to better externalize SLR results to a wider community (e.g., software industry) (CARTAXO *et al.*, 2018; BADAMPUDI *et al.*, 2019), and techniques to improve the SLR conduction using supporting tools (FELIZARDO; CARVER, 2020). After carefully examining these studies, we extracted *61 text passages* that contained information possibly associated to sustainable SLR. After that, we conducted group meetings to discuss the pertinence of those passages. Appendix B, Table 25 lists the selected passages.

4.3.4 M.4 - Determination of Relation Between Concepts

This step aimed to demonstrate that it is possible to correlate software systems and SLR regarding their sustainability. For this, this step was responsible for finding matches in the text passages of both areas by outlining correlations and similarities between them. For this, we applied Thematic Analysis (CRUZES; DYBA, 2011a) to extract a total of *72 codes* on sustainable software systems, additionally, we highlighted the main keywords and metaphors[8] used in text passages from both areas. Next, we grouped similar codes to facilitate data interpretation and removed duplications. For SLR, we extracted *56 codes*, also grouped those similarly and removed duplications. Following, we used keywords and metaphors identified to juxtapose the codes from both areas looking for the best connections between codes using baseline similarity to further depict which aspects of sustainable software systems and SLR are comparable. Section 4.4.1 details the correlation of 13 major codes (which we refer to *concepts* and are common in both areas) and how they can be interpreted in each area.

[8] A metaphor refers to a phrase used in a given area and could match with a keyword for which it does not correspond exactly but, based on empirical data and other evidence, it could be interpreted as similar to that keynote.

4.3.5 M.5 - Translation of Concepts

This step was responsible for creating an interpretation to the sustainability in SLR. For this, we applied reciprocal translations that consist in analyzing each code individually and generating a new interpretation (as recommended by Atkins *et al.* (2008)). Additionally, we used lines of arguments technique (NOBLIT; HARE, 1988) to combine our findings into a new piece of our theory, i.e., the understanding of sustainability in SLR.

Reciprocal translations are presented in Appendix C, D, and E (Tables 26, 27, and 28, respectively). These appendixes and tables respectively present translations to the *characteristics* for sustainable SLR, translations to the *critical factors* for sustainable SLR, and translations to the *guidelines* for sustainable SLR. In Table 26, the translations are structured in 1st order statements (text passages on sustainable software systems and SLR, found in the second and third columns, respectively) and 2nd order statements (translations, in the fourth column). In other words, the 1st order statements are a summary with one or more text passages that contains characteristics of sustainable software systems and SLR aligned throughout similarity. The 2nd order statements represent a possible interpretative adaptation to the SLR area. We proceeded similarly to Tables 27 and 28.

4.3.6 M.6 - Synthesis of Translations

This step was responsible for first removing noises (which could hinder the comprehension, such as duplications or keywords that lost their meaning during the translation process) and after validating the translations.

During the noise removal, we observed that, for instance, some studies highlighted the importance of software that uses more efficient computing resources and fosters their energetic efficiency (GARCIA-MIRELES *et al.*, 2018). However, there is a poor connection between SLR conduction and electric energy consumption or computing resources, leading us to remove it to avoid bias in our interpretation. In other cases, we borrowed concepts that are still a novelty for SLR but have already been discussed in other areas like medicine or software development. For instance, we borrowed the concept of "e-waste" (which refers to electronic residues originated from software production/execution (FATHI *et al.*, 2022)) and defined it as "research waste" (ROBERTS; KER, 2015) to refer to SLR results that are misleading for researchers who waste their efforts conducting unuseful investigations. As a main result, we found *41 translations*[9] that characterize sustainable SLR. After that, we validated them, as described following in V1, V2, and V3[10].

[9] These translations are available at <http://dx.doi.org/10.17632/v7bbs5cx7p.1>

[10] The supplementary material containing details of the steps, raw data, and data synthesis of validation are available at <http://dx.doi.org/10.17632/v7bbs5cx7p.1>

4.3.6.1 V.1 - Validation Setup

This step was responsible for setting up the validation. We first converted the 41 translations into a preliminary set of *15 characteristics* and *15 critical factors*. Additionally, we synthesized *16 guidelines* to deal with the critical factors. We also defined a semi-structured interview script and a list of potential participants.

4.3.6.2 V.2 - Delphi Method Application

This step applied the Delphi method to appraise our translations (15 characteristics and 15 critical factors) by gathering opinions from participants. This method is widely used to investigate problems or phenomena not fully understood yet (WANG *et al.*, 2022). Through an iterative process (SKULMOSKI *et al.*, 2007), it aims to find a consensus among participants (OKOLI; PAWLOWSKI, 2004) and includes four key features (SKULMOSKI *et al.*, 2007): (i) anonymity that allows participants to freely express their opinions without undue social pressures to conform from others in the group; (ii) iteration that allows participants to refine their views based on multiple perspectives; (iii) controlled feedback that informs the participants about the other participant's perspectives and provides the opportunity to clarify or change their views; and (iv) statistical aggregation that allows quantitative analysis and interpretation of data. Figure 9 illustrates the five steps we applied:

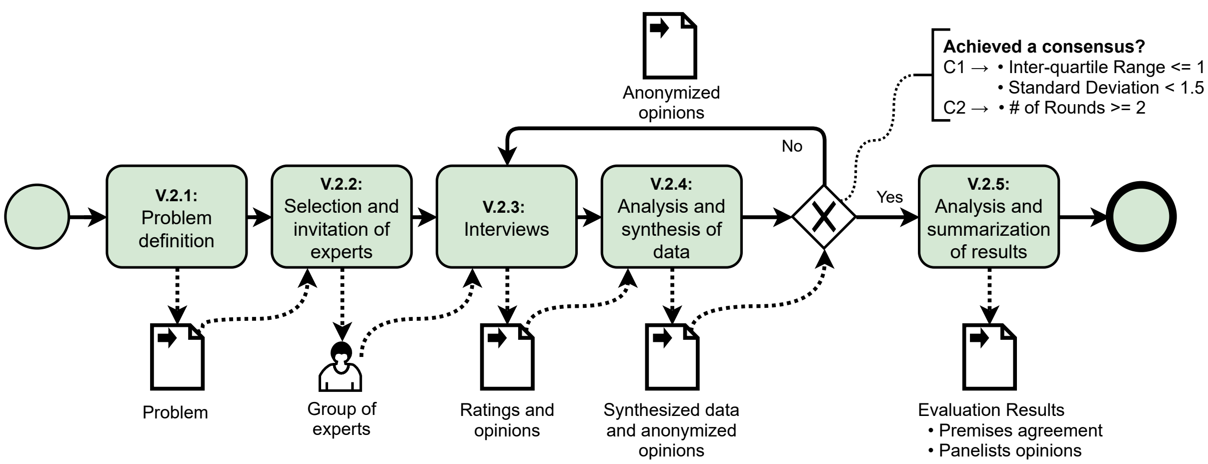

Figure 9 – Details of the validation process conducted based on Delphi Method

- V.2.1 - Problem definition: The objective of the validation is to observe from multiple perspectives and outline a consensus on the understanding of sustainability in SLR.

- V.2.2 - Selection and invitation of experts: We invited 14 researchers based on their profiles whose main criterion was their experience in conducting/updating SLR. We had nine valid participants who had conducted on average 6.8 secondary studies (ranging from 3 to 20), while one formally rejected the invitation.

- V.2.3 - Interviews: We conducted the first round of the semi-structured interviews and collected ratings using a five-point Likert scale (JOSHI *et al.*, 2015). Semi-structured interviews were chosen due to the possibility of guiding the conversation with the interviewee, alternating between closed and open questions, evaluating the data collected but also considering the researcher's impressions by answering questions that aimed to understand how sustainability could be interpreted (WOHLIN *et al.*, 2012);

- V.2.4 - Analysis and synthesis of data: We transcribed the participants' opinions and summarized them into an anonymous set of opinions. We applied the interquartile range (IQR) and standard deviation (SD) metrics to appraise consensus, as often used in Delphi execution (MESHKAT *et al.*, 2014). We considered a consensus when IQR achieved less than 1 (indicating that over 50% of the participants' opinions were focused on a certain point) and SD is less than 1.5 (indicating that the participants reached a consensus). The second round of interviews presented a summary of all data collected in the previous round to the participants. Participants reviewed each item focusing on those with a lack of consensus. Only two rounds were necessary to achieve a consensus about all of the premises. The interviews (first and second rounds) were remotely conducted from July to August 2022 and took approximately one hour each.

- V.2.5 - Analysis and summarization of results: We synthesize and group similar opinions collected in interviews by applying Thematic Analysis (CRUZES; DYBA, 2011a). The agreement with premises and the summary of opinions were the main output of this step.

4.3.6.3 V.3 - Data Synthesis

To synthesize data obtained using the Likert scale, we calculated mode to depict the most frequent opinion from the participants. Moreover, the agreement with premises was expressed by combining the frequency of participants that "Strongly Agree" and "Agree". For qualitative data collected from participants, we transcribed the participants' interviews and applied Grounded Theory (CORBIN; STRAUSS, 1998). In particular, we used open coding to identify codes that are separated into discrete parts for analysis, and axial coding to handle connections between codes and group them according to their similarities. In our case, we grouped the arguments of participants in each characteristic and critical factor related to sustainable SLR.

4.3.7 M.7 - Expression of Synthesis

Finally, this step was responsible for reporting our main results, i.e., our final set of *characteristics*, *critical factors*, and *guidelines* (presented respectively in Sections 4.4.2, 4.4.3, and 4.4.4). Additionally, we report practical insights obtained throughout the translation and validation process that could help researchers to foster sustainability in their SLR.

4.4 Results

This section presents the results obtained throughout the meta-ethnography process. We first discuss how key concepts from both areas (software system and SLR) are comparable in Section 4.4.1. Next, Section 4.4.2 presents the core characteristics of sustainable SLR, Section 4.4.3 offers the critical factors that may impact those SLR, and Section 4.4.4 shows guidelines that could contribute to achieving sustainable SLR.

It is worth highlighting our goal is not to argue that the processes of building software systems and conducting SLR are identical. Indeed, the meta-ethnography provided a systematic way to comprehend better which characteristics, activities, tools, practices, or even quality attributes from both contexts can be similar. Our goal was to focus on relevant aspects of sustainable software systems that could serve as a basis to pave the way for sustainable SLR.

4.4.1 Key Concepts

As a result of Step M.4 (detailed previously), this section outlines the 13 major concepts used in both software systems and SLR and the correlation between them.

The first key concept used in a similar way is the **development process**. A well-defined process is essential in both areas to ensure that the final product will be delivered on time and with quality. Over the years, software development processes evolved to better fit with the needs of the software industry. A classic example is the Waterfall Model that breaks down project activities into linear sequential phases, in which each phase depends on the deliverable from the previous one (PETERSEN *et al.*, 2009). SLR process is also broken down into three main phases (planning, conduction, and reporting (KITCHENHAM *et al.*, 2004a; KITCHENHAM; CHARTERS, 2007)), which generate sub-products used afterward. In the SLR *planning*, requirements for the SLR conduction are defined and documented in an SLR protocol, which is comparable to the requirements and design documents of software systems. In the SLR *conduction*, the protocol is realized to search, extract, and synthesize data (KITCHENHAM; CHARTERS, 2007), similarly to the software implementation. In the SLR *reporting*, internal and external stakeholders appraise the key findings that is comparable to the software acceptance by the clients. Following, the SLR results are published, similarly when software systems are deployed to the clients. Means to maintain available the SLR data/materials for future reanalysis/maintenance are created; this is comparable to creating means to maintain software systems.

Software development processes like Waterfall, Unified Process (KRUCHTEN, 2003), and others sequential usually rely on heavy documentation, becoming often inadequate for some scenarios. Iterative processes such as Scrum (ANWER *et al.*, 2017) that were leveraged from the agile movement have then been recognized as beneficial (ROYCE, 1987) and became popular due to their flexibility for changes (LARMAN; BASILI, 2003). Hence, **iteration** can be considered a core feature of current software development processes. In a similar perspective, iteration is reported as an important characteristic of SLR process (KITCHENHAM *et al.*, 2004a), since such an iteration can produce better and more accurate results in each iteration, reflecting the progressive gain in expertise during the SLR conduction, especially when researchers are novices (LAVALLEE *et al.*, 2014).

The software industry has understood that **collaboration** among stakeholders (e.g., developers, customers, and even researchers) is essential for the success of software projects. In particular, integrating customers within software projects proved to be effective in reducing the delay in response to changes and improving the understanding of developers on the project complexity (RASHID; KHAN, 2018). Likewise, while conducting SLR, researchers work actively and collaboratively (BRERETON *et al.*, 2007), aiming to avoid bias, considering multiple perspectives in decision-making and, consequently, improving the usefulness and accessibility of SLR results to a wider community (BUDGEN *et al.*, 2018a).

Another key concept frequently mentioned in both areas is **knowledge management** (KM). Whereas knowledge is one of the main assets in software companies (LEVY; HAZZAN, 2009), its management improves the teams' capacity to communicate and solve problems more effectively. Accordingly, KM proposes practical solutions, such as for knowledge sharing and acquisition (NAZIR *et al.*, 2020a; NAZIR *et al.*, 2020b) that involve activities like the creation of knowledge bases (NAZIR *et al.*, 2020b). In the same way, an effective KM in SLR is essential, especially to better handle information and exchange experiences (FELIZARDO *et al.*, 2020a). Despite the knowledge transfer being more intense during the SLR conduction and somehow restricted to the involved team, it goes beyond that boundary when: (i) open science practices are adopted to make available research artifacts (MENDEZ *et al.*, 2020); and (ii) when knowledge acquired is fully transferred to other researchers and SE practitioners (BADAMPUDI *et al.*, 2019).

Another key concept is the **usage of tools** to improve productivity (NAZIR *et al.*, 2020b; WOHLIN *et al.*, 2020). Software developers need an environment composed of different kinds of tools, for instance, IDEs (IntelliJ, Visual Studio, etc.), databases (Oracle, MySQL, etc.), project management tools (Jira, IBM RTC, etc.), or even communication tools (Slack, Microsoft Teams, etc.). Likewise, the researchers' productivity is tightly related to the usage of efficient tools to conduct SLR. For instance, search engines (Scopus, Scholar, etc.), data manipulation tools (spreadsheets, reference managers, etc.), statistical data analysis software (NVivo, IBM-SPSS, etc.) or text editors (LaTex, Microsoft Word, etc.) are necessary (AL-ZUBIDY; CARVER, 2018). They play an essential role in supporting researchers' workflow and reducing human errors along the process and, as a consequence, increasing the reliability of the SLR results.

The processes to develop software and conduct SLR deliver **products**, which are respectively the software systems and the SLR results themselves. A common concern in both areas is to improve the quality of these products, e.g., by managing resource limitations (professionals, tools, monetary cost, etc.) or applying practices to create and maintain these products. A recurrent quality attribute in software products is **reliability** (MOURÃO *et al.*, 2018), which fosters the long-term operation of these products and contributes to their sustainability (AMSEL *et al.*, 2011). Likewise, reliability is also mentioned as an important aspect of SLR that count on diverse means (or guidelines) to mitigate possible biases and increase confidence in the results (e.g., involvement of two or more reviewers in studies selection and data extraction). Hence, when SLR follows such guidelines, the results are prone to be reliable (MACDONELL *et al.*, 2010a). Another quality attribute mentioned in both areas is **maintainability**. In software systems, it refers to the capability of introducing changes in software with a small impact on the development and users, making the software's life even longer. To produce maintainable software, techniques like componentization and modularization can minimize cost and effort for maintenance (RASHID; KHAN, 2018). At the same time, maintainability is mentioned in SLR domain referring to the capability of being easily updated. Besides, good software must present **reusability** (GARCIA-MIRELES *et al.*, 2018), i.e., the capability to be reused in other projects. In SLR, reuse of elements (e.g., search strings, selection criteria, and quality criteria) should be promoted, saving time and effort and avoiding conducting SLR from scratch (AMPATZOGLOU *et al.*, 2020). Another attribute mentioned in both areas is **auditability** (or **assessability**), which refers to the possibility of auditing/assessing the artifacts generated and promoting transparency. Several solutions exist to enhance the auditability of software systems, for instance, the usage of metrics (CURRY; DONNELLAN, 2012; BOZZELLI *et al.*, 2013), usage of tools to monitor sustainability indicators (e.g., Greentrackers that measure energy consumption (AMSEL; TOMLINSON, 2010)), or even reference models (e.g., (NAUMANN *et al.*, 2011)) whose suggest the usage of development processes that continuously assess the process and the final product to optimize their sustainability. In a similar way, the auditability of SLR can ensure their reproducibility and reliability. Such auditability mainly depends on the rigor in following the guidelines for conducting, reporting, and maintaining the data/information collected and generated available.

We also found similarities concerning resources consumed to produce software and conduct SLR. In software development, one of them is **human resource** that refers to people involved in software projects, e.g., developers, engineers, analysts, and designers, who assume responsibilities in diverse activities of software development from planning to maintenance. Another resource is **monetary** referring to the cost of human resource and other costs, for example, energy, computers, servers, tools, training, etc. Finally, **time** to build, test, and maintain software is also considered a resource, and its consumption is associated with several aspects of development process (e.g., the development team's experience, tools used, etc). Similarly, SLR also consume these three resources (NUSSBAUMER-STREIT *et al.*, 2021): human resource, monetary, and time. Researchers and practitioners involved in the SLR conduction represent the **human resource**. **Monetary** refers to the cost to use and maintain facilities such as electronic equipment and services (e.g., computers, internet, and electronic databases access), which are usually supported by universities and funding agencies. Finally, **time** consumption is a resource frequently mentioned as the major problem in SLR (FELIZARDO; CARVER, 2020).

4.4.2 Core Characteristics of Sustainable SLR

Table 11 presents the 15 core characteristics of sustainable SLR (C1 to C15). To facilitate traceability, the third column of this table provides the ID of the translations (presented in Appendix C, Table 26). As a result of our validation process, we also provide the Mode (most frequent answer in the Likert scale), the consensus level represented by the interquartile range (IQR), standard deviation (SD), and the percentage of agreement i.e., number of participants that agree or strongly agree with each characteristic. We can observe a strong agreement within our assertions, since in 13 characteristics the mode was five (Strongly Agree). Furthermore, our consensus measurement (IQR and SD) remained between the acceptable limits, which means that despite some participants highlighting sensitive points in those assertions, overall they achieved a consensus.

In more detail, we notice reliability in the SLR results can promote sustainable SLR (**C1**), as reliability can add value to such results and, as a consequence, avoids future researchers from spending efforts to reconduct unnecessarily the same or similar SLR. Currently, the best way to assure reliability is following rigorously the best practices, guidelines, and processes proposed by specialists (**C2**). The literature provides evidence that SLR process is robust and reliable (MACDONELL *et al.*, 2010a), but due to its complexity, it is not possible to assure that it is rigorously followed (WOHLIN *et al.*, 2013). Hence, such guidelines, practices, and process need to be reviewed from the sustainability perspective.

The iteration in the SLR process is another characteristic of sustainable SLR (**C3**). Iteration can minimize the impact of wrong decisions taken mainly during planning and conduction phases and allows researchers to learn from their mistakes. Iteration also makes possible researchers use the knowledge acquired while reading studies and adapt the SLR protocol to better fit the research scope. Hence, iteration is beneficial to the sustainability in SLR, but it still remains open mainly regarding the number of iteration and the required additional resources and different issues (e.g., teammates' experience and available timespan).

Another characteristic is SLR results have real impacts on the target audience over a long-term period (**C4**). The "long-term period" means the timespan whose results remain useful and last as much time as possible according to specificities and volatility of each research area. For this, the SLR are well-built (structure, protocol, insights) and well-structured, serving as a baseline for further investigations, allowing researchers to reuse elements, and providing a snapshot or a historical register of the research topic evolution. From a sustainability perspective, the scientific community should look beyond a current SLR limitation (in particular, the constant obsolescence of SLR) and should move toward proposal new solutions to preserve the SLR results for a longer period of time.

Regarding the economic aspect, sustainable SLR address ways to deal with resource consumption (e.g., time, human effort, and monetary cost). Hence, the responsible use of

Table 11 – Characteristics of Sustainable SLR

ID	Characteristic	Translation	Mode	IQR	SD	Agreement
C1	SLR should report reliable results, mitigating threats to validity and reducing the uncertainty level of researchers.	TR1	5	0	0	100%
C2	SLR should follow a process that complies with recommended practices and guidelines to ensure quality.	TR3	5	1	0.5	100%
C3	SLR should be conducted using a process that is iterative and considers changes in the SLR protocol during the pilot test.	TR3	5	0	1.01	88%
C4	SLR should have long-term goals aiming at impacting the research area and the community around (including SE researchers and practitioners) over a long period of time.	TR4	4	1	1	77%
C5	SLR should be produced with responsible use of resources (e.g., time, human effort, and monetary cost) and aiming at reducing the time consumption (e.g., by adopting automation tools).	TR4, TR5, TR9	5	0	0.44	100%
C6	SLR should have a documentation that is detailed, easily understandable, auditable, and accessible to others (researchers, practitioners, and others).	TR9	5	0.25	0.21	100%
C7	Items of SLR protocol (e.g., search string, selection criteria, included studies, etc.) and other items (e.g., raw data extracted) should be learnable, accessible, reusable, modifiable, and adaptable.	TR9	5	0.5	0.65	95%
C8	SLR over their whole life cycle (i.e., while the SLR is conducted, update, and/or replicated) should be continuously assessed and continuously documented, i.e., SLR should be continuously updated.	TR8	4	1	0.88	66%
C9	Components of SLR (e.g., search string, selection criteria, and other elements like raw data) should be reused in the update of SLR.	TR8	5	0	0.44	100%
C10	Researchers should make accessible all evidence found in the SLR (i.e., selected studies) as well as all associated data (e.g., raw extracted data) aiming to keep the viability of future SLR updates.	TR10	5	1	0.53	100%
C11	Stakeholder needs should be translated into research questions to be answered by SLR.	TR1	5	0	1.33	88%
C12	SLR should provide results that are useful to a wider community not just fulfilling their particular needs or from specific research groups.	TR2, TR5	5	1	0.73	88%
C13	SLR results should positively impact a research area.	TR7	5	1	1.12	77%
C14	Researchers should avoid research waste: (i) evaluating if an updated SLR already exists on the same topic, (ii) reusing components from the previous SLR; and (iii) publishing outputs that are useful for target audience.	TR6	5	1	0.72	92%
C15	Researchers while conducting SLR should make decisions and adopt practices that minimize the negative impact on current (e.g., saving efforts by automating tasks) and future researchers (e.g., documenting correctly and providing accessible recommendations for practitioners).	TR7	5	0.5	0.42	100%

Legend: IQR - Interquartile Range; SD - Standard Deviation

resources emerges as another characteristic (**C5**), highlighting the importance of managing better resources during the SLR conduction. We intentionally use the term "responsible" aware of the inherent trade-off between the resources' usage and gains in the SLR quality, i.e., "responsible" means creating ways of applying the resources without sacrificing quality. A possible solution is to handle the technical support for SLR better since more efficient tools simultaneously leverage the researchers' productivity and quality of SLR.

The documentation of SLR is tightly connected to their sustainability since it represents the final product (i.e., SLR results and all items that document an SLR). Such a product was generated through a complex process containing multiple issues (phases, tasks, decisions) that directly impact the quality, reliability, and usefulness. Such documentation is detailed enough to be easily understood, auditable, and accessible to researchers, practitioners, and others (**C6**). In addition, all documentation items (e.g., SLR protocol, studies included and excluded, data extracted, results synthesis) are available to reuse, modification, or adaptation (**C7**).

From the sustainability view, an SLR is continuously assessed and documented through continuous update (**C8**), which refers to periodically appraising the need for updates, considering the community's interest in that SLR research topic. Detailed documentation of the SLR life cycle (i.e., a macro vision of the SLR) becomes essential. The SLR life cycle begins when the idea for its conduction emerges, goes through its conduction and publication, continues with continuous updates or its reuse when an update is not possible or not necessary, and finishes with its deactivation/decayment. A good comprehension of the SLR life cycle could also support the community in understanding better understand when and how to update them.

Another characteristic of a sustainable SLR is the possibility of reusing its components (e.g., search string, selection criteria, studies excluded in the previous conduction) to update it (**C9**). For that, detailed documentation of that SLR is essential. Similarly, another characteristic is the availability of all evidence (i.e., the included studies) and other relevant data (e.g., raw data) (**C10**). While detailed documentation can consume more resources from researchers (especially time and effort), such detailing can promote reproducibility/auditability and, consequently, reliability in SLR. Hence, the means to document SLR better should be reviewed to comply with the sustainability view, which intends to replace the "simplistic" view that has recommended exhaustive documentation without considering the existing trade-offs.

Sustainable SLR present research questions that answer the needs of stakeholders through evidence from the scientific literature (**C11**). Hence, sustainable SLR also address the usefulness of their results. These results can have several purposes, e.g., providing a baseline for controlled experiments or an overview of a given topic to guide future research. Sustainable SLR must also impact of research area by offering results useful for a wider community, i.e., by adopting an altruistic view, fulfilling (whenever possible) the needs of a whole community instead of only fulfilling particular needs (**C12**). Consequently, sustainability prioritizes publishing results that positively impact the research area (**C13**). At the same time, several challenges emerge. Social aspects such as the limited experience of researchers can hamper the capacity of SLR to contribute to the area. Another issue is to define what "wider community" means precisely since there is no consensus about the optimum size of the target audience. These three characteristics (C11, C12, and C13) intend to develop a new mindset on the research work and promote SLR results useful for academia and industry.

Lastly, sustainable SLR minimize the waste of time and effort and maximize the results' usefulness for the target audience (**C14**). The commitment to sustainability premises is essential for researchers to understand the impact of their decisions on projects and on the whole scientific research scenario (**C15**). In a broader sense, sustainability in SLR aims to support the scientific community in achieving the EBSE's original objective, which is to provide reliable evidence to support the decision-making process.

4.4.3 Critical Factors for Sustainable SLR

We identified 15 critical factors for achieving sustainable SLR, as listed in Table 12 (CF1 to CF15), and grouped them into four categories (Communication & Collaboration, Knowledge Management (KM), Technical Support, and Resources Usage). The third column of this table provides the factors (which resulted from the translations presented in Appendix D, Table 27). The result of our validation also provided us with the Mode (most frequent answer in the Likert scale), the consensus level represented by the interquartile range (IQR), standard deviation (SD), and the percentage of agreement with each factor. Overall, most participants strongly agreed with the critical factors, and a consensus was also achieved by observing IQR and SD. Following, we discuss the reasons why these factors are critical to achieving sustainable SLR and, therefore, they should be addressed during the SLR conduction.

Table 12 – Critical Factors of Sustainable SLR

Category	ID	Critical Factor	Translation	Mode	IQR	SD	Agreement
Communication & Collaboration	CF1	Rich communication	TR11	5	0	0	100%
	CF2	Effective participation of stakeholders (researchers and SE professionals) in SLR process	TR12	5	0	0	100%
	CF3	Knowledge of stakeholders about the research domain	TR13	5	1	1.27	77%
	CF4	Experience of team members in SLR conduction	TR14	4	1	0.93	88%
KM	CF5	Improvement of SLR reusability	TR15	5	1	0.5	100%
	CF6	Usage of refactoring techniques	TR16	5	1	0.88	77%
	CF7	Efficient knowledge sharing and transfer	TR17	5	0	0.44	100%
Technical Support	CF8	Usage of tools to support SLR	TR18	5	0	1.01	88%
	CF9	Maturity of support technology	TR19	5	1	1.27	77%
	CF10	Accessibility of support technology	TR20	5	0	1.01	88%
Resources Usage	CF11	Usage of techniques that minimize the resources consumption	TR21	5	1	1.01	88%
	CF12	Efficient management/usage of resources	TR22	5	0	1	88%
	CF13	Usage of feasibility tests	TR23	4	0	1.32	77%
	CF14	Maintenance of SLR	TR24	5	0	0.44	100%
	CF15	Usage of iterative process	TR25	5	1	0.53	100%

Legend: KM - Knowledge Management; IQR - Interquartile Range; SD - Standard Deviation

In the context of this work, a "critical factor" refers to sensitive points in a process (in our case, the process to conduct SLR) that have a major impact on deliverables (in our case, the SLR life cycle, SLR results, and associated documentation) and, therefore, an adequate management of them may lead to success or failure.

The first critical factor is communication & collaboration, which addresses several aspects of the relationship between researchers and SLR stakeholders (in the context of this work, stakeholders are those that supported the SLR process and somehow benefited from results, e.g., researchers and SE practitioners). Rich communication (**CF1**) and effective participation of stakeholders (**CF2**) were unanimously considered critical factors, especially because of their impacts on the social dimension of sustainability. Rich communication fosters knowledge sharing among team members and minimizes wrong decisions that could occur due to lack of experience or knowledge about the research domain, while the effective participation of stakeholders can promote the conduction of SLR whose results are more useful. For software industry, the adoption of agile methodologies has promoted communication & collaboration between development teams and customers, while some researchers have pointed out that aligning the industry needs and academic research is still a challenge. Hence, sustainability plays an important role in reaffirming the need for exploring more efficient ways to align industry and academia in both software development and SLR conduction.

Another critical factor is the knowledge of stakeholders about the research area associated with SLR (**CF3**). In our validation process, participants highlighted that previous knowledge about the domain is crucial to making the right decisions leading to more reliable and valuable results for the target audience. At the same time, team members also learn from studies during the SLR process (IWAZAKI *et al.*, 2022) and acquire deeper knowledge about the research area. The team's experience in conducting SLR is another critical factor (**CF4**) since the SLR conduction process is complex, and the experience with this process can result in better results. During our validation, we also found that hybrid/balanced teams (with experienced researchers and novices) are acceptable and somehow desirable due to the opportunities of sharing knowledge with novices and preserving SLR educational purposes (i.e., novices gain experience in research).

Effective KM can promote sustainability in SLR, particularly when increasing the possibility of reusing SLR components (e.g., SLR protocol, selected studies, datasources, search strings, etc.) (**CF5**). Such reuse may reduce the effort and time needed to conduct/update SLR, impacting mainly the economic dimension of sustainability. Sustainability reaffirms the need for approaches like Open Science that have drawn attention by highlighting the need to document SLR artifacts correctly and, consequently, making possible reusing them. In this context, the usage of refactoring techniques emerges as another critical factor (**CF6**) to ensure that knowledge produced during SLR conduction is used as most as possible in further works. An example of a refactoring technique is forward snowballing to update SLR (FELIZARDO *et al.*, 2016). When using this technique, the initial set of studies is fully reused aiming to find new evidence with possibly fewer effort. Refactoring techniques could also be applied to multiple artifacts of SLR, e.g., search strings construction, database selection, definition of quality assessment criteria, supporting researchers to reuse these elements, reinterpret data and answer research questions.

An important activity of KM is knowledge sharing/transfer (**CF7**). It can occur at least in three different moments during the SLR conduction: (i) when researchers and stakeholders share their knowledge/experience during consensus meetings; (ii) when researchers provide open access to data (including raw data) and report all decisions made in the SLR process; and (iii) when researchers report lessons learned and insights about the applicability of results in practice and provide recommendations for SE professionals. Hence, knowledge sharing/transfer needs to be considered, since the main premise of sustainability in the social dimension is to handle human aspects of the SLR process and maintain the capability of social groups to exchange information in a reliable and effective way.

Another critical issue for achieving sustainable SLR is the technical support, which impacts mainly the technical and economic dimensions. The usage of tools is then a critical factor for successful SLR (**CF8**), minimaly, authors should use tools to support data manipulation like spreadsheets (e.g., Excel, Google sheets), reference managers (e.g., Jabref, Mendeley), and text editors (e.g., Word, LaTex) . At the same time, the lack of adoption of tools that are specific for SLR (e.g., StArt, Parsifal) is a well-known problem, since these tools usually suffer from limitations, lack of integration, and slow and fragmented development (FELIZARDO; CARVER, 2020). Al-Zubidy and Carver (2018) state that tools to support the entire SLR process are immature. It is also observed that universities/research groups have difficulty maintaining these tools and ensuring their quality. For all these reasons, the maturity of supporting tools is required (**CF9**).

Another factor is the accessibility of the technologies (or tools) that support the SLR conduction (**CF10**). For instance, several problems were already reported in the publications databases regarding their front-end and back-end (including user interfaces, search presentation, filters, syntax, algorithms, and optimization), impacting the researchers' productivity while conducting SLR (AL-ZUBIDY; CARVER, 2018). Another problem found in SLR specific tools is their long learning curve (MARSHALL *et al.*, 2018). Otherwise, accessible tools can promote the researchers' productivity and efficiency and, as a consequence, reduce the amounts of effort required to conduct SLR. Hence, tools' accessibility can be considered a critical factor for the sustainable SLR.

The main contribution of the economic dimension of sustainability is to diminish the barriers that cause the excessive consumption of resources to conduct or update SLR. It is important to mention that by "excessive" we refer to all inefficiencies that make SLR a time-consuming method. Hence, another critical factor is the creation and usage of strategies to minimize resources consumption (**CF11**) and efficiently manage them (**CF12**). An example of these strategies is snowballing that, when compared with a traditional database search, provides a significant gain in efficiency (WOHLIN, 2016; JALALI; WOHLIN, 2012). At the same time, other studies like Felizardo *et al.* (2016), Mendes *et al.* (2019) evaluated its efficiency and concluded that snowballing is suitable to update SLRs and can be used in combination with database searches, resulting in a more complete search strategy (WOHLIN *et al.*, 2022). Therefore, it is worth highlighting that the adoption of any technique to minimize resource consumption while conducting SLR (e.g., automation techniques (FELIZARDO; CARVER, 2020; WATANABE *et al.*, 2020)), only makes sense whereas the quality of SLR is not negatively impacted.

In general, pilot testing is used to assure the quality of SLR, for instance, to avoid wrong research questions, usage of inappropriate terms in the search string, or adoption of wrong inclusion criteria. When sustainability is considered, it is necessary not only performing pilot test but also appraising the feasibility of SLR (i.e., comprehending better the scenario where the SLR will be inserted). Hence, another factor is the usage of feasibility tests (**CF13**). Feasibility tests should consider sustainability perspective to evaluate SLR in its early stages better comprehending which are the expected impacts on stakeholders and readers, reinforcing the need for a new SLR. This critical factor was a bit controversial in the validation process since feasibility tests adds one more step in SLR process and could consume even more resources; furthermore, in areas considered new (i.e., with few studies available) feasibility tests are prone to be less critical. We achieved a consensus that those tests are useful when the research area has many studies available and this analysis provides a more precise analysis about the need of a new SLR avoiding unnecessary research work.

Finally, sustainability advocates the need to change the way SLR are maintained/updated (**CF14**). An SLR update process shall avoid unnecessary updates and minimize efforts while keeping the SLR results still relevant to stakeholders. Sustainability also states the need to adopt an iterative process to conduct SLR (**CF15**). An iterative process can reduce the impacts of changes in the protocol during the SLR conduction and reduce drawbacks created by social aspects (e.g., the lack of experience of team members).

4.4.4 Guidelines to Conduct Sustainable SLR

Several SLR guidelines and processes already exist and have been recommended to avoid pitfalls during the SLR conduction and update (e.g., (KITCHENHAM; CHARTERS, 2007; BUDGEN *et al.*, 2018a; AMPATZOGLOU *et al.*, 2020; WOHLIN *et al.*, 2020)). This section presents complementary guidelines from the sustainability perspective that could contribute to leveraging sustainable SLR. Table 13 lists these guidelines (G1 to G16), which were obtained from the translations presented in Appendix E (Table 28) and are organized into four categories: Communication & Collaboration, Knowledge Management, Technical Support, and Resources Usage..

Existing guidelines intend to strengthen the information exchange e.g., using techniques to solve disagreements (PETERSEN; ALI, 2011), performing brainstorming sessions to reach multiple interpretations of results, or connecting SLR results to SE practitioners by using evidence

Table 13 – Guidelines for sustainable SLR

Category	ID	Guideline	Translation
Communication & Collaboration	G1	Schedule regular meetings to **encourage communication/-collaboration** to brainstorm with researchers and stakeholders through to better comprehend the research domain and share solutions to improve the SLR process.	TR26
	G2	Always use **two or more reviewers** to double check the results of the conduction process	TR27
	G3	**Feedback from target audience** during SLR conduction should improve the overall quality of SLR	TR28
	G4	SLR design must **comprise all stakeholder's requirements** and provide evidence that is **user-friendly that meets users' needs**.	TR29
	G5	Follow short conduction/update **cycles/iterations** managing the SLR redesign along the process according to the stakeholders requirements and **using their opinion to validate the results in each iteration**	TR30
	G6	**Consider multiple perspectives** while conducting SLR identifying with stakeholders issues, objectives, constraints, considerations, and **apply acceptance tests** to validate requirements with stakeholders and target audience.	TR31
Knowledge Management	G7	Enhance **knowledge management** by using **sharing platforms** and effective strategies to **avoid knowledge loss** and **improve sharing among team members**.	TR32
	G8	Create a **knowledge base** to improve knowledge management and allow SLR stakeholders to **learn from previous experiences** of the research group.	TR33
	G9	Include in the research team or **get support from skillful and experienced researchers/stakeholders** to acquire knowledge about the research domain and SLR process.	TR34
Technical Support	G10	Provide **sustainable SLR tools and methods** to researchers and practitioners	TR35
	G11	**Avoid tools** and methods that **affect** researchers and practitioners in their **work negatively**	TR36
	G12	**Demonstrate the current tools and methodologies applicability** to researchers and practitioners	TR37
Resources Usage	G13	**Develop pilot tests for SLR that are reusable** and allow researchers to build them incrementally, avoiding throwing away the information and wasting resources reconducting unnecessarily any SLR step.	TR38
	G14	**Adopt a "component-based" development** strategy to maximize the **reuse of existing schemas/assets** and **minimize time, cost, and efforts** of the research team whenever a refactor is needed.	TR39
	G15	Conduct SLR that has a **lower impact when introducing changes** and improve the **usability** of results making its lifecycle longer	TR40
	G16	Enhance the quality of SLR concerning **maintainability, usability, and reliability** of results positively interacts with sustainability and promotes SLR **perdurability**	TR41

briefings (CARTAXO *et al.*, 2016). Hence, communication seems obvious and inherent in the SLR process, but it is complex and involves multiple variables (e.g., culture, language skills, background, etc.). As mentioned earlier, **communication** can contribute to achieving sustainable SLR; hence, a guideline is to use strategies to improve communication (e.g., online or face-to-face meetings) and using supporting technologies to create a productive environment (**G1**). Likewise, another guideline is to use two or more reviewers to perform SLR tasks reducing bias in activities involving sensitive judgments (**G2**).

The process to create sustainable SLR requires continuous **collaboration** between researchers and the target audience (or SE practitioners) to add value to the review. Guidelines **G3** to **G6** could ensure SLR results are aligned with the needs of the target audience. That way, sustainability could avoid, for example (i) designing SLR questions with a narrow mindset; (ii) simplistic and wrong views on SE practice; (iii) presenting SLR findings in a format that is difficult to understand by professionals. In this sense, researchers should collect from the target audience their insights about the results (**G3**) and use this feedback as a baseline to improve the quality and usefulness of results. For this, stakeholders should actively participate in the definition of SLR goals (**G4**) and follow the process of evaluating and pointing out inconsistencies of results in each iteration (**G5**). While analyzing results, multiple perspectives are important to observe results in different lights, in this stage acceptance tests can be applied to validate whether requirements defined in planning were covered (**G6**).

As mentioned in previous sections, complete and coherent documentation (which is a result of transforming tacit knowledge into explicit knowledge and includes artifacts generated, procedures, and decisions made during the SLR process) represents a fundamental pillar of sustainable SLR. Guidelines G7, G8, and G9 address the **knowledge management** associated with SLR. In particular, **G7** emphasizes the importance of using knowledge-sharing platforms, and **G8** recommends creating knowledge bases to make it possible to learn from accumulated experiences with previous SLR. **G9** recommends getting support from more experienced researchers outside of the research team (which is conducting the SLR), thus promoting the experience exchange.

We observe sustainable SLR depend directly on good **technical support**. **G10** advocates the need of sustainable tools (i.e., mature, accessible, reliable, and with long-time support) and methods that facilitate the SLR conduction. For this, the SE community needs to cooperate to transform current tools (mostly still prototypes) into reliable and mature tools. Henceforth, universities, funding agencies, and other stakeholders need to join efforts to consolidate tools that should also present long-time support, like in medicine with Cochrane[11] that offers a collection of tools, e.g., RevMan. At the same time, tools or methods that could negatively affect the SLR conduction should be avoided (**G11**). After achieving mature tools and methods, coordinated work is necessary to disseminate them to stakeholders and demonstrate they are applicable, effective, and trustworthy (**G12**).

Several authors propose solutions to reduce the excessive **resource usage** in the SLR process (e.g., (WATANABE *et al.*, 2020; MERGEL *et al.*, 2015; FELIZARDO *et al.*, 2017a; FELIZARDO *et al.*, 2011)). In sustainable SLR, pilot tests over SLR protocol should be conducted, and their results should always be reused, i.e., previous efforts should not be discarded (**G13**).

[11] <https://community.cochrane.org/help/tools-and-software>

The sustainability perspective refuses the idea that each new SLR designs new and specific components (e.g., the items of the SLR protocol); instead, it announces reusability should be a central element of SLR design. Currently, some works already propose to evaluate SLR components like quality assessments (ZHOU *et al.*, 2015) or threats to validity (AMPATZOGLOU *et al.*, 2019; ZHOU *et al.*, 2016)) and propose checklists to support researchers' reuse of these items, another example is presented in Napoleão *et al.* (2021) that establishes a reusable search string to search for secondary studies. Henceforth, SE community should move towards designing other reusable SLR components and adopting an SLR component-based design to reduce excessive resource usage (**G14**). Other guidelines associated with resource usage are **G15** and **G16** which address the usability, maintainability, and reliability of SLR. Aiming at these qualities during the whole SLR life cycle and taking decisions that promote them, resource usage could be reduced and, as a consequence, sustainable SLR could be achieved. In particular, usability refers to the usefulness of SLR components and results for further research, hence, even if SLR could not be updated it is still possible with minor efforts refactor elements to avoid starting from scratch. Maintainability refers to keeping SLR well-documented and open access to generated artifacts to improve its capability of being updated. Finally, reliability preserves the rigor in SLR process avoiding that additional efforts being wasted to double-check unreliably elements or even reconduct the whole SLR.

The 16 guidelines presented in this section refer to the first set of guidelines focused on sustainable SLR and obtained from our meta-ethnography study. This section provides clues about which guidelines have been applied in other areas and could work in SLR. Further, as the SLR research area advances regarding guidelines and practices to achieve sustainable SLR, this set needs to be refined and updated.

4.5 Discussion

The SE community has unquestionably benefited from SLR, but its conduction and update (as they should be) are sometimes hampered, mainly due to high resource consumption (specifically human resources) and lack of effective impact of SLR results on SE practices. These and other drawbacks reveal problems that have been neglected or cannot be treated in isolation, so motivating us to introduce the concept of sustainability in SLR. This concept intends to provide a brand-new perspective of SLR, and *the novelty is a more integrated view on the SLR problems addressing the multiple sustainability dimensions (i.e., social, economic, and technical) and aiming at ultimately achieving sustainable SLR*. Following, we discuss the main findings, future work, and threats to the validity of our work.

4.5.1 Main Findings

Our findings resulted from a meta-ethnographic study that is considered a best-suited method for generating models or higher-order theories that interpret findings across multiple studies (ATKINS *et al.*, 2008), as it is the case of this work. We identified the following five main findings:

- **Correlation between software systems and SLR**: A main finding is that SLR must be treated not only as a research method but also as a complex social-technical process (like in software development) that must consider cost-benefit trade-off to create artifacts (documents including those that contain answers to RQ), which must also be managed accordingly to provide real impact on a target community. This finding resulted from the correlation of key concepts widely used in software systems (e.g., process, KM, usage of tools, qualities like reliability and maintainability, etc.) with those in SLR. We also observed that new concepts/terms (e.g., refactoring, components, reuse of components, and stakeholders) from the software systems context could be appropriated to SLR, as discussed previously in Section 4.4.1.

- **Characterization of sustainable SLR**: Another important finding is that it seems to make sense to say about the main characteristics of sustainable SLR and disseminate them to the EBSE community. Such characteristics (summarized previously in Table 11) were based on multiple understandings of sustainability deeply discussed for more than ten years in the software systems area. The awareness of these characteristics could change the way SLR are conducted.

- **Critical points in SLR process**: Another finding is that there are sensitive points (or critical factors, as summarized previously in Table 12) in the SLR process that will require special attention when aiming at becoming SLR sustainable. This work brings to light such points and intends to draw the attention of the community to them. At the same time, this work also suggests a set of guidelines (summarized previously in Table 13) to deal with these points.

- **Inclusive SLR process**: We also observed that SLR processes should be inclusive in the sense decisions should consider the whole context, i.e., goals and constraints of stakeholders and target audience (like the industry). Besides reinforcing the importance of following current SLR guidelines, our work introduces pieces in the SLR process that enable researchers to revise those guidelines to encompass the sustainability vision by considering a fair balance among the three sustainability dimensions (social, economic, and technical).

- **Less rigid SLR conduction**: The view on SLR from the sustainability perspective shows SLR conduction can be even less rigid than "traditional" process and guidelines that the literature has somehow imposed. The sustainability view states that the conduction of each SLR should consider its context; for instance, if the results of an SLR will be used just once, detailed documentation could be unnecessary. In other words, the sustainability view states that a trade-off between cost and benefit should be performed.

- **Suitability of a meta-ethnographic study**: A general finding is that the results of a meta-ethnographic study are sometimes broader and generic (like in other works and ours), as this kind of study maps concepts and characteristics from one area to another (NOBLIT; HARE, 1988; FU *et al.*, 2019). At the same time, in works similar to ours that intend to introduce novel concepts to a given area, the meta-ethnographic study seems to be an adequate method since it can systematically and rigorously collect, analyze/interpret, synthesize, juxtapose, and transpose/translate concepts and characteristics from a context to another (NOBLIT; HARE, 1988; FU *et al.*, 2019). As far as we know, this is the only method that could have thoroughly supported us in introducing the concept of sustainability to SLR.

As we systematically conducted our research method, which is very detailed and composed of different approaches, we believe our findings and results are correct. We applied well-known and experimented approaches to search and identify relevant literature (i.e., SLR). Hence, our results are exclusively based on the most recent evidence (studies from both software systems and SLR areas) published in relevant peer-reviewed venues. Our results are also based on the up-to-date challenges and problems reported by specialists in SLR. To interpret the results, we applied a meta-ethnographic study, and finally, we validated the results using the Delphi method and involving researchers with an extensive experience in SLR conduction.

4.5.2 Future Work

As this work introduces a novel concept in the SLR area, it will demand further investigations and could pave the way for new research directions. Some possible future works are:

- **Dissemination of the sustainability concept in the SLR community**: Considering that this work offers a new perspective about how to deal with SLR, for future work, it is necessary to foster discussions and gather evidence (through case studies and experiments) to observe the impact of conducting SLR in accordance with the sustainability perspective, i.e., considering the characteristics of sustainable SLR and drawing attention to their conduction and update processes. As a result, this work could be the starting point to change community's mindset about how to address SLR.

- **Refinement of the set of characteristics of sustainable SLR**: This work proposes an initial set of characteristics of sustainable SLR. For future work, these characteristics need to be validated empirically, while other characteristics can emerge or change; hence, this initial set needs to be adjusted after investigations. It is also necessary to conduct other exploratory studies to check the critical factors and guidelines presented in this chapter aiming at refining them. In turn, this work provide a wide view showing a variety of characteristics, critical factors, and guidelines for sustainable SLR; thus, each of them need to be deeply investigated requiring a considerable number of further investigations.

- **Organization to leverage sustainable SLR**: It is worth highlighting that the medicine area has benefited from a not-for-profit organization — Cochrane[12] — that supports clinicians, patients, carers, researchers, policy-makers, and others interested in making health decisions using high-quality information (including results of SLR). Such organization: (i) provides an environment for everyone that needs to collaborate (i.e., Cochrane Engage[13]); (ii) minimizes unnecessary efforts by providing means for avoiding research overlapped with existing ones; and (iii) provides integrate tools that aid the SLR conduction[14]. Following this example, EBSE community should consider to establish an international integrated organization to deal with several SLR that have been conduced and many others that will emerge. We also believe that the existence of such an organization can leverage SLR to become sustainable, since this organization will be a means to share knowledge, promote reuse, minimize duplicated effort, and so on.

- **Continuous process for SLR**: Another future work is to define a continuous process for SLR (inspired in agile methods and continuous software engineering (FITZGERALD; STOL, 2017)) considering the trade-off among cost, benefit, and effort required and aiming at keeping results continually updated. Initiatives in that direction already exist in medicine (e.g., living systematic reviews (ELLIOTT *et al.*, 2017)) and recently emerged in SE (NAPOLEÃO *et al.*, 2022). In this scenario, we believe that the concept of sustainability (as discussed in this work) can make it possible to comprehend better the importance of adopting a continuous process for SLR.

In addition to the aforementioned future works, other more punctual issues could be revisited using the sustainability viewpoint, including: (i) SLR stakeholders should be better comprehended aiming at finding solutions to foster collaboration and information exchange; (ii) the trade-offs between SLR quality and time/effort consumption for SLR conduction and update should be investigated; and (iii) SLR supporting tools with interoperability among them as well as maintainability are also necessary. Additionally, some issues have been poorly discussed or do not even exist formally in the SLR context, for instance, metrics to measure the SLR execution or the quality of SLR documentation (differently from software development area). We believe that the introduction of sustainability into SLR context can lead to the need of exploiting reliable metrics.

[12] <https://www.cochrane.org/about-us>
[13] <https://community.cochrane.org/help/tools-and-software/cochrane-engage>
[14] <https://community.cochrane.org/help/tools-and-software>

4.5.3 Threats to Validity

This section presents the potential threats to the validity of this work and the countermeasures we adopted to mitigate such threats. We divided these threats into four categories (as proposed by Wohlin *et al.* (2012)):

- **Construct validity**: A possible threat is to miss relevant studies during the process of searching studies (from both software systems and SLR areas); consequently, some concepts, terms, characteristics, and guidelines may not have appeared in our results. To mitigate this threat, we pilot-tested the protocol of the tertiary study on sustainability in software systems. For studies associated with SLR studies, we performed group meetings with experts (with more than 10 years of experience in SLR) to decide which studies would be included. As a result, we believe our work considered possibly all relevant studies.

- **Internal validity**: During the data extraction from studies of both software systems and SLR areas, some studies did not provide clearly information or enough details. Since meta-ethnographic study is an interpretive approach to synthesis information, we performed our interpretation that may have led to misconceptions. To minimize this threat, we carefully revisited all studies and conducted consensus meetings with all authors, specially when identifying terms and concepts, determining relations among them, and generating translations. Additionally, we collected the opinion of external domain experts. We also made available the raw data to make it possible to audit our work.

- **External validity**: A potential threat is the number of participants in our validation process. To mitigate this threat, we applied the Delphi method that can improve the data collection, considering different participants' views and extracting as much as possible quantitative and qualitative data for our analysis. We believe our validation process was able to capture important feedback to refine our results and ensure that they are somehow correct.

- **Conclusion validity**: A threat associated with the conclusion is the relation we defined from software sustainability to SLR that may have resulted in imprecise mapping and translations. To mitigate this threat, we systematically followed the steps of a meta-ethnographic study and also built our discussions strictly based on the current state of the art of software sustainability and SLR; therefore, topics not explicitly discussed in literature was not considered (for instance, the usage of metrics in SLR), remaining them to future work.

4.6 Final Remarks

The last years have been stuffed with many SLR in SE and other computing areas. While several well-known problems associated with the SLR conduction and update still exist, several initiatives often isolated have emerged without complete success in solving those problems. In this scenario, this work explores a novel concept for the SLR area that could deal with the SLR problems in an integrated way. This work offers a characterization of sustainable SLR as well as points in the SLR process that should receive special attention. In turn, this concept can add value to the existing SLR research and contribute to advancing the state of the art of the area. More importantly, this work intends to change the mindset of the community regarding the way to manage SLR. From this first work that addresses SLR sustainability, several open issues should still be investigated, mainly regarding improving the SLR process by incorporating trade-offs between cost/effort and benefits in a more conscious way and, ultimately, ensuring that SLR continue to offer their benefits better.

CHAPTER

5

Practical Actions to Leverage Sustainability in Systematic Literature Reviews

DOS SANTOS, V.; NAKAGAWA, E.Y.; Practical Actions to Leverage Sustainability in Systematic Literature Reviews, 2023.

Status of this publication: This paper soon will be submitted to a high-ranked journal in SE area.

Abstract

Context: Systematic Literature Reviews (SLR) over the years consolidated as an essential source of information for Evidence-Based Software Engineering (EBSE) providing several benefits in the whole process. At the same time, SLR method also suffers from problems, especially the high consumption of time and effort from researchers, and the lack of effective impact of results in Software Engineering (SE) practitioners. In this context, Sustainability for SLR was proposed to address open issues from a more integrated view, however, there is still a lack of more practical recommendations on how to transpose the current SLR process into a more sustainable one.

Objective: The main goal of this paper is to present practical actions to leverage sustainability in the current SLR.

Method: Based on sustainability characteristics, we defined 19 indicators to provide us clues about the state of the practice of current SLR. Next, we selected a sample of ten relevant studies published in the last decade in the SE area. Following this, we carefully extracted data from these studies and synthesized which factors have a major impact on SLR sustainability. Finally, we synthesize our insights about possible actions for leveraging the sustainability of SLR.

Results: Our results indicate the existence of several weaknesses in the current SLR, for instance, flaws in the research process, documentation with a lack of important details, shallow indications

of internal and external collaboration, poor knowledge management and usage practices that facilitate information exchange, lack of usage of support tools, lack of support for maintenance, and lack of interest on including practical insights for practitioners.

Conclusion: Some of the problems impacting the current sustainability of SLR are solvable with simple actions that depend exclusively on the commitment of researchers. Hence, we claim that our recommendations should be followed by the research community to improve SLR sustainability.

5.1 Introduction

Systematic Literature Reviews (SLRs) were introduced to assist researchers in aggregating results from empirical studies, supporting the decision-making process in the software development industry (KITCHENHAM *et al.*, 2004a). The SLR process is well-defined and comprises three main phases (KITCHENHAM *et al.*, 2015): (i) Planning: This phase aims to establish the review protocol, which contains the definition of research aims and scope, research questions, inclusion and exclusion criteria, sources of studies, search string, quality assessment, and data extraction/synthesis strategy; (ii) Conduction: This phase aims to execute the search strategy defined in data sources to select relevant studies, extract, and synthesize data; (iii) Reporting: This phase aims to answer research questions and report results to disseminate them to the interested community.

Usage of the SLR method offers several advantages and disadvantages for stakeholders (FELIZARDO *et al.*, 2020b). For researchers, the benefits include the improvement of research skills (PEJCINOVIC, 2015; BORREGO *et al.*, 2015) and the possibility of learning from studies (BABAR; ZHANG, 2009; ZHANG; MUHAMMAD, 2012); while for the SE community, benefits include the possibility of dealing with information from different studies in an unbiased manner (KITCHENHAM *et al.*, 2015; NIAZI, 2015), production of auditable and repeatable results (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a), identification of research gaps and also perspectives for future investigations (KITCHENHAM *et al.*, 2015; BABAR; ZHANG, 2009). In parallel, disadvantages reported in SLR process are often connected with human factors, such as the lack of compromise with well-experimented guidelines (KUHRMANN *et al.*, 2017), poor documentation (ZHOU *et al.*, 2016; AMPATZOGLOU *et al.*, 2019) or even lack of specialists to support critical activities in review (CARVER *et al.*, 2013). Besides, other drawbacks are caused by technical factors likewise the lack of maturity of tools that still require authors to perform a lot of manual work (AL-ZUBIDY; CARVER, 2018).

SLR problems encouraged the search for solutions capable of alleviating the drawbacks of the method, for instance, usage of iterative process (FABBRI *et al.*, 2013) or the creation tools to automate time-consuming SLR activities (FELIZARDO; CARVER, 2020); however, solutions proposed are predominantly and punctual incapable of addressing the problem as a whole. In this context, the sustainability concept was introduced into the SLR domain to propose a new and disruptive vision about how to deal with these problems and is defined as: "*the process and a set of actions to make it possible to preserve SLR that endure over time (i.e., longevity) with less possible time and effort consumed and an effective impact to the industry*" (SANTOS *et al.*, 2021). In this sense, sustainability raises awareness about the role of SLR method within scientific development, additionally, it uses a long-term view and considers multiple perspectives to propose more integrated solutions.

In the SLR domain, sustainability considers three main dimensions (SANTOS *et al.*, 2021): social, economic, and technical. Social dimension addresses the human aspects associated with SLR, such as communication and collaboration, and ensures the effective participation of stakeholders in the SLR process. Economic dimension preserves the resources (time and effort) during the conduction, audition, and update or replication of SLR. Finally, the technical dimension preserves the reliability of supporting tools used to conduct SLR, facilitating their use for the update or re-conduction of those SLR and ensuring technical ways to minimize the efforts by automating tasks and supporting their execution. These dimensions were proposed to illustrate how SLR problems are somehow interconnected and the importance of proposing solutions that are aware of these connections.

Despite sustainability of SLR providing a way to classify and study how these problems could be interconnected, it is still necessary to outline the most sensitive points that researchers should pay attention during conducting/updating of SLR to make them more sustainable. Therefore, the main research question that guided us is: **which are the practices that could leverage the sustainability of SLR?**?

The main goal of this chapter is to provide useful recommendations for SE community to conduct SLR using sustainability-driven approaches. In particular, our main contributions are (i) a snapshot of current practices adopted in SLR; (ii) an analysis of the impacts of current practices on the sustainability of SLR; and, (iii) 13 practical recommendations for researchers that should be used conduct more sustainable SLR.

For this, we established a set of 19 indicators of sustainability based on a list of sustainability characteristics defined in Santos *et al.* (2023), which provided us clues about the current state of the practice of SLR conducted in SE. Next, we selected ten studies published in the last decade in renowned journals from the SE area; following, we carefully analyzed those studies from social, economic, and technical perspectives looking at how problems detected impact sustainability in these dimensions. As the main result, we indicate the flaws in current studies that without the sustainability perspective could not be observed; we highlight key points that are currently being misunderstood or neglected by authors during SLR conduction; finally, we propose as a takeaway of this work a set of 13 practical recommendations that could be adopted immediately that certainly would be beneficial to sustainability in SLR.

The rest of the chapter is organized as follows: Section 5.2 presents the background and related works about sustainability in different areas including SLR. Section 5.3 describes in detail the methods used to depict sustainability in the current SLR. Section 5.4 presents the analysis of our results. Section 5.5 presents the discussions including the major threats to the validity of this study. Finally, in Section 5.6, the conclusions are presented and future work is outlined.

5.2 Background and Related Work

Sustainability was first introduced in the Ecology area mainly to preserve natural resources (i.e., water, air, land) and better comprehend the impacts of human actions on society, environment, and economy (PURVIS *et al.*, 2018b). Over time, sustainability comprehension expanded becoming a multidisciplinary topic that involved ecologists, economists, and sociologists (PURVIS *et al.*, 2018b). These discussions defined a wide range of terms, such as "sustainable development" that states that sustainability must take account of social and ecological factors, as well as economic ones to solve current problems (NATURE; FUND, 1980); just as "sustainable conditions" that are those who ensure the existence of the human race on the Earth for as long as possible (DALY, 1973). Due to the escalation of serious problems such as climate change a need for more effective solutions emerged, hence, in the 1980s, sustainability term gained popularity (PURVIS *et al.*, 2018b) representing a bootstrap for major changes in population mindset that would serve as a baseline for the formulation of environmental policies research (BROWN, 1982).

Advances achieved in the Ecology area drew the attention of the SE community who decided to expand its comprehension of software production and introduce the sustainability concept to better understand its impacts on society, considering that software systems underpin all aspects of societal life from commerce, communication, and education, to energy, entertainment, finance, governance, and defense. Modern society's reliance on software systems has resulted in technology being considered the backbone of virtually all the solutions designed to support sustainability (DEEK *et al.*, 2005).

Sustainable Software comprehension evolved over time, but the first glimpse of this topic started in a special session included in International Conference on Software Engineering - ICSE'2009 called "Software Engineering for the Planet special session"[1]. Several authors in the following years contributed to expanding the sustainability definition in SE, for instance, Penzenstadler *et al.* (2014) defined sustainable systems as (i) the software code being sustainable and agnostic of purpose or (ii) the software purpose being to support sustainability goals, i.e., improving the sustainability of humankind on our planet. In 2015, a manifesto referred to as "The Karlskrona Manifesto for Sustainability Design" (BECKER *et al.*, 2015a) was produced by 13 specialists and signed by more than 250 researchers worldwide and still counting. This document describes the importance of sustainability in SE, underpin the main premises that

[1] ICSE'09 - Special session <https://www.cs.uoregon.edu/events/icse2009/specialSessions/#planet>

sustainability is based on, proposed a set of commitments for software engineers, and attributes responsibilities for stakeholders to get started with immediately to promote sustainability in their projects.

Literature in SE indicates a need for multiple dimensions to address sustainability issues (LAGO *et al.*, 2015b). Despite the number of dimensions may vary among studies, important contributions were made to define them, for instance, in Razavian *et al.* (2014b) and Lago Lago *et al.* (2015b) that define four dimensions, namely: social, environmental, technical, and economic. Later in 2015, Karlskrona Manifesto (BECKER *et al.*, 2015a) considered it important to include individual dimension in the sustainability definition.

Social dimension focuses on ensuring that current end future generations have the same or greater access to social resources by pursuing generation equity (LAGO *et al.*, 2015b). It encompasses the direct support of social communities in any domain, as well as activities or process that indirectly create benefits for social communities. **Environmental dimension** has a strong correlation to the ecology domain, i.e., it revolves around the concept of natural resources and how to make fair use of these resources including electric energy. This dimension is concerned to improve human welfare while protecting natural resources such as water, land, air, and minerals (BECKER *et al.*, 2015a). **Technical dimension** addresses the long-term use of software systems and their appropriate evolution in a constantly changing execution environment (RAZAVIAN *et al.*, 2014b; LAGO *et al.*, 2015b). According to Karlskrona Manifesto, technical sustainability refers to the longevity of information, systems, and infrastructure and their adequate evolution with changing surrounding conditions, including maintenance, innovation, obsolescence, and data integrity, among others (BECKER *et al.*, 2015a). **Economic dimension** focuses on preserving and maintaining capital and financial values (LAGO *et al.*, 2015b; BECKER *et al.*, 2015a). Razavian *et al.* (2014b) believes that economic sustainability is about creating a positive economic value exchange between all the stakeholders participating in software systems. Hence, economic sustainability should consider budget constraints, development, and operational costs, market information when available, and long-term business goals that can relate to one or more functional and/or non-functional requirements (OJAMERUAYE *et al.*, 2016b). **Individual dimension** addresses the well-being of humans as individuals including their mental and physical well-being, education, self-respect, skills, mobility, etc (BECKER *et al.*, 2015a). Furthermore, the individual dimension covers a wide range of topics that can be related to individual freedom and agency, human dignity, and fulfillment. It includes individuals' ability to thrive, exercise their rights, and develop freely (PENZENSTADLER *et al.*, 2018).

Sustainability in SE paved the way for this concept to be adapted into other contexts. Besides, it proved that, is possible to benefit from a systemic vision that covers a wider range of aspects involved in software production abandoning a simplistic view that proved to be harmful to society. The possibility of dealing with such a complex problem drew the attention of researchers that face a similar problem around the production of secondary studies.

The need for sustainability in SLR domain arose meanwhile researchers observed that problems reported a decade ago (e.g, (CARVER *et al.*, 2013)) still impacts current SLR, for example, the fact of SLR consumes much time and effort (FELIZARDO; CARVER, 2020); lack of efficient tools to support SLR (AL-ZUBIDY; CARVER, 2018); or, the lack of effective impact on software industry (CARTAXO *et al.*, 2018). To address these problems, Santos *et al.* (2021) performed a deeper analysis using six important studies published from 2010 to 2020 (AMPATZOGLOU *et al.*, 2019; BUDGEN *et al.*, 2018a; ZHOU *et al.*, 2016; RIAZ *et al.*, 2010b; KITCHENHAM; BRERETON, 2013; IMTIAZ *et al.*, 2013) which reported the main barriers of SLR. This study contributed to highlighting the researcher's difficulties/threats to validity providing a fresh view of the problems and establishing sustainability as a key concept and proposing better ways to deal with SLR. Its results problems of SLR are considered social barriers (mainly associated with human aspects of SLR, such as communication, and culture); economic barriers (associated with the high consumption of time and effort while planning, conducting, or updating SLR); and technical barriers (address problems with tools and technologies used to support SLR conduction).

A deep discussion about what means to be sustainable regarding SLR domain is presented in Santos *et al.* (2023) that outline the main characteristics of sustainable SLR based on more consolidated evidence presented about the sustainable software systems. Santos *et al.* (2023) proposes a set of 15 characteristics and 15 critical factors to build more sustainable SLR. Characteristics reinforce from a sustainability perspective, for example, the importance of providing reliable results and rich documentation, using a process that complies with standards, and adopting techniques such as interactivity and pilot testing. Nevertheless, these characteristics looked beyond a common ground already established for SLR studies, for instance, stating the importance of thinking about SLR using a long-term view, using resources responsibly, maintaining SLR continuously updated, or avoiding that research be wasted. Moreover, sustainability also includes as characteristics the usefulness of results and the importance of providing evidence for a wider community including academic communities and the software industry. Another contribution was the critical factors that state the most sensitive points of SLR have a strong correlation with sustainability. These factors state the importance of (i) communication and collaboration among stakeholders to exchange experience about the research domain and SLR conduction; (ii) wisely management of knowledge generated by SLR mainly creating efficient ways of sharing knowledge and by reusing, refactoring former results; (iii) usage of reliable, mature, and accessible tools; and, (iv) support of efficient management of resources using techniques that mitigate excessive consumption and establish means to alleviate consequences of introducing changes in the project (e.g., interactivity) or even avoiding that unnecessary research be conducted.

Despite the efforts of transposing sustainability into SLR context, this is not enough to make the current SLR sustainable and make it achieves its original goals of impacting society (academia and industry) and supporting the decision-making process. Hence, there is still a lack of comprehension about the current state of practice of SLR and which leveraging points need to be treated immediately to move towards sustainability.

5.3 Research Method

This chapter answers the *RQ4 How to leverage sustainability of SLR?*. An overview of the research method is presented in Figure 10), which is composed of four main phases. In short, the setup phase established the main goals, restrictions, and formal methods to extract information from studies. In the second phase, we performed searches in electronic databases and study selection according to criteria defined in the protocol. In the third phase, we piloted our data extraction form to confirm which information was available, next, the data extraction occurred. Finally, in the fourth phase, we synthesized and reported results including an analysis using a sustainability perspective. The following sections describe in detail each of these activities.

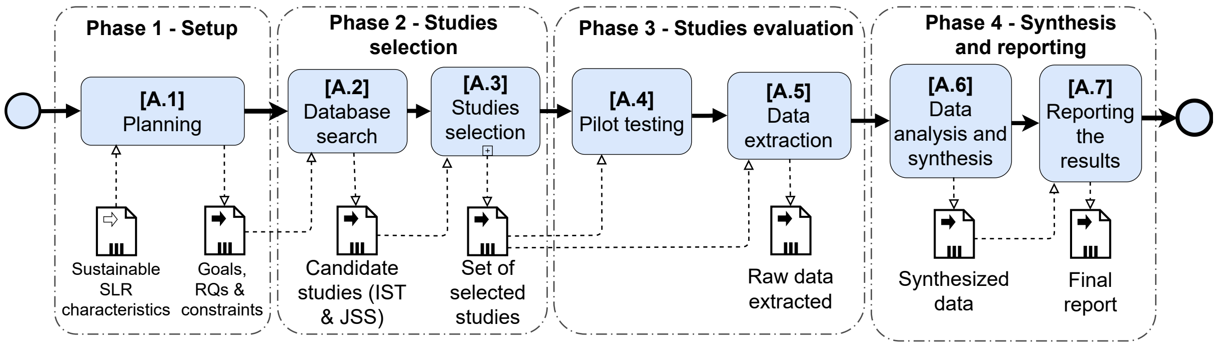

Figure 10 – Research method used to select studies, extract data, synthesize and report results

A.1 - *Planning*

This activity was responsible for establishing a protocol that defined the main goals, research questions, constraints, search strategies, extraction forms, and data extraction/synthesis techniques. To achieve our goals, we defined two research questions:

1. RQ4.1 – What is the current state of practice of sustainability in SLR conducted in SE??

   Rationale: despite SLR having well-established guidelines, its success does not relies only on the observance of these guidelines, instead, but also relies on multiple aspects that impact directly its whole life cycle. This question provides an overview of the studies selected regarding different aspects such as processes, documentation, resource usage, knowledge management, communication, tools, maintenance, and research usefulness/impacts likewise relevant information for analyzing sustainability.

2. RQ4.2 – Which impacts do current practices adopted by researchers have on the social, economic, and technical sustainability of SLR?

Rationale: practices adopted currently by researchers have different types of impact on SLR in its whole life-cycle. This question uses data collected in the previous question to reinterpret them using the social, economic, and technical perspectives to highlight potential threats to the sustainability of these studies observing potential hooks that make these problems somehow integrated.

Two essential constraints must be applied in this study to limit its scope and allow a deeper analysis of results. Firstly, we analyzed only studies that claim to be Systematic Literature Reviews, hence, variations likewise Systematic Mappings (PETERSEN; ALI, 2011), Multivocal Reviews (GAROUSI *et al.*, 2019) or Rapid Reviews (CARTAXO *et al.*, 2018) were not considered in this analysis. Furthermore, we selected as our data source studies published in two major journals, namely Information and Software Technology (IST) and the Journal of Systems and Software (JSS). These journals appreciate and encourage the submission of SLR, consequently, these reviews were approved in a rigorous peer-review process which endorses their reliability.

Regarding data extraction strategy, we designed it to put together information about potential threats to sustainability in current SE studies. For this, we synthesized sustainability characteristics and critical factors proposed by Santos *et al.* (2023) into 19 sustainability indicators using thematic analysis (CRUZES; DYBA, 2011a) i.e., this process juxtaposed keywords to generate a unique set of indicators (For more details, see Appendix A). Henceforth, we adopted these indicators as guidelines to define questions that could provide us clues about how current studies address these topics. Next, based on the aforementioned information we built a preliminary version of our extraction form.

Concerning data synthesis, in short, the main techniques used were thematic analysis to process qualitative data and descriptive statistics for quantitative data. Besides, to facilitate data extraction/synthesis and further visualization, the seven categories proposed by Santos *et al.* (2023) were adopted, namely:

1. **Research process and documentation** - deal with topics about SLR processes followed by researchers during conduction including guidelines and techniques used to document the study.

2. **Resources usage** - include topics about resource consumption during SLR conduction, mainly evaluating how the time and effort of researchers were managed.

3. **Communication and Collaboration** - addresses how internal and external stakeholders exchange information and collaborate while executing SLR tasks.

4. **Knowledge Management** - embrace topics about internal information management and how stakeholders share their experience about SLR process and research domain; and,

external information exchange appraising which techniques are used to provide access to data, improve reusability and avoid wasting information.

5. **Tool Support** - tackle topics about the tools used to support the SLR process highlighting aspects like their accessibility to other researchers.

6. **Update / Maintenance** - covers how studies selected were maintained i.e., whether they were updated, and also evaluates whether their components were reused as a basis for other studies.

7. **Research impacts, usefulness, and long-term goals** - tackle how research work was useful for academia and industry and the efforts taken to present results for practitioners in an understandable way.

Finally, using the classification system aforementioned, indicators were assigned to their respective categories as presented in Table 15.

A.2 - *Database Search*

This activity was responsible for executing the database searches and identifying the candidate studies for selection. For this, we used Science Direct[2] database that is the official search engine of papers published in IST and JSS considering the following metadata fields: title, abstract, and keywords. We used the following search string:*"literature review", "systematic review", "systematic literature review"* that returned a total of 412 candidate studies (see Figure 11) published between 2004 and 2022 which were submitted to the selection process.

A.3 - *Studies Selection*

In this activity, we performed the selection process systematically and extracted a sample to be further analyzed. For this, we first defined a set of two inclusion criteria (IC) and five exclusion criteria (EC) that are described below:

1. IC_1 – Study is a Systematic Literature Review (Systematic Mappings, MLR, Rapid Reviews, and tertiary studies are not allowed);

2. IC_2 – Study is about software engineering or any subarea of SE.

1. EC_1 – Study is not a Systematic Literature Review or is a different type of Secondary study;

2. EC_2 – Study discusses the Systematic Literature Review as a method and does not apply it;

[2] Science Direct: <https://www.sciencedirect.com/>

3. EC_3 – Study is a shorter/older version of another study already included;
4. EC_4 – Record is only a call for papers or a collection of abstracts.

Figure 11 presents the selection process used. In A3.1, we processed metadata resulting in 412 studies, next, in A3.2 filtered these studies considering a ten-year time span resulting in 354 valid studies. In A3.3, we applied the aforementioned selection criteria considering the title, abstract, and keywords, and whenever it was necessary we considered the full text, resulting in 238 studies. Finally, in A3.4 a sample of ten studies was extracted to compose our final set. As a metric, we used the studies with the highest number of citations per year mapped by Scopus database. Our final set of studies is presented in Table 14.

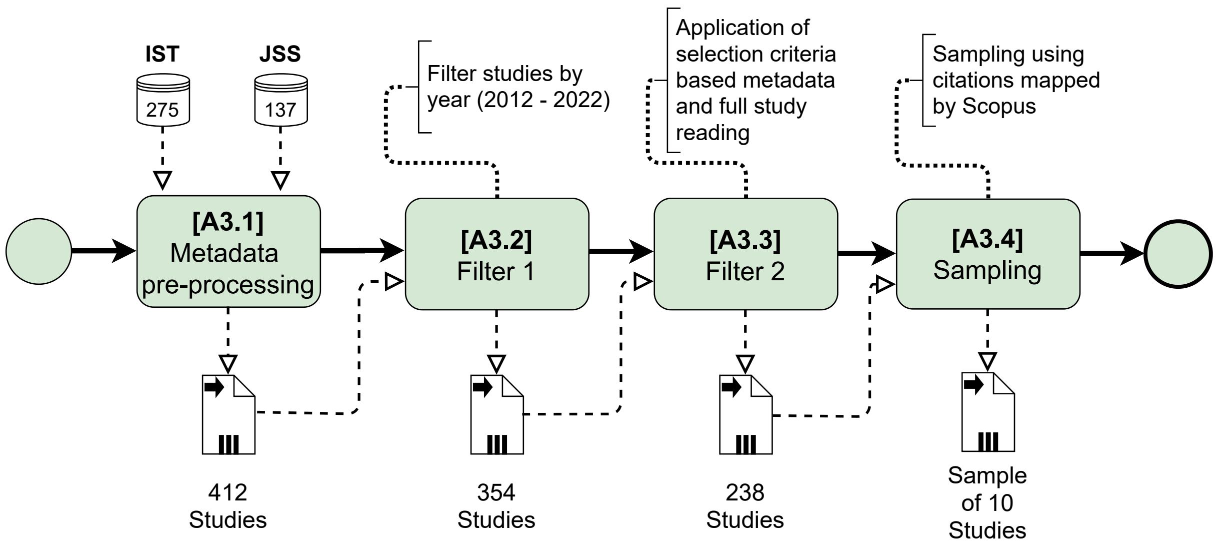

Figure 11 – Details of selection process used

A.4 - *Pilot testing*

This activity is responsible for running a pilot study over the data extraction procedures to understand the capability of our artifacts of capturing evidence about the current scenario of SLR. For this, the preliminary data extraction form generated in the first activity (A.1) was used to gather data from two studies (S1 and S2) from our final set. This activity was performed iteratively i.e., we progressively adjusted the form by adding or removing questions that could not be answered only by reading study documentation. This process was repeated until we define a more stable version of the form amenable to be applied to all studies selected. Table 15 presents the final version of the extraction form.

A.5 - *Data extraction*

In this activity, each study was carefully appraised to better comprehend its motivation, research method, contributions, and relevance. We answered questions defined in data extraction form and explored the studies metadata to find relevant information about authors' profiles, supplementary materials, or any other associated data that could help us to answer our questions. To

Table 14 – Studies selected for analysis

ID	Title	Year	Venue	Citation (scopus)	Citations per Year	Ref.
S1	Systematic literature review of machine learning based software development effort estimation models	2012	IST	301	27.36	(WEN *et al.*, 2012)
S2	Software fault prediction metrics: A systematic literature review	2013	IST	346	34.6	(RADJENOVIĆ *et al.*, 2013)
S3	A systematic literature review of software requirements prioritization research	2014	IST	268	29.8	(ACHIMUGU *et al.*, 2014)
S4	Exploring principles of user-centered agile software development: A literature review	2015	IST	189	23.6	(BRHEL *et al.*, 2015)
S5	Challenges and success factors for large-scale agile transformations: A systematic literature review	2016	JSS	396	56.6	(DIKERT *et al.*, 2016)
S6	Static analysis of android apps: A systematic literature review	2017	IST	163	27,2	(LI *et al.*, 2017)
S7	Test case prioritization approaches in regression testing: A systematic literature review	2018	IST	118	23.6	(KHATIBSYARBINI *et al.*, 2018)
S8	Machine learning techniques for code smell detection: A systematic literature review and meta-analysis	2019	IST	93	23.3	(AZEEM *et al.*, 2019)
S9	A systematic review of unsupervised learning techniques for software defect prediction	2020	IST	69	23.0	(LI *et al.*, 2020)
S10	A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges	2021	JSS	33	16.5	(VACCA *et al.*, 2021)

support data storage, manipulation, and analysis, we generated artifacts using Excel Spreadsheets which are publicly available here [3].

A.6 - *Data Synthesis*

This activity applied summarized data extracted from studies to answer questions defined in our extraction form (Table 15 and 16 – fourth column). For this, we applied Thematic Analysis (CRUZES; DYBA, 2011a) to summarize text passages with relevant qualitative data. Furthermore, we used descriptive statistics techniques (e.g., mean, mode, and standard deviation) to synthesize quantitative data and collect relevant insights. This process has particularities since some items required additional effort to ensure a more reliable answer, these particularities are detailed as it follows.

SLR reliability evaluation should consider that it relies on actions taken in planning, conduction, and reporting. In former studies, MacDonell *et al.* (2010b) used some parameters to evaluate the reliability of SLR, for instance, comparing search strategies (e.g., including

[3] <https://www.doi.dx/1941234>

sources, search terms, documents included/excluded), data extraction approach, and techniques used for data summarization/analysis. An alternative is proposed by Ampatzoglou *et al.* (2019) that provides a comprehensive checklist to evaluate threats that impact SLR and actions that could be adopted to increase the reliability of their studies. To appraise the reliability of our set of studies, we reused former experiences mentioned and adapted the checklist proposed by Ampatzoglou *et al.* (2019) establishing 49 unique actions to increase SLR reliability (presented in Table 17). Next, we appraised each study individually to collect evidence of practices adopted, not adopted, and not informed. For this, we strictly based on documentation and supplementary files to comprehend the reliability of SLR based on double-checking mechanisms implemented. Results are presented in Section 5.4.1.1.

A particular challenge of this chapter was to evaluate resource consumption since few details are provided in the documentation. We used three items available in the final reports that were used to measure the effort level needed to conduct the study, they are (i) a number of candidate studies processed; (ii) relevant studies analyzed during SLR conduction; and, (iii) techniques used to minimize the efforts within most time-consuming activities (results are available in Section 5.4.1.2).

Collaboration/communication among researchers or previous knowledge about the research domain is usually not explicitly asserted in the final report. Hence, we used all information publicly available in the public profiles of authors and all metadata of studies to gather more evidence about stakeholders' participation. We analyzed 43 public profiles (using DBLP [4] and Google Scholar [5]) collecting 3789 studies that were used to evaluate the experience of researchers in scientific research. Using title, abstract, and keywords we filtered those which were secondary studies (i.e., Systematic Mappings, SLR, Grey Literature Reviews, etc.) to comprehend researchers' expertise in SLR conduction. Additionally, we applied the classification proposed by Budgen and Brereton (2022) which uses three categories: inexperienced (no previous secondary studies conducted); limited experience (between 1 and 5 studies published); and, Experienced (more than five studies published). Results of this analysis are available in Sections 5.4.1.3 and 5.4.1.4.

Evaluation of SLR update was performed using a Forward Snowballing (FS) technique (MOURÃO *et al.*, 2020) to identify possible updates derived from our set of studies. For this, we used the Scopus database since more generic search engines are considered a more suitable option due to their coverage (FELIZARDO *et al.*, 2018). Scopus mapping identified a total of 2102 references, unfortunately, 111 studies were unavailable/inaccessible, hence, our classification process started considering 1991 studies. First, we removed 39 duplication's resulting in 1952 unique papers. Next, we evaluated those studies using the title, abstract, and keywords looking exclusively for SLR, in this step 514 studies remained. Next, a full-text

[4] <https://dblp.org/>
[5] <https://scholar.google.com/>

analysis was conducted remaining only 284 studies, finally, after analyzing these papers only three papers (ALI; GRAVINO, 2019; TIEPPO *et al.*, 2021; CASTRO-CABRERA *et al.*, 2020) were considered updates and the results of this analysis are presented in Section 5.4.1.6.

Usefulness of SLR studies for academia and industry is appraised using two metrics: (i) Number of citations/year received over years; (ii) presence of practical recommendations to support practitioners in applying knowledge summarized into practice. It is worth mentioning that citation per year was already used by Garousi and Fernandes (2016) to quantify the usage of the study. They collected data from 71,668 SE papers indexed by Scopus [6], aiming to analyze the top 100 most cited SE papers. Their results presented that highly cited papers have an average citation value was 6.82 per paper. This metric was also used by Mendes *et al.* (2020b) to evaluate how much interest the community has in this topic, consequently, its impact and usefulness for the development of the research domain making it a cut-off parameter to decide the feasibility of update. Finally, another parameter used to appraise usefulness is the presence of recommendations for practitioners. Results of this analysis are presented in Section 5.4.1.7.

A.7 - *Reporting the Results*

This activity was responsible for creating a comprehensive report for data synthesized to disseminate results to the audience.

5.4 Results

Studies selected addresses topics that grab much attention from SE community, such as: Machine Learning (S1, S8, S9), Software Metrics (S2), Software Requirements (S3), Agile (S4, S5), Mobile Apps (S6), Software Testing (S8), Blockchains and Smart Contracts (S10). We assume that these studies followed the best practices for planning, conducting, and reporting SLR, and were peer-reviewed by experts that compose the editorial team of two journals respected journals SE area. Moreover, their quality was endorsed by the community which widely used its results making undeniable the relevance and influence of these studies in their area.

5.4.1 RQ_1 – State of the Practice of SLR in SE

This research question is answered using the seven categories defined in the planning stage, which are discussed in detail as follows.

5.4.1.1 Research Process and Documentation

Regarding the guidelines adopted by researchers to conduct their SLR, Kitchenham *et al.* (2004a), Kitchenham and Charters (2007) is still considered a common basis for studies in our sample

[6] <https://www.scopus.com>

Table 15 – Extraction form used based on sustainability indicators

Cat.	SI	Questions	Data Extracted
Research process documentation	**SI1**	**Compliance with standards defined for conduction/update**	
		Q1.1 - Which guidelines do authors claim to use?	SLR guidelines
	SI2	**Iterativity and pilot testing**	
		Q2.1 - Does the process that was followed contains iterations? Q2.2 - Does the pilot test was conducted? Q2.3 - Which SLR stages were refined in the pilot test?	Yes/No Yes/No SLR stages refined
	SI3	**Documentation quality**	
		Q3.1 - Which guidelines SLR documentation were used?	Guidelines for documentation
	SI4	**Study reliability**	
		Q4.1 - Which actions to ensure reliability were applied? Q4.2 - Does a quality assessment was conducted?	Actions adopted Yes/No
Resources usage	**SI5**	**Resources usage**	
		Q5.1 - Is the study produced with responsible use of resources making use of techniques that minimize resource consumption? Q5.2 - How many studies were reviewed and analyzed?	Yes/No + techniques Number of studies reviewed/analyzed
Communication and Collaboration	**SI6**	**Communication among stakeholders**	
		Q6.1 - Any methods/strategies used to support rich communication among researchers, share knowledge, and solve conflicts?	Communication methods/strategies
	SI7	**Participation/collaboration of stakeholders**	
		Q7.1 - How research team is composed? Q7.2 - How stakeholders contributed?	Authors details Stakeholders contributions
Knowledge Management	**SI8**	**Knowledge of stakeholders about the research domain**	
		Q8.1 - How much experience research team have?	Number of published papers
	SI9	**Experience of team members in SLR conduction**	
		Q9.1 - How much experience research team have in SLR conduction?	Number of secondary studies conducted
	SI10	**Knowledge sharing/transfer**	
		Q10.1 - Any method to share knowledge between team members was used?	Knowledge sharing techniques
	SI11	**Accessibility of SLR artifacts**	
		Q11.1 - A protocol is available including all its versions generated in SLR packaging? Q11.2 - Does the study provides a complete replication kit containing raw data to be audited/replicated Including, for instance, included/excluded studies (with motivations for exclusion, agreement), raw data extracted, and data synthesis?	Yes/No Yes/No
	SI12	**Research waste**	
		Q12.1 - Do authors avoid research waste by prior undertaking an SLR, researchers evaluated whether an updated SLR already exists on the same topic?	Yes/No
	SI13	**Improvement of SLR reusability**	
		Q13.1 - Did the study reuse/refactor any element of the previous SLR (e.g., search string, selection criteria, and other elements like raw data)?	Yes/No + refactored elements

Legend: Cat - Category; SI - Sustainability Indicator. **Continues...**

(S1, S2, S3, S4, S5, S6, S7, S8, S10), additionally, S4 and S9 also mention the usage of two additional studies that evaluate SLR conducted in SE (KITCHENHAM *et al.*, 2009) and update the original guidelines (KITCHENHAM *et al.*, 2015). Besides, S2 and S4 also considered recommendations and lessons learned from Brereton *et al.* (2007), Brocke *et al.* (2009), Staples and Niazi (2007), Webster and Watson (2002). Finally, only S8 adopted more specific guidelines for snowballing (WOHLIN, 2014b).

Our analysis showed that iterativity was mentioned only in three studies (S2, S4, S5) and most of the refactoring occurred during the initial stages of SLR. Besides, only two studies

Table 16 – Continuation of extraction form used based on sustainability indicators

Continuation of Table 15			
Cat.	**SI**	**Questions**	**Data Extracted**
Tool Support	**SI14**	**Usage of tools to support SLR**	
		Q14.1 - Which tools were used?	Tools used
	SI15	**Accessibility of support technology**	
		Q15.1 - Tools used are available for usage? Q15.2 - Are tools used free? Q15.3 - How tools used are maintained?	Yes/No Yes/No Tools details
Update / Maintenance	**SI16**	**Continuous update**	
		Q16.1 - Was the study updated?	Yes/No
	SI17	**Components reuse**	
		Q17.1 - Does study components were used as a basis for other SLR (replications)?	Yes/No
Research impacts, usefulness, and long term goals	**SI18**	**Research usefulness and impacts over community**	
		Q18.1 - Do authors present practical recommendations that are useful for a wider community (researchers and SE practitioners)?	Yes/No + practical recommendations
	SI19	**Long-term goals and research impact over time**	
		Q19.1 - How many citations the study received in total and per year?	Number of citations & Number of citations per year
Legend: Cat - Category; SI - Sustainability Indicator.			

claimed explicitly the usage of pilot testing in their review process (S1, S2). It is worth mentioning that S5 and S7 authors mentioned the conduction of "preliminary searches" which by itself does not fully comply with the rigor required of this kind of testing, however, we recognize that this assertion may indicate the informal conduction of pilot testing. Concerning parts of the protocol refined through pilot testing, the search string is the most mentioned item (S2, S3, S5, S7). Moreover, other items like selection criteria (S1, S2), database selection (S7), and data extraction strategy (S1, S2) were also refined through pilot tests. Despite results showing that researchers are using pilot tests in compliance with guidelines (KITCHENHAM; CHARTERS, 2007) that state the importance of evaluating mainly study selection criteria and data extraction forms before proceeding with SLR, we notice that there is still a lack of adoption of these tests within our sample since six studies (S3, S4, S6, S8, S9, S10) do not mention any kind of testing in their process.

More specifically about documentation, all studies have used Kitchenham's guidelines (KITCHENHAM *et al.*, 2004a; KITCHENHAM; CHARTERS, 2007; KITCHENHAM *et al.*, 2009; KITCHENHAM *et al.*, 2015) to document their SLR. In addition, only S4 adopted more specific recommendations for documentation proposed by Brocke *et al.* (2009). Although all reviews in our studies set had good quality documentation, there is a discrepancy regarding the detail level in reported items. For instance, the identification and selection of primary studies are often quite detailed, at the same time, other processes that are equally important (e.g., data extraction and quality assessment) are described with fewer details. Good examples are found in S2, S4, and S10 which provided many details about the classification used in the data extraction, meanwhile in S3 and S7 fewer details are provided and readers only understand how data was summarized while reading the results section. Despite more updated recommendations have been published to improve SLR documentation e.g., (BUDGEN *et al.*, 2018a; KITCHENHAM

Table 17 – Actions taken to increase reliability

ID	Actions	Studies
A1	Selection of the most-known digital libraries or specific publication venues or usage of broad search engines or indices	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10
A2	Documentation of inclusion/exclusion criteria been explicitly in the protocol	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10
A3	Usage of snowballing	S1, S2, S3, S4, S5, S6, S8, S10
A4	Prescription decision rules set for study inclusion/exclusion	S2, S3, S4, S6, S7, S8, S9, S10
A5	Usage of a strategy for systematic search string construction	S1, S2, S3, S4, S5, S7, S8, S10
A6	Exhaustive search for related work to (a) familiarize with the field, (b) identify comparable studies, and (c) identify relevant publication venues and influential papers	S2, S3, S4, S5, S6, S7, S8, S10
A7	Conduction of article quality assessment as inclusion/exclusion criterion	S1, S2, S3, S7, S8, S9, S10
A8	Involvement of more than one researcher	S1, S2, S3, S5, S6, S8, S10
A9	Usage of statistics to deal with quantitative data for answering research questions	S2, S3, S6, S7, S8, S9, S10
A10	Selection of existing initial classification schema	S1, S2, S4, S5, S6, S9, S10
A11	Development of a consistent strategy (e.g., keep the newer one or keep the journal version) for selecting which study should be retained in the list of primary studies	S1, S3, S5, S6, S7, S9
A12	Discussion about inclusion/exclusion of selected articles in case of conflict	S1, S2, S5, S6, S8, S10
A13	Definition of quality thresholds for inclusion/exclusion	S1, S2, S3, S6, S8, S9
A14	Execution of paper screening to cross-check data extraction	S1, S2, S3, S6, S8, S10
A15	Usage of broad search process without an initial starting date	S2, S3, S5, S6, S7, S8
A16	Comparison of findings and compliance with those of existing studies	S4, S5, S6, S7, S8
A17	Usage of broad search process in generic search engines or indices (e.g., Google Scholar) to ensure the identification of all relevant publication venues	S1, S3, S7, S9
A18	Usage of tools to facilitate the review process	S3, S4, S6, S9
A19	Evaluation of search results and documentation of the outcomes	S2, S4, S8, S9
A20	Continuous update of classification schema until it becomes stable and capable of classifying all primary studies in one or more classes	S2, S4, S5, S6
A21	Usage of a formal data synthesis method	S1, S4, S5, S8
A22	Execution of random screening of articles among authors	
A23	Manual selection and scan venues to check if they publish articles related to your secondary study	S1, S6, S8
A24	Usage of scientific quality of primary studies to draw conclusions	S4, S8, S9
A25	Provision of public access to all gathered data	S6, S8, S9
A26	Discussion and brainstorm with authors to reach possible interpretations of the findings when there is a lack of related studies	S3, S5, S7
		Continues...

et al., 2022), in our sample they were not used.

In Table 17 we present the ten most used MA's whose five refer to study selection validity indicating that improving SLR coverage was a priority for analyzed studies. Similarly, data validity mitigation actions were also used in studies analyzed aiming to ensure the quality of studies and avoid misinterpretation of results using more than one researcher to perform analysis and apply statistical techniques. Finally, an important indicator of reliability is the fact that 9 of 10 studies analyzed conducted some kind of quality assessment, consequently, preventing biases derived from primary studies impact critically the results.

5.4.1.2 Resources Usage

Regarding the number of studies selected by manual searches and in electronic databases, the values range from 707 to 181,829 studies. This variation drew our attention because most of the

Table 18 – Continuation of actions taken to increase reliability

Continuation of Table 17		
ID	**Actions**	**Studies**
A27	Execution of pilot searches to train your search string	S2, S5
A28	Comparison of primary studies list to a gold standard or to other secondary studies	S2, S5
A29	Usage of summaries of candidate primary studies to guarantee the correct identification of all duplicate articles	S3, S6
A30	Discussion of inclusion/exclusion criteria revising them after pilot studies or experts' suggestions	S1, S2
A31	Selection of variables among authors to guarantee that the research questions can be answered	S1, S2
A32	Appraisal of primary studies validity and their impact applying statistics	S1, S9
A33	Execution of pilot data extraction to test agreement between researchers	S1, S2
A34	Holistic discussion and brainstorm about research questions coverage regarding the goal of the study	S3, S4
A35	Revision of independent experts over search process	S2
A36	Comparison of the number of primary studies in different languages with the population	S8
A37	Comparison of the number of studies with missing full texts with the population	S8
A38	Quantification of experts' disagreement with the kappa statistic	S2
A39	Inclusion of grey literature	S2
A40	Execution of pilot data extraction to test the existence of relationships	S1
A41	Execution of pilot data analysis and interpretation	S2
A42	Conduction of reliability checks like post-SLR surveys with experts	S4
A43	Discussion of the research method used (SLR or SMS) to fit the goals/research questions of the study and justification of the purpose and scope of the methods	S7
A44	Usage of systematic voting	-
A45	Usage of sensitivity analysis	-
A46	Usage of experts or external reviewers' opinion in case of conflicts	-
A47	Development, supervision, and documentation of the protocol and their possible deviations	-
A48	Consultation of the target audience for setting up study goals	-

studies are composed of small teams (4 or 5 members) and predominately they did not mention the usage of tools to support nor automation techniques indicating that possibly much manual effort from researchers was consumed. Meanwhile, the number of studies considered relevant to be deeply analyzed was somehow more stable ranging from 15 to 124 studies which seems more reasonable since data collection can be very time and effort-consuming depending on the depth, and the amount of data to be extracted, tabulated, and processed. Nevertheless, as this number increases more efforts are needed to mitigate biases inserted by researchers' interpretations. We can find evidence of these additional efforts by observing that seven studies involved more than one researcher in the process (S1, S2, S3, S5, S6, S8, S10), six cross-checked data extraction among authors (S1, S2, S3, S6, S8, S10); and, nine studies conducted quality assessments over primary studies (S1, S2, S3, S4, S5, S7, S8, S9, S10).

Only two studies mentioned techniques to reduce the consumption of efforts while conducting the SLR. In S2, the authors said that they limited the use of human efforts in data extraction and quality assessment by checking only 10 randomly selected studies. In S6, the authors said that they have used Python scripts to help process data extracted from the databases, in addition, they mentioned that the workload was balanced with the authors of the work, making resource management more efficient. Broad searches in SLR are acceptable and often recommended to attend to the need for coverage. However, given the laborious nature of review studies, and the fact that SLR analyzed in this study were conducted by small teams, it is difficult to comprehend how these large amounts of data were processed without the use of tools. The lack of details about which techniques were used hindered us to comprehend exactly the amount of effort consumed in the process.

5.4.1.3 Communication and Collaboration

Analyzing team composition, all 43 are formally affiliated with some universities or research centers, and only one is affiliated with both university and industry. Concerning external stakeholders collaboration in protocol elaboration or other steps of the process, the only evidence available is presented in S6 when the authors contacted primary studies authors to self-check their data extraction and point out any inaccuracies.

To support collaboration among researchers the most popular technique was consensus meetings (S1, S2, S3, S5, S6, S8, S10), nevertheless, this statement is quite generic since it does not provide clues about which and how decisions were taken. In some cases, studies combined consensus meetings with more pragmatic techniques, such as Kappa Coefficient (PÉREZ *et al.*, 2020) (S2) or Krippendorff Alpha Kr_a (KRIPPENDORFF, 2018) (S8) or even data extraction cards (S1). In three studies nothing is mentioned about how authors communicated internally, being possible that it still occurs informally or is very limited to some participants, which makes this aspect a threat to validity that is omitted by authors.

In S2 and S4 reported team participation through the suggestion of important studies or in activities related to SLR planning, such as, database selection, piloting selection criteria, or protocol review. In conduction stage, researchers reported more intensive team collaboration in studies selection (S1, S2, S5, S6, S8), data extraction (S1, S2, S3, S6, S8) and quality assessment (S1, S3, S8). In reporting stage only S10 stated that all authors collaborated to write the final report.

5.4.1.4 Knowledge Management

Our results revealed that the research teams were predominantly heterogeneous regarding their experience in research work. In eight studies analyzed (S1, S2, S3, S4, S5, S8, S9, S10) the first author had the lowest number of studies published (avg. of six studies published) indicating a relative lack of experience. At the same time, the remainder of the authors were more experienced (avg. 194 studies published) indicating that most of the SLR incorporated into the research team experienced researchers to avoid bias in the process. Our results are in consonance with those presented in Budgen and Brereton (2022) observed the number of inexperienced authors conducting SLR increased over time due to the value of beginning postgraduate study by conducting a formal literature review. Furthermore, experienced researchers are often involved as additional authors in studies led by inexperienced authors, probably being a combination of student and supervisor in postgraduation.

Concerning the experience of authors in SLR conduction, our results revealed that 10 (out of 43) authors could be considered experienced in SLR conduction and the remainder of the 33 authors were considered with limited experience in SLR conduction. Additionally, we noticed that eight studies (S2, S3, S4, S5, S6, S7, S8, S9) have in their team at least one author which was considered experienced in SLR conduction, once again, reaffirming the heterogeneity of research teams.

Few details were provided about internal knowledge sharing/transfer among researchers. Consensus meetings were used by 7 out of 10 studies (S1, S2, S3, S5, S6, S8, S10) as the main technique to exchange information, though no additional methods to exchange information internally or externally to the research group were mentioned. Another indicator of sustainability analyzed was the information exchange with future generations of researchers by analyzing the availability of SLR artifacts which are essential for updates, replications, and audition of results. In this sense, the availability of the SLR protocol including all versions and clear statements of changes that occurred during the review is essential to ensure the exchange of know-how about the process executed. Nonetheless, in all studies, the protocol is presented only in its final version and no additional details are provided in external repositories, supplementary materials, or associated technical reports. This can be considered a problem when compared to SLR conducted in other areas (e.g., medicine), whose researchers are encouraged to register their research protocol in open science tools like PROSPERO[7], or Open Science Framework[8]. Hence, these approaches may foster researchers' collaboration, avoid wasting efforts on research that is already being conducted, and reuse elements from other reviews. Additionally, we noticed that raw data provided by authors are often incomplete and include only few artifacts generated by SLR conduction, for example, all studies included the list of studies selected for data analysis, but none provided a complete list of excluded studies which makes it more difficult to reproduce, update or audition of the study.

Our analysis showed that few studies reported the reuse of previous SLR elements. Most of the reuse is related to punctual artifacts of SLR, for instance, in S4 authors reused a set of keywords and S7 reused a database set. Only in S8 authors used previous results to compare their findings and draw more reliable conclusions. In spite of the results indicating a lack of reuse of elements, we noticed that researchers are aware of the importance of not duplicating SLR studies. Our analysis concluded that nine studies (S1, S2, S3, S4, S5, S6, S7, S8, S10) assert that there are no previous SLR already existing addressing the same topic. However, a common flaw in studies is a lack of more concrete evidence i.e., a systematic search to ensure that authors have not missed any similar SLR that could be reused.

[7] <https://www.crd.york.ac.uk/PROSPERO/>
[8] <https://osf.io/>

5.4.1.5 Tool Support

During our analysis, we noticed that only four studies mentioned the usage of some tools to support the review process. In S6 authors used Python scripts to support collecting studies from databases; S4 used MAXQDA2 [9] to process qualitative data; S9 claimed to use R scripts to support data analysis and Mendeley [10]; and, S3 claimed to use a visualization tool without mentioning which one. Considering that the studies analyzed treated a high amount of information, probably the six studies that did not mention the usage of tools to support the review (S1, S2, S5, S7, S8, S10) omitted this information in the final report.

Accessibility of techniques used in S6 and S9 is questionable since they mention the use of programming languages (Python and R) to support in SLR process, which may represent a barrier to future researchers due to the long learning curve of coding techniques to reproduce the results. The adoption of MAXQDA2 to summarize qualitative data is more accessible since this tool was validated by empirical studies with positive remarks about its easy and interactive design (SAILLARD, 2011), additionally, it has more solid documentation (KUCKARTZ; RäDIKER, 2019; GIZZI; RÄDIKER, 2021) to support researchers to use it. However, it is worth mentioning that MAXQDA2 is developed and maintained by a private company and operates only under a paid license i.e., it is not freely available for everyone which may pose as a barrier to updating/reproduction. Other tools like Mendeley are also supported by private companies (Elsevier), but it is free for usage. Meanwhile, Python and R are open-source programming languages i.e., they are free for use and are maintained by the community. Despite the existence of more specific tools to support the SLR process (MARSHALL; BRERETON, 2015), there is evidence in the literature that these tools are not used and researchers prefer to use more generic tools like google/excel spreadsheets, Jabref, Mendeley, etc. (AL-ZUBIDY; CARVER, 2018), which is in consonance with our results.

5.4.1.6 Update/Maintenance

Forward snowballing process identified over two thousand candidate studies that after a classification process two studies (ALI; GRAVINO, 2019; CASTRO-CABRERA *et al.*, 2020) remained as updates.

Ali and Gravino (2019) performed an SLR based on the empirical studies published in the time period of January 1991 to December 2017. Authors mention that their coverage of studies allows them to update results already provided by Wen *et al.* (2012). Due to the substantial amount of work done in the last 7 to 8 years based on software effort estimation using ML methods, the main objective of Ali and Gravino (2019) is to review studies that used and discussed the software effort estimation models built using ML techniques outlining conclusions that changed in further years. Five research questions proposed by Wen *et al.* (2012)

[9] MAXQDA website: <https://www.maxqda.com/>
[10] Mendeley website - <https://www.mendeley.com/>

were maintained and some additional ones are proposed to provide a new analysis of trends of applying ML techniques for effort estimation.

Castro-Cabrera *et al.* (2020) perform a systematic literature review (SLR) on the test case prioritization (TCP) that updates the results provided in S7. Authors provide the latest developments in TCP specifically using the taxonomy proposed in S7 including the most important publications from 2017 to 2019. Results showed an increasing interest in the topic analyzing 320 papers published in this short period of time dealing with many different TCP approaches.

Despite the last years providing more guidelines to update SLR (e.g., (WOHLIN *et al.*, 2020; MENDES *et al.*, 2020b; FELIZARDO *et al.*, 2018; GARCÉS *et al.*, 2017)) both studies (ALI; GRAVINO, 2019; CASTRO-CABRERA *et al.*, 2020) analyzed that claim to update S1 and S7 respectively do not use a systematic approach to execute the update. According to Nepomuceno and Soares (2019) SLR updates should justify any changes in the original SLR protocol, in addition, they should be made only when really necessary. In both studies analyzed, there is a lack of information about the update process, Ali and Gravino (2019) only state to cover the same period and answer the same research questions and it is not clear which modifications were done in the protocol. Meanwhile in Castro-Cabrera *et al.* (2020) authors state that the years covered in their review are complementary to S7, but it does not describe how much of the former study was reused.

Our investigation also revealed that the studies selected inspired somehow at least 284 other SLR (not including Systematic Mappings, Ad-hoc literature reviews, or Grey Literature Reviews). These inspirations are presented in different manners across studies, for instance, Tieppo *et al.* (2021) performed minor adaptations in protocol proposed by S1, and Pan *et al.* (2021) claims the whole research process was inspired by S5. Meanwhile, other studies just mention studies in related work or reused smaller components of previous SLR, for instance, S1 inspired the search strategy (IDRI *et al.*, 2016), data extraction (IDRI *et al.*, 2016; PAPAVASILEIOU *et al.*, 2021), quality criterion (IDRI *et al.*, 2016; PAPAVASILEIOU *et al.*, 2021), and results analysis (IDRI *et al.*, 2016). Nevertheless, we did not consider the reuse of components as updates as concluded by Nepomuceno and Soares (2019).

5.4.1.7 Research Impact and Usefulness

Table 19 presents that the average of citations per year ranged between 16.5 and 56.6 citations, which means that studies analyzed are far beyond from cut-off metric previously established by Garousi and Fernandes (2016) making them useful for their respective areas. Another important aspect is the timespan that information remains useful for SE community. Overall, studies published remain widely used by the academic community over the years

Table 19 – Studies citations over years

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Mean
S1	1	9	9	17	29	27	32	47	**52**	46	32	27,4
S2	-	6	12	33	30	32	46	**57**	50	45	35	34,6
S3	-	-	10	14	20	35	42	**49**	30	34	34	29,8
S4	-	-	-	5	18	29	**34**	23	33	26	21	23,6
S5	-	-	-	-	3	26	50	64	82	**90**	81	56,6
S6	-	-	-	-	-	6	16	34	34	**45**	28	27,2
S7	-	-	-	-	-	-	6	24	24	31	**33**	23,6
S8	-	-	-	-	-	-	-	7	24	27	**35**	23,3
S9	-	-	-	-	-	-	-	-	7	28	**34**	23,0
S10	-	-	-	-	-	-	-	-	-	14	**19**	16,5

indicating that SLR has a long-term impact on SE the community, and its information remains as a knowledge base for future researchers. In addition, some studies (S1, S2, S3, S5) achieved their peak of citations between 5 and 8 years after being published, after that, there is a tendency for decay of citation numbers indicating that an update may be necessary.

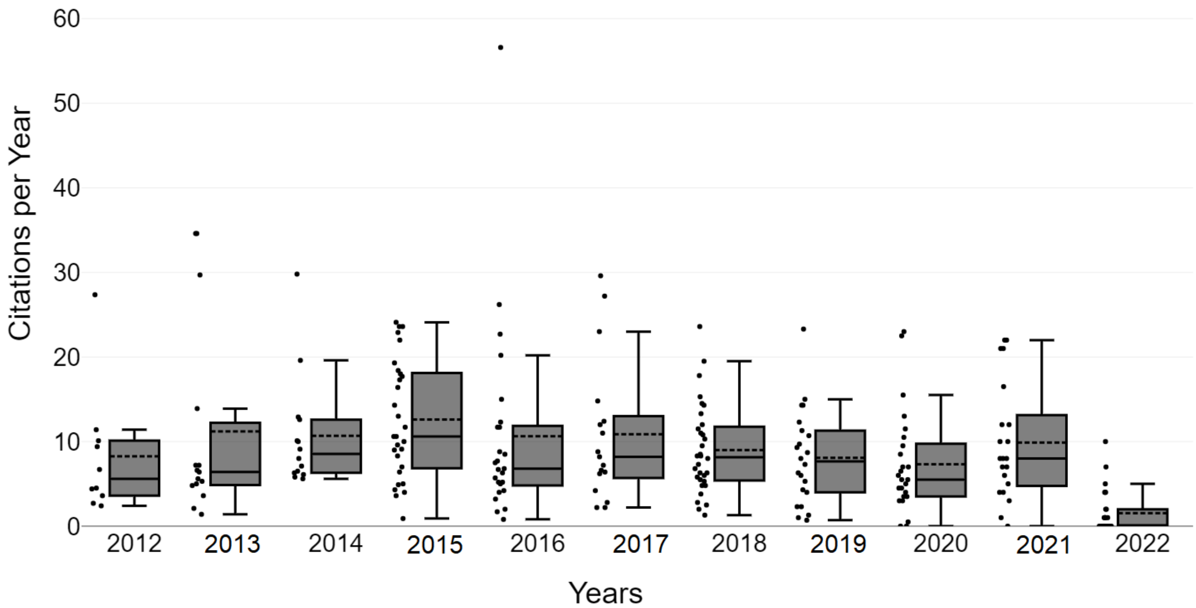

Figure 12 – Overview of citations/year of studies between 2012–2022

Our sample contains relevant studies published in the SE area which proved their value to the academic community through their mentions in further investigations. However, it is important to recognize that our sample may not represent the entire population of SLR studies published in the IST and JSS journals. To mitigate this bias, in Figure 12 we present the distribution of citations per year of our complete population of SLR (238 studies). We can observe that SLR studies receive an average of 9.28 citations per year, therefore, studies selected (S1 - S10) stand out as outliers, because they significantly deviate from the average citation/year. Furthermore, we identified that 40.7% (97 out of 238 studies) still remain below the cut-off metric proposed by Garousi and Fernandes (2016), therefore, improvements are needed to enhance the utilization of SLR in SE.

Finding clear practical recommendations for industry practitioners was quite difficult for most of the studies. Only in S1 a section was especially dedicated to providing insights for practitioners making it clear that researchers have the industry as a target audience. Overall, studies dedicated a few text passages to communicate with practitioners e.g., S2 suggested useful studies and included in their classification scheme means for flagging research created from industry data that may be relevant for these practitioners. Similarly, S3 mentioned that results provide insight for both academia/industry; S4 claimed that their results might contribute to support practitioners applying techniques described in practice delivery of useful and usable software; and, S5 defined its research questions believing that they represent the viewpoints that are likely to provide insights to practitioners as well as researchers. In the remainder of the studies, there are no clear indications about which piece of evidence is particularly valuable for practitioners, however, it doesn't necessarily mean that the results provided are useless, but the final report does not communicate with this audience.

5.4.2 RQ_2 – Impacts of Current Scenario in Social, Economic, and Technical Sustainability of SLR

To apply the sustainability vision, the following sections analyze our results into three different dimensions: social, economic, and technical.

5.4.2.1 Social Perspective

To preserve the human aspects, the social dimension proposes to better comprehend how communication and collaboration occurred throughout SLR process. Firstly, we observed that a potential threat is a lack of diversity in the research team regarding their composition of academics and industry professionals. Despite there are few evidence available, we can observe that a low number of team members are associated with the software industry, additionally, studies did not report concerns about collaboration with external stakeholders. In this scenario, we have a majority of studies with a research team possibly biased with the academic mindset and this lack of diversity can make research design attend only a limited community instead of breaking through barriers and making results more useful for the entire community.

Communication among researchers is reported through generic techniques in SLR report, for instance, consensus meetings or brainstorming sessions, only in a few cases, more specific techniques to achieve consensus (e.g., Kappa Coefficient) were used. Although these efforts remain valid communication remains only a secondary aspect that is poorly documented. Similarly, few details are provided to describe the collaboration among team members. Most of the studies report more intense collaboration within manual tasks mostly to avoid bias (e.g., studies selection, data extraction), however, we did not find concrete evidence of collaboration that adds value to SLR, for instance, defining relevant questions, interpreting results, and validating insights. An approach to document the collaboration among researchers was CreDiT which started to be discussed in 2012 attributing to each author their contribution, distinguishing the role of each stakeholder, and mitigating possible conflicts of interest, however, only one paper (S10) used this taxonomy despite the encouragements made by journals (e.g., IST). In addition, since CreDiT covers multiple types of studies and this statement often provides only generic information it hampers readers to draw precise conclusions about the collaboration of stakeholders. Hence, we believe that a more refined version of this taxonomy could be designed for SLR studies. Foremost, SLR has a social role to bring more reliable evidence to stakeholders, however, this social aspect seems to be neglected while analyzing our dataset.

Conduction of pilot studies, for instance, is a good opportunity to interact with team members, validate the research work, and overall exchange experiences, however, only three studies reported the conduction of pilot studies and iterations. In a restricted view, this information does not invalidate the results from ones that didn't use these techniques even though evidence exists about its benefits (e.g., (FELIZARDO *et al.*, 2017a; FABBRI *et al.*, 2013)). Nevertheless, in a wider view not using pilot studies impacts the social aspect since the research team lost opportunities to interact with experienced researchers who certainly have a deeper comprehension of the research domain. Since many of the current SLR are conducted by post-graduate students (BUDGEN; BRERETON, 2022), interacting with experienced researchers represents an opportunity to catalyze its learning process and weave new and successful collaborations.

A main concern of the social dimension in the SLR domain should be supporting researchers to improve internal their processes to foster rich communication and also consider the social impacts derived from their actions while conducting the study. In this sense, SLR reports play an important role in disseminating information, however, most of the studies analyzed do not provide open access to artifacts generated in SLR conduction limiting the access to information and impacting negatively knowledge sharing making it harder to reuse, audit and update this research. Therefore, the social dimension should also raise awareness in the community that the impacts of decisions taken in SLR studies are not restricted to the research team, but can propagate to a wider community over a long period of time.

5.4.2.2 Economic Perspective

Preserving SLR economically means reinforcing actions that maintain/improve its quality while reducing the resource consumption (mainly time and effort) taken in its whole life cycle. For this, it is necessary that authors apply the best practices available for SLR and update their processes to comply with novelties that could, in parallel, improve SLR quality and efficiency. A threat to the economic aspect is wasting resources dealing with problems that were already solved or mitigated, for instance, using search mechanisms that are inefficient instead of adopting better ones.

We observed in our analysis that practices adopted prioritize to increase reliability, coverage, and quality of results usually implementing double-checking mechanisms to ensure data correctness. Nevertheless, very few techniques were adopted to diminish the number of resources spent in the process. In this sense, the lack of evidence about the use of techniques, tools, or even task automation raises some doubts, for instance, about the reliability of results since manual work is more susceptible to errors. Additionally, the number of candidate studies in SLR analyzed ranged between hundreds to hundreds of thousands of studies, hence, from an economic perspective the implementation of double-check mechanisms with specialists become a bottleneck for this method and duplicates the consumption of resources.

Economic dimension should encompass a long-term perspective and consider that decisions made during SLR planning, conduction, and reporting could impact current and future generations of researchers. That way, it is economically unacceptable that research is conducted due to flaws in earlier research or due to the unavailability of generated artifacts. In studies analyzed the need of avoid duplicating research seemed to be clear for authors since most of them invested time looking for previous works and justified why they are conducting a new SLR. However, the poor documentation of the current SLR should be better discussed. We noticed that Kitchenham and Charters (2007) is still the most used guideline to report SLR and more specific guidelines (e.g., (BUDGEN *et al.*, 2018a; KITCHENHAM *et al.*, 2022)) are still being developed needing more time for researchers adopt them. Following a more complete standard for documentation would enable researchers to not omit important SLR details, consequently, it would improve studies' auditability and facilitates researchers' decision about the need of conducting new research.

5.4.2.3 Technical Perspective

To preserve SLR from technical perspective authors should prioritize the usage of tools that facilitate conduction/update. Our results point out that many technical aspects of SLR were omitted in reporting, but evidence collected indicates that the most popular tools were data analysis/visualization software and reference managers, additionally, no studies reported the usage of tools specifically for managing the SLR process. This scenario indicates that the technical perspective should identify the main causes for this lack of adoption, propose solutions fostering the creation of better tools that attend to the needs of researchers, and create better ways to maintain these tools.

Tools mentioned are partly created by private companies (e.g., MAXQDA, Mendeley) and partly are open-source projects (e.g., Python, R). Using proprietary software on SLR can impose a technical barrier to audition/replication due to the need of acquiring licenses to reuse data. A possible alternative to mitigate this problem is to develop better ways to integrate and exchange data more easily, for example, using of neutral data formats (e.g., JSON, XML) and developing standards for new tools.

Despite the usage of open-source software seeming to eliminate the economic barriers of tool acquisition, it imposes technical challenges that are well-known in open-source projects, for instance, the need for an engaged community of developers that are constantly working on updates. We argue that building a strong community to manage the development of SLR tools is essential for SE. Good examples of non-profit organizations (e.g., Cochrane[11]) are already well-established in the medicine area and are an interesting solution to provide a more sustainable life for SLR tools.

Another technical aspect neglected in the studies are Open Science Policies. Currently, many Journals and Conferences in SE encourage in their guidelines the use of open science practices (e.g., IST [12],JSS[13], ICSE[14], ESEM[15]), nevertheless, only one study provided an external link with their dataset, and nine did not provide any supplementary material or links for other artifacts. In this sense, the technical perspective should also support researchers to store their data in an accessible and reliable way. Furthermore, tools developed for SLR should be able to import/export data to comply with open science principles and facilitate researchers' work in the update.

Regarding tools for automation of SLR tasks and their potential benefits for supporting SLR, only one study used scripts to support study collection. However, there are many other stages of SLR that are amenable to automation (FELIZARDO; CARVER, 2020) that researchers preferred to perform manually. Technical dimension should also foster the adoption of automation techniques, however, we admit that the current scenario of SLR automation is still immature and needs more empirical evidence to be widely adopted.

5.5 Discussion

This section presents our insights highlighting possible flaws in sustainability and provides more practical recommendations for researchers. For this, we summarized our suggestions into 13 leveraging points that could be applied in further studies to increase sustainability of SLR:

1. **LP**$_1$: Improve team knowledge about best practices of SLR process and prioritize the adoption of standards defined and validated by the community (SI1);

2. **LP**$_2$: Design SLR to evolve iteratively making use of pilot testing (SI2);

[11] <https://www.cochrane.org/>
[12] <https://www.elsevier.com/journals/information-and-software-technology>
[13] <https://www.elsevier.com/journals/journal-of-systems-and-software>
[14] <https://conf.researchr.org/track/icse-2023/icse-2023-open-science-policies>
[15] <https://conf.researchr.org/info/esem-2021/open-science>

3. **LP**$_3$: Provide a detailed SLR report making use of best documentation standards to ensure its quality (SI3);
4. **LP**$_4$: Identify, prioritize, and apply the most reasonable number of techniques to mitigate threats to validity (SI4);
5. **LP**$_5$: Prioritize the preservation of resources and avoid propagating problems that hamper future researchers' work (SI5);
6. **LP**$_6$: Define and document the roles played by each stakeholder making use of strategies to improve internal/external communication (SI6) and foster their participation/collaboration (SI7);
7. **LP**$_7$: Prefer hybrid teams combining experience in the SLR process (SI9) and knowledge about the research domain (SI8);
8. **LP**$_8$: Make sure that knowledge acquired/generated along the process is shared/transferred to interested parts (SI10), and all artifacts are fully accessible for readers fostering the open-science (SI11);
9. **LP**$_9$: Make effort to avoid duplicating unreasonably the research and reduce research waste (SI12), ultimately, reuse as many as possible components from the previous SLR (SI13, SI17);
10. **LP**$_{11}$: Use tools to support SLR process to its entire extent (SI14, SI15) and document experience to foster the development/improvement of these tools;
11. **LP**$_{12}$: Design SLR to enable the continuous update (SI16);
12. **LP**$_{13}$: Improve the research usefulness by designing studies that cover multiple perspectives that could benefit the SE community (SI18) and have long-tail impacts (SI19).

Analyzing the leveraging points mentioned above, they reveal that the community seems to have difficulties taking a step further to keep its process in consonance with novelties in SLR process. Hence, simple and well-known solutions are still not applied in practice creating an ineffective environment for scientific research review. In this sense, it is necessary to foster the adoption of updated guidelines, for instance, the standards for documentation like SEGRESS (KITCHENHAM *et al.*, 2022).

Improvement of the reliability of current research has a positive interaction with sustainability, hence, application of techniques to mitigate human bias in research is essential. At the same time, researchers should consider that preserving resources is also a priority to build more sustainable studies. Therefore, it is necessary to balance the application of techniques to mitigate threats to the validity within a responsible use of resources avoiding wasting them unnecessarily.

Roles played by each stakeholder should be defined making sure that a more intensive collaboration occurs. In this sense, hybrid teams i.e., including academia and industry members with different levels of knowledge, seem to be a feasible option that can produce high-quality results and from a social perspective, creates a more inclusive environment and provides opportunities for novice researchers to develop their skills.

Knowledge management is a weakness in current studies and researchers should explore more efficient ways to transfer knowledge for interested parties should be used. In this sense, open science techniques are recommended to promote transparency of studies, which could avoid that research be duplicated due to a lack of information about the conduction process.

Most of the studies preferred to use generic tools since there is a lack of specific and reliable tools for SLR. This is evidence of a systemic problem with current SLR tools which makes their development mostly experimental (prototypes) and unreliable. A possible solution is to strengthen the researcher's collaboration and solidify the guidelines to develop new tools or refactor those that already exist. Similarly, the update process of SLR studies still needs to be improved to reduce the workload required and make this process continuous. Recently a new approach called "Continuous Systematic Literature Reviews" was proposed inspired by DevOps and Open Science practices creating a new approach aligned with sustainability goals (NAPOLEÃO *et al.*, 2022).

Making SLR more sustainable also includes to consider multiple perspectives to plan, conduct, and interpret their results. SLR can have a long-term impact inspiring many other studies, unraveling research gaps, and fostering empirically based discussions. Hence, researchers should focus on improving the usefulness of results to create studies that generate impacts over time.

5.5.1 Threats to Validity

This section presents the potential threats that can have affected the validity of our study divided in four categories proposed by Wohlin *et al.* (2012).

Internal validity: A possible threat is the time span chosen of a decade (2012-2022) because it is possible to lose some information published in previous studies. This risk is acceptable since there is a high probability of problems reported before 2012 having been solved or being reported in the last decade of studies, consequently, being captured by our analysis. Another possible threat is the researcher's subjectivity while answering the questions defined in data extraction, accordingly, we only considered as valid answers that are verifiable in SLR reporting, i.e., are explicitly claimed by authors.

Construct validity: A possible threat to construct is the usage of Ampatzoglou *et al.* (2020) checklist to appraise reliability since it's not mandatory that SLR implement all mitigation actions for a study being considered reliable. To mitigate this threat we performed a more generalistic analysis that does not assume any arbitrary number to appraise reliability. Another potential threat while analyzing only IST and JSS studies is the lack of representativeness of these studies over the quality of SLR being produced currently. The trade-off between coverage and deepness was analyzed and we considered it more important to appraise sustainability in high-quality studies instead of drawing conclusions based on a sample that does not implement best practices.

External validity: a possible threat is the low number of studies selected do not represent the current state of the practice, consequently, limiting the generalizability of our results. To mitigate this we compared our results to other studies that diagnose the current scenario (e.g., (KITCHENHAM; BRERETON, 2013; BUDGEN *et al.*, 2018a; BUDGEN; BRERETON, 2022)). Despite we recognize that results may contain imprecision, we are confident that the evidence gathered is important to reevaluate those studies from a different perspective.

Conclusion validity: A possible threat is the lack of information in reporting about some aspects of SLR. To mitigate this threat we considered not only the paper itself but all information available about it e.g., metadata, supplementary material, external links, and Scopus/Google Scholar reports. Another threat to our conclusions is the lack of information about team composition since researchers' roles can change over time making it difficult to appraise, for instance, their connection with the industry. To mitigate this we cross-checked information available from different sources (DBLP, Google Scholar, and personal pages) and adopted for our analysis those which were more complete.

5.6 Final Remarks and Future Work

This study presents the main leveraging points found in current SLR that should be treated by researchers to transpose them into a more sustainable one. For this, we systematically searched in literature and selected a sample of ten high-quality SLR published in the last decade. Afterward, we extracted data based on 19 sustainability indicators, synthesized them, and reported the results.

Our main contribution is to provide an overview of the current state of the practice of SLR in SE. In addition, we analyzed these studies from a broader perspective and formalized our recommendations in 13 practical actions that may guide current and future researchers to conduct SLR. After conducting this study, we believe that a long journey to achieve sustainability in SLR is still afterward. There is a need for more evidence about the extension of the impacts caused by neglecting the points mentioned in this paper. Furthermore, further investigation is still necessary to provide more pragmatic ways on how researchers could create/adapt current guidelines and propose new approaches to support researchers to comprise with sustainability goals. Our research agenda aims to empirically investigate how to include sustainability as a priority in SE research.

CHAPTER 6

Conclusions

This chapter concludes this thesis by providing answers to the research questions previously defined and discussing the directions for future research about sustainability of SLR.

Chapter 1 states our main research problem: *Is there a way to tackle the fragmentation of the solutions proposed for SLR and transpose it into more integrated thinking?* To address this problem, Chapter 1 defined four specific questions to guide us through our research that was answered in Chapters 2 to 5. A reminder of these questions is presented in Table 20.

Table 20 – Specific research questions defined for this project

RQ_ID	Question	Answered in
RQ1	Which is the state of the art of sustainability in SE area?	Chapter 2
RQ2	How could sustainability be applied to the SLR context?	Chapter 3
RQ3	Which are the characteristics of sustainable SLR?	Chapter 4
RQ4	How to leverage sustainability of SLR?	Chapter 5

6.1 Main Contributions

In our studies, we addressed the aforementioned questions, gaining valuable insights, making contributions, and drawing important lessons. Following, we summarize our main contributions and answer the research questions based on our findings about sustainability of SLR.

RQ_1 - What is the state of the art of sustainability in the SE area?

The field of SE has been greatly influenced by the concept of sustainability. To fully embrace sustainability, SE has matured and developed coordinated efforts to address the impact of software systems on society, individuals, environment, economy, and technological development. Recent literature has highlighted the importance of sustainable SE in modern society, meaning that software engineers must propose solutions to reduce the resources consumed by software systems throughout their life cycle, while understanding the positive and negative impacts of software on other fields.

Valuable contributions of sustainability to software have inspired researchers to observe problems from a wider perspective. Adapting the concept of sustainability for software has paved the way for other areas to embrace the holistic mindset to solve their problems. The main findings about sustainability in SE are presented as follows:

- *Growth in research interest*: The past decade witnessed a surge in studies addressing sustainability in SE, with a notable concentration of secondary studies between 2017 and 2018. This trend underscores the community's growing interest in comprehending the application of sustainability principles to software.

- *Multiple understandings over sustainable software definition*: While there are varying interpretations of sustainable software, a majority of studies converge on the idea that sustainability encompasses multiple dimensions, including social, economic, environmental, technical, and individual aspects. Additionally, sustainable software is those that manages their impacts not only the software production process itself (sustainability within software and first-order effects), but also consequential effects resulting from its application (sustainability through software and second- and third-order effects). It is crucial to consider these side effects for a comprehensive understanding of sustainability in software.

- *Alignment with agile and quality practices*: Numerous agile software development practices, such as continuous improvement, pursuit of quality attributes (e.g., maintainability, modifiability, reusability), and resource efficiency, are inherently aligned with sustainable software development principles.

- *Holistic resource consumption*: Building software involves an extensive chain of resource consumption, encompassing economic, human, and environmental resources. The quest for sustainability must encompass all dimensions to effectively address the impacts of software development.

- *Critical considerations for industry*: Software development encompasses critical points that warrant attention from the industry, including human factors, resource utilization, software features, and the development process. Addressing these aspects is pivotal in promoting sustainable software practices.

- *Community engagement and contributions*: In recent years, the SE community has demonstrated a strong commitment to fostering sustainable software practices. Efforts have been made to share guidelines, disseminate knowledge, and propose models and frameworks that provide guidance for developing sustainable software.

In summary, there is a growing interest in better comprehending what means sustainability in SE. Despite there are still diverse understandings, researchers converge that sustainability causes multi-level effects of software in our society. SE community has advanced much toward discussing the systemic effects of software development on society in recent years, several approaches were proposed to address sustainability goals in software development.

RQ_2 - How could sustainability be applied to the SLR context?

Having delved into the state of the art of sustainability and its implications for SE, our focus on this question is to understand the underlying issues associated with the SLR process highlighting how sustainability can be interpreted in this specific context. Our research has contributed to better comprehending the following aspects:

- *Categorization of SLR problems*: we identified that the key challenges in current SLRs can be better understood when classified into social problems (pertaining to human aspects), economic problems (related to resource consumption), and technical problems (concerning tools and technologies).
- *Interconnection of barriers*: it became evident that these identified barriers influence one another, illustrating their interconnected nature. Each barrier affects others in some capacity, emphasizing the need to address them holistically.
- *Resource preservation and impactful results*: sustainability entails not only resource efficiency but also the production of valuable outcomes, hence, to achieve sustainability, researchers must strive to preserve the resources invested in constructing SLR and produce results that have a meaningful impact on the software industry.
- *Balancing dimensions and visions*: applying sustainability within the SLR context necessitates finding viable approaches to balance or integrate the concerns of each dimension and assessing the impact of adopting different perspectives.

We identified that sustainability can be effectively applied within the SLR context as a means to highlight the systemic effects of current SLR problems. This innovates proposing a unique unifying thread, guiding discussions around existing challenges and shifting researchers' mindsets towards proposing solutions that consider the entire SLR landscape rather than focusing solely on specific aspects. By adopting a sustainability lens, we aim to foster a broader perspective that encourages holistic problem-solving and promotes sustainable practices within the SLR.

RQ_3 - Which are the characteristics of sustainable SLR?

After taking the first steps towards sustainability, our goal in this question is to dig deeper into sustainability principles defined in SE area to adapt them into the context of SLR. Our research identified 15 main characteristics, 15 critical factors, and 16 guidelines that define what means sustainable SLR and contribute to advancing current research towards sustainability.

Following, we present the key findings that helped us to define the main characteristics of sustainable SLR:

- *Alignment between sustainable software development characteristics and sustainable SLR*: Our results presented a reasonable parallel between several sustainable software development topics and SLR. Hence, we outline using meta-ethnography that some of the sustainable software characteristics, such as a well-defined software development process, usage of knowledge management techniques, usage of support tools, support for quality attributes (e.g., reliability, modifiability, accessibility), application of good practices, or even a conscious resource consumption (human, monetary, time, and effort) can be considered as sustainable SLR characteristics.

- *Successful Application of Meta-Ethnographic Method*: A noteworthy advancement for EBSE proposed in this work is the utilization of the meta-ethnographic method. Our experience not only underscores the feasibility of using meta-ethnography to transpose successful concepts from diverse domains into the realm of SLR, but also sets a precedent for future endeavors. By using this method, we open precedents for future studies that can serve as inspiration to SE to assimilate well-established solutions which were validated and endorsed by experts of various disciplines.

- *Reinforcement of well-established good practices*: Our results showed that sustainability characteristics often align with well-established best practices in SLR. This reinforces the importance of preserving the great achievements in evolving SLR method and incentives the community to apply those practices to their research. At the same time, it draws attention to the need of revisiting certain topics (e.g., the need for heavy documentation, or the intensive usage of double-checking mechanisms) to ensure that the process and its final product are guided by sustainability objectives.

- *Introduction of new terminology to designate sustainability characteristics*: Our research innovates by introducing several terms that refer to topics that had limited attention in the SLR context. These terms include the importance of observing the long-term goals of SLR, practicing responsible resource usage, understanding the concept of components within SLR, or even the existence of research waste related to SLR conduction. By defining these terms, we bring into light structural problems of the SLR process that require more attention, furthermore, the new terminology also enables us to discuss and propose different solutions that are aligned with compose sustainable SLR characteristics.

- *Addressing Critical Points for Sustainable SLR*: Our work highlight which are the critical points to conducting sustainable SLR emphasizing that sustainability needs a strong social connection between stakeholders, highly efficient knowledge management, effective utilization of technological support, and the application of resource optimization techniques in the SLR process. Similarly, the collected guidelines reaffirm as a characteristic of sustainable SLR the commitment to adopting practical strategies that mitigate the challenges associated with these critical points.

- *Inclusive Approach and Broader Context*: Another characteristic of sustainable SLR is the proposal of a more inclusive SLR process, i.e., a process that considers a broader context that incorporates the needs and constraints of academia and the software industry. This balanced approach accounts for social, economic, and technical perspectives, guiding the community to create reviews that are conscious of their impact on society.

While sustainability serves as a bootstrap to evolve SLR into the next level, it is crucial to continuously refine the identified characteristics and adapt to emerging needs. Henceforth, we believe that the pursuit of new solutions to enhance quality and potentially automate processes are aligned with the scope of sustainability. Therefore, future works should focus on investigating innovative solutions that evolve the SLR process while upholding its sustainability principles.

RQ_4 - How to leverage the sustainability of SLR?

After defining the characteristics of sustainable systematic reviews, we evaluated the ten most cited SLR published in two reputable journals in the SE field over the past decade through the lens of sustainability. Our analysis revealed opportunities for researchers to enhance the sustainability of their SLR that we present as leveraging points as follows:

- *Strengthen the use of good SLR practices*: researchers should apply the best practices suggested in classic literature about SLR method and stay updated to put into practice techniques that could improve the sustainability of their reviews. A leveraging point that could be adopted immediately is to foster the conduction of pilot tests to refine the protocol and iterative development that are aligned with sustainability characteristics.

- *Provide quality documentation*: SLR documentation should adopt the best documentation standards to prevent errors and enable easy updates by future researchers. A leveraging point for current SLR is to explore better ways of documenting details that are important like the procedures taken for team collaboration or even standards adopted for data extraction and quality assessment.

- *Application of techniques to improve SLR quality*: techniques to mitigate validity threats are applied in an ad-hoc manner, consequently, it becomes difficult to measure the real impacts and benefits on quality and resource consumption of applying these techniques. Therefore, a leveraging point of sustainability is to use more systematic methods to identify, prioritize, and apply techniques to mitigate threats to validity to prevent SLR flaws always prioritizing high-quality results.

- *Research team composition and industry support*: SLR often lack support from industry members in planning, conducting, or validating study results. Insufficient commitment to cover multiple perspectives (academic and industry) weakens social connections and limits the impact of SLR. We state a leveraging point to strengthen the communication among stakeholders and compose hybrid teams capable of addressing problems from different perspectives and providing richer results.

- *Enhancements in collaboration and communication*: Collaboration in SLR should extend beyond consensus in manual tasks like data extraction or study selection. We state as a leveraging point the effective communication of the whole team during the process especially focusing on sharing experiences while planning, conducting, or documenting SLR.

- *Accessibility and knowledge sharing*: providing open access to artifacts generated during the review process is essential to ensure that SLR is accessible to anyone. This lack of knowledge sharing impedes reusability, auditability, and updates, thus limiting long-term impact. We state a leveraging point to enhance knowledge management in SLR studies to avoid propagating problems and facilitate future research work, for this, open science practices in SLR should be fostered enabling free access to data.

- *Efficiency enhancement*: SLR studies should prioritize reducing the effort required by adopting appropriate tools, automating tasks, reusing components from previous SLR, and avoiding unnecessary duplication of research work. We state as a leveraging point of sustainability consistently using techniques to enhance the efficiency of the SLR process and reduce the amount of resources needed in the extent of its life-cycle.

Even in high-quality SLR that were reviewed rigorously several leveraging the points were observed using the sustainability perspective, since it shed light on important issues that are overlooked currently. For instance, the importance of enhancing the quality and efficiency of current SLR or the internal communication among researchers and industry stakeholders. Addressing these leveraging points could foster the sustainability of SLR in SE, moreover, we pave the way for future investigations aimed at proposing solutions that align with this vision.

6.2 Limitations and Future Directions

This study represents an initial step toward understanding sustainability in a specific context. However, limitations were encountered during the investigation. One limitation pertains to the applicability of the sustainability characteristics identified in this study to other types of secondary studies, such as systematic mapping or gray literature reviews. The diversity of secondary study types makes it challenging to generalize the findings beyond the scope of this research.

Furthermore, the method used to adapt the concept of sustainability (meta-ethnography) relies on the current state of the art in both relevant areas. As a result, certain well-developed concepts in one area may not have been fully explored or discussed in the other. This limitation emphasizes that our conclusions are based on consolidated literature and identified trends, with potential directions for future research in SLR.

In terms of the research approach, the validation of characteristics and critical factors, as well as the leverage points for sustainability in SLR, involved a limited number of researchers and a sampling process. These decisions were intentional, prioritizing qualitative analysis to promote insightful discussions among researchers and obtain diverse perspectives on the topic. While this approach allowed us an in-depth exploration, it also constitutes a limitation, as a larger participant pool and quantitative methods would have provided a more comprehensive view of the community's acceptance of sustainability. Additionally, the focus on studies that underwent rigorous peer review and garnered significant citations may have resulted in overlooking valuable insights from studies published in less prominent outlets. A broader survey incorporating studies from various media, such as journals and conferences, could have shed light on leverage points within different communities.

The journey toward sustainability in SLR is ongoing, and more evidence is needed to understand its impact on SE research. Despite the limitations, this work revealed several open issues, such as how to improve the SLR process by incorporating cost-benefit trade-offs and ensuring that SLR continues to offer benefits for academia and industry. We argue that future research should investigate pragmatic ways to include sustainability as a priority in SE research, in combination with massive dissemination of the importance of sustainability for SLR and how it could be a catalyst for scientific research.

This project introduced several keywords that have been adapted to compose a new way of how to deal with SLR, for example, the concept of research waste, life cycle, or the existence of components for reviews that could be reused. Despite this work contributing to unraveling the meaning of these terms, there are still gaps that need to be filled and issues that need to be answered, for instance, which are the negative impacts of outdated or wrong conclusions in the community, how much research is wasted due to misleading results, which the main components of SLR are, and how suitable they are for reuse. Therefore, we believe that an important contribution is to dig deeper into these concepts and further build a new taxonomy fully adapted to support discussions about sustainability of SLR.

An example of a promising concept is the utilization of "quality attributes" (which are widely known in the domain of SE) to leverage the sustainability of SLR studies. Taking advantage of this insight, we proposed a lightweight maturity model (named MM4SLR) to diagnose and improve the maturity of SLR process in a systematic way (discussed in depth in the work presented in Appendix A.4). MM4SLR is designed as a proof-of-concept project that shows how researchers could use sustainability to guide their future research. MM4SLR is fully aligned with sustainability principles since it allows researchers with different levels of knowledge about SLR evaluate and progressively evolve their process making it more sustainable. Nevertheless, we recognize that further research is still required to ensure that quality attributes remain valid in different research contexts or define efficient measurement methods for these attributes.

Sustainability has already sparked the interest of other researchers and led to the proposal of innovative solutions aimed at optimizing the review process and addressing the social, economic, and technical barriers associated with it. Napoleão *et al.* (2022) proposed an innovative type of SLR called Continuous Systematic Literature Review (CSLR), it offers a practical and systematic approach to continuously integrate new evidence into SLR in SE. This method directly impacts the economic aspect by proposing a method that reduces the current friction in updating studies mainly associated with accessing and processing raw data, consequently, decreasing the workload at the end of the process. Moreover, CSLR also addresses the technical dimension by fostering the development of tools for automating, managing, and sharing information effectively. CSLR also impacts the social dimension since it delivers updated results, enables their application in industry decision-making processes, and facilitates communication among researchers through a notification mechanism that alerts them when a study becomes outdated.

Another insight that emerged from our research was the existence of trade-offs exists involving in the construction of reviews that warrant further exploration. Certain parameters of systematic reviews, such as the composition of the research team, level of documentation detail, adopted search strategy, and the implementation of double-checking mechanisms to mitigate bias, can be tailored according to the specific objectives of the review. For instance, while thorough documentation is crucial for ensuring the replicability and reliability of the study, it demands significant effort and time from researchers. A future research should investigate the advantages and disadvantages (trade-offs) associated with investing time and effort into extensive documentation, as well as the tangible benefits it brings to research.

Ultimately, we contend that the development of strategies for organizing knowledge on sustainability and transposing it into practice is very important. As a future work, a framework could be built comprising models, best practices, guidelines, and metrics to form a robust collection of resources that can guide researchers throughout the entire life cycle, i.e., from cradle to grave. We believe that future works should focus on changing the community mindset over literature reviews and comprehend the importance of using the sustainability view to support researchers. Henceforth, it will be possible to overcome current SLR barriers and create ways to achieve sustainability using a pragmatic and progressive.

6.3 Final Remarks

The initial purpose of the SLR process was to provide a reliable means of extracting information from the literature to aid SE researchers and practitioners in decision-making. However, the process's inherent challenges, such as resource-intensive demands and a lack of connection between the problems investigated and the needs of the software industry, have posed a barrier for SLR to achieve its objective. In this thesis, a new approach is proposed to address the current challenges of SLR process. Our results revealed that these issues are somehow interconnected and the only way to address them is to propose an integrated solution.

Sustainability is introduced as a key concept to address the current challenges with the potential to unlock connections among the different dimensions of SLR process guiding researchers toward more sustainable literature reviews. This project advances the state-of-art of SE proposing that being sustainable in SLR context means observing problems from three different perspectives, i.e., social (researchers and industry stakeholders), economic (resources consumption), and technical (tools and support technologies) aiming to find solutions that cover as most as possible the needs from these three perspectives. Beyond only presenting key characteristics and critical factors of a sustainable review, we advocate that researchers should look forward to comprehending the impacts of creating SLR that complies with sustainability principles. We also contributed by providing an empirical evaluation of the SLR practices currently used in SE. Our results demonstrated that even good reviews still need to fill some important gaps that can only be observed through a sustainability lens, hence, we propose an initial set of guidelines and leverage points. Overall, this project provides a new perspective on the SLR process and its challenges, hence, we expect that it become a trigger to the construction of more sustainable reviews that are capable of addressing the current issues and offer a new pathway to leverage EBSE.

This thesis contributes toward developing a novel perspective on the SLR process. A major challenge encountered in defining sustainability was to adapt the specificities that were previously only understood within the ecological or software development contexts and make them more accessible to the target audience of this work. This project seeks to introduce the concept of sustainability into the field of SLR and therefore has an exploratory nature, at times even didactic, in order to bring previously abstract concepts closer to the researcher's reality. We believe that this work provides valuable insights into the importance of sustainability in the SLR process, and it could serve as a foundation for future research in this area.

References

ACHIMUGU, P.; SELAMAT, A.; IBRAHIM, R.; MAHRIN, M. N. A systematic literature review of software requirements prioritization research. **Information and Software Technology**, Elsevier BV, v. 56, n. 6, p. 568–585, Jun. 2014. Citation on page 95.

AHMAD, R.; BAHAROM, F.; HUSSAIN, A. A systematic literature review on sustainability studies in software engineering. In: **Knowledge Management International Conference (KMICe)**. Langkawi, Malaysia: , 2014. p. 1–6. Citations on pages 13, 18, 51, 59, 61, and 159.

AL-ZUBIDY, A.; CARVER, J. Identification and prioritization of SLR search tool requirements: an SLR and a survey. **Empirical Software Engineering**, Springer Science and Business Media LLC, v. 24, n. 1, p. 139–169, May 2018. Citations on pages 67, 74, 86, 90, 104, 176, 177, 182, and 191.

ALI, A.; GRAVINO, C. A systematic literature review of software effort prediction using machine learning methods. **Journal of Software: Evolution and Process**, Wiley, v. 31, n. 10, Jul. 2019. Citations on pages 97, 104, and 105.

ALLEN, C.; SMITH, M.; RABIEE, M.; DAHMM, H. A review of scientific advancements in datasets derived from big data for monitoring the sustainable development goals. **Sustainability Science**, Springer Science and Business Media LLC, v. 16, n. 5, p. 1701–1716, Jun. 2021. Citation on page 31.

AMPATZOGLOU, A.; BIBI, S.; AVGERIOU, P.; VERBEEK, M.; CHATZIGEORGIOU, A. Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. **Information and Software Technology**, v. 106, p. 201–230, 2019. Citations on pages 36, 37, 38, 39, 40, 41, 42, 50, 51, 78, 86, 90, 96, 175, 186, and 188.

AMPATZOGLOU, A.; BIBI, S.; AVGERIOU, P.; CHATZIGEORGIOU, A. Guidelines for managing threats to validity of secondary studies in software engineering. In: **Contemporary Empirical Methods in Software Engineering**. 2020. p. 415–441. Citations on pages 60, 61, 62, 68, 75, 112, 149, 150, 151, 154, 156, 157, 164, 165, 166, 167, and 168.

AMSEL, N.; IBRAHIM, Z.; MALIK, A.; TOMLINSON, B. Toward sustainable software engineering. In: **33rd International Conference on Software Engineering (ICSE) - NIER track**. 2011. p. 976–979. Citation on page 68.

AMSEL, N.; TOMLINSON, B. Green tracker. In: **Conference on Human Factors in Computing Systems (CHI)**. 2010. p. 3337–3342. Citation on page 68.

ANWAR, H.; PFAHL, D. Towards greener software engineering using software analytics: A systematic mapping. In: **43rd Euromicro Conference on Software Engineering and Advanced Applications**. Vienna, Austria: , 2017. p. 157–166. Citations on pages 13, 18, 24, 51, 59, 61, and 159.

ANWER, F.; AFTAB, S.; SHAH, S.; WAHEED, U. Comparative analysis of two popular agile process models: Extreme programming and scrum. **International Journal of Computer Science and Telecommunications**, v. 8, n. 2, p. 1–7, 2017. Citation on page 66.

ATKINS, S.; LEWIN, S.; SMITH, H.; ENGEL, M.; FRETHEIM, A.; VOLMINK, J. Conducting a meta-ethnography of qualitative literature: Lessons learnt. **BMC Medical Research Methodology**, v. 8, n. 1, p. 1–10, 2008. Citations on pages 63 and 79.

AZEEM, M. I.; PALOMBA, F.; SHI, L.; WANG, Q. Machine learning techniques for code smell detection: A systematic literature review and meta-analysis. **Information and Software Technology**, Elsevier BV, v. 108, p. 115–138, Apr. 2019. Citation on page 95.

BABAR, M.; ZHANG, H. Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. In: IEEE COMPUTER SOCIETY. **3rd International Symposium on Empirical Software Engineering and Measurement (ESEM'09)**. 2009. p. 346–355. Citations on pages 50 and 86.

BADAMPUDI, D.; WOHLIN, C.; GORSCHEK, T. Contextualizing research evidence through knowledge translation in software engineering. In: **Evaluation and Assessment in Software Engineering (EASE)**. 2019. p. 306–311. Citations on pages 2, 36, 42, 51, 60, 61, 62, 67, 164, 165, and 191.

BECKER, C.; CHITCHYAN, R.; DUBOC, L.; EASTERBROOK, S.; MAHAUX, M.; PENZENSTADLER, B.; RODRIGUEZ-NAVAS, G.; SALINESI, C.; SEYFF, N.; VENTERS, C. *et al.* The Karlskrona manifesto for sustainability design. In: **37th International Conference on Software Engineering (ICSE)**. 2015. p. 467–476. Citations on pages 53, 88, and 89.

BECKER, C.; CHITCHYAN, R.; DUBOC, L.; EASTERBROOK, S.; PENZENSTADLER, B.; SEYFF, N.; VENTERS, C. C. Sustainability design and software: The karlskrona manifesto. In: **37th International Conference on Software Engineering (ICSE)**. Florence, Italy: , 2015. v. 2, p. 467–476. Citations on pages 12, 13, and 24.

BECKER, C.; CHITCHYAN, R.; DUBOC, L.; EASTERBROOK, S.; PENZENSTADLER, B.; SEYFF, N.; VENTERS, C. C. Sustainability design and software: The karlskrona manifesto. In: **37th International Conference on Software Engineering (ICSE)**. Florence, Italy: , 2015. v. 2, p. 467–476. Citations on pages 38, 39, 43, 44, and 45.

BELKHIR, L.; ELMELIGI, A. Assessing ict global emissions footprint: Trends to 2040 & recommendations. **Journal of Cleaner Production**, v. 177, p. 448–463, 2018. ISSN 0959-6526. Citation on page 12.

BERKHOUT, F.; HERTIN, J. **Impacts of information and communication technologies on environmental sustainability: Speculations and evidence**. 2001. v. 21. Citation on page 22.

BERNTSEN, K. R.; OLSEN, M. R.; LIMBU, N.; TRAN, A. T.; COLOMO-PALACIOS, R. Sustainability in software engineering - a systematic mapping. In: MEJIA, J.; MUÑOZ, M.; ROCHA, Á.; FELIU, T. S.; PEÑA, A. (Ed.). **6th International Conference on Software Process Improvement**. Zacatecas, Mexico: , 2017. p. 23–32. Citations on pages 18, 59, 61, 159, and 160.

BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. **Systematic review in software engineering**. 2005. v. 679, n. 05, 45 p. Citation on page 188.

BOOCH, G. Morality and the software architect. **IEEE Software**, Institute of Electrical and Electronics Engineers (IEEE), v. 25, n. 1, p. 8–9, Jan. 2008. Citation on page 3.

BORREGO, M.; FOSTER, M.; FROYD, J. What is the state of the art of systematic review in engineering education? **Journal of Engineering Education**, Wiley Online Library, v. 104, n. 2, p. 212–242, 2015. Citations on pages 50 and 86.

BOWES, D.; HALL, T.; BEECHAM, S. Slurp: A tool to help large complex systematic literature reviews deliver valid and rigorous results. In: **2nd International Workshop on Evidential Assessment of Software Technologies**. Lund, Sweden: , 2012. p. 33–36. Citations on pages 176 and 191.

BOZZELLI, P.; GU, Q.; LAGO, P. **A systematic literature review on green software metrics**. Amsterdam, Netherlands, 2013. Citations on pages 18 and 68.

BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of Systems and Software**, v. 80, n. 4, p. 571–583, 2007. Citations on pages 60, 61, 62, 67, 98, 164, 165, 167, and 168.

BRHEL, M.; METH, H.; MAEDCHE, A.; WERDER, K. Exploring principles of user-centered agile software development: A literature review. **Information and Software Technology**, Elsevier BV, v. 61, p. 163–181, May 2015. Citation on page 95.

BRIDGES, J.; NICHOLSON, C.; MABEN, J.; POPE, C.; FLATLEY, M.; WILKINSON, C.; MEYER, J.; TZIGGILI, M. Capacity for care: meta-ethnography of acute care nurses experiences of the nurse-patient relationship. **Journal of Advanced Nursing**, v. 69, n. 4, p. 760–772. Citation on page 55.

BROCKE, J. v.; SIMONS, A.; NIEHAVES, B.; NIEHAVES, B.; REIMER, K.; PLATTFAUT, R.; CLEVEN, A. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: **17th European Conference on Information Systems (ECIS 2009)**. Verona, Italy: , 2009. p. 2206–2217. Citations on pages 98 and 99.

BROWN, B. J.; HANSON, M. E.; LIVERMAN, D. M.; MERIDETH, R. W. Global sustainability: Toward definition. **Environmental Management**, v. 11, n. 6, p. 713–719, 1987. Citations on pages 39 and 52.

BROWN, L. R. Building a sustainable society. **Society**, v. 19, n. 2, p. 75–85, 1982. Citation on page 88.

BUDGEN, D.; BRERETON, P. Short communication: Evolution of secondary studies in software engineering. **Information and Software Technology**, Elsevier BV, v. 145, p. 106840, May 2022. Citations on pages 2, 96, 102, 108, and 113.

BUDGEN, D.; BRERETON, P.; DRUMMOND, S.; WILLIAMS, N. Reporting systematic reviews: Some lessons from a tertiary study. **Information and Software Technology**, v. 95, p. 62–74, 2018. Citations on pages 2, 36, 37, 38, 39, 40, 41, 42, 50, 60, 61, 62, 67, 75, 86, 90, 99, 100, 109, 113, 165, 166, 167, 169, 186, and 188.

BUDGEN, D.; BRERETON, P.; DRUMMOND, S.; WILLIAMS, N. Reporting systematic reviews: Some lessons from a tertiary study. **Information and Software Technology**, v. 95, p. 62 – 74, 2018. ISSN 0950-5849. Citations on pages 149, 152, 153, 154, and 155.

CABEZAS, H.; DIWEKAR, U. **Sustainability: Multi-disciplinary perspectives**. : Bentham Science Publishers, 2012. Citation on page 3.

CAIRO, L.; CARNEIRO, G.; MONTEIRO, M.; ABREU, F. Towards the use of machine learning algorithms to enhance the effectiveness of search strings in secondary studies. In: **SBES**. 2019. p. 22–26. Citations on pages 2, 36, and 51.

CALERO, C.; BERTOA, M. F.; MORAGA, M. A. A systematic literature review for software sustainability measures. In: **2nd International Workshop on Green and Sustainable Software**. San Francisco, CA, USA: , 2013. p. 46–53. Citations on pages 18, 59, 61, 159, 160, and 161.

CALERO, C.; MANCEBO, J.; GARCIA, F.; MORAGA, M. A.; BERNA, J. A. G.; FERNANDEZ-ALEMAN, J. L.; TOVAL, A. 5ws of green and sustainable software. **Tsinghua Science and Technology**, v. 25, n. 3, p. 401–414, 2020. Citation on page 13.

CALERO, C.; MORAGA, M. Á.; BERTOA, M. F.; DUBOC, L. Green software and software quality. In: **Green in Software Engineering**. : Springer International Publishing, 2015. p. 231–260. Citation on page 22.

CALERO, C.; MORAGA, M. Ángeles; GARCíA, F. Software, sustainability, and un sustainable development goals. **IT Professional**, v. 24, n. 1, p. 41–48, 2022. Citation on page 22.

CALERO, C.; PIATTINI, M. Puzzling out software sustainability. **Sustainable Computing: Informatics and Systems**, Elsevier BV, v. 16, p. 117–124, Dec. 2017. Citation on page 12.

CAMPBELL, R.; POUND, P.; MORGAN, M.; DAKER-WHITE, G.; BRITTEN, N.; PILL, R.; YARDLEY, L.; POPE, C.; DONOVAN, J. Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. **Health Technology Assessment**, v. 15, n. 43, p. 1–164, 2011. Citation on page 55.

CARTAXO, B.; PINTO, G.; SOARES, S. The role of rapid reviews in supporting decision-making in software engineering practice. In: **EASE**. 2018. p. 24–34. Citations on pages 2, 3, 36, 42, 51, 54, 60, 61, 62, 90, 92, 164, 167, 186, and 191.

CARTAXO, B.; PINTO, G.; SOARES, S. Rapid reviews in software engineering. In: **Contemporary Empirical Methods in Software Engineering**. : Springer, 2020. p. 357–384. Citation on page 54.

CARTAXO, B.; PINTO, G.; VIEIRA, E.; SOARES, S. Evidence briefings. In: **10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2016. p. 1–10. Citation on page 76.

CARVER, J. C.; HASSLER, E.; HERNANDES, E.; KRAFT, N. A. Identifying barriers to the systematic literature review process. In: **International Symposium on Empirical Software Engineering and Measurement**. 2013. p. 203–212. ISSN 1949-3770. Citations on pages 2, 54, 86, 90, 178, 186, and 191.

CASTRO-CABRERA, M. d. C. de; GARCÍA-DOMINGUEZ, A.; MEDINA-BULO, I. Trends in prioritization of test cases: 2017-2019. In: **35th Annual ACM Symposium on Applied Computing**. New York, NY, USA: Association for Computing Machinery, 2020. (SAC ’20), p. 2005–2011. ISBN 9781450368667. Citations on pages 97, 104, and 105.

CHRISSIS, M.; KONRAD, M.; SHRUM, S. **CMMI for development**. 3. ed. Boston, MA: Addison-Wesley Educational, 2011. Citations on pages 186 and 187.

CORBETT, J. Unearthing the value of green it. In: **31st International Conference on Information Systems (ICIS)**. Saint Louis, USA: , 2010. v. 198, p. 1–22. Citation on page 24.

CORBIN, J.; STRAUSS, A. **Basics of qualitative research: Techniques and procedures for developing grounded theory**. 2. ed. : Sage publications, 1998. Citation on page 65.

CRAWFORD, J. **Project Management Maturity Model**. : Auerbach Publications, 2021. Citation on page 188.

CRUZES, D.; DYBA, T. Recommended steps for thematic synthesis in software engineering. In: **5th International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2011. p. 275–284. Citations on pages 62, 65, 92, and 95.

CRUZES, D. S.; DYBA, T. Recommended steps for thematic synthesis in software engineering. In: **5th International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Banff, Canada: , 2011. p. 1–10. Citations on pages 19, 54, 55, and 188.

CRUZES, D. S.; DYBA, T. Recommended steps for thematic synthesis in software engineering. In: **International Symposium on Empirical Software Engineering and Measurement (ESEM'11)**. Banff, Canada: IEEE, 2011. p. 275–284. Citation on page 188.

CURRY, E.; DONNELLAN, B. Sustainable information systems and green metrics. **Harnessing Green IT: Principles and Practices**, p. 167–198, 2012. Citation on page 68.

DALY, H. **Toward a steady-state economy**. illustrated edition edition. San Francisco, United States: , 1973. ISBN 0716707934. Citation on page 88.

DEBBARMA, T.; CHANDRASEKARAN, K. Green measurement metrics towards a sustainable software: A systematic literature review. In: **International Conference on Recent Advances and Innovations in Engineering**. Jaipur, India: , 2016. Citation on page 18.

DEEK, F.; MCHUGH, J.; ELJABIRI, O. **Strategic Software Engineering**. : Auerbach Publications, 2005. Citations on pages 12 and 88.

DEWAELE, A.; VANDAEL, K.; MEYSMAN, S.; BUYSSE, A. Understanding collaborative interactions in relation to research impact in social sciences and humanities: A meta-ethnography. **Research Evaluation**, v. 30, n. 2, p. 179–190, 2021. Citation on page 55.

DIKERT, K.; PAASIVAARA, M.; LASSENIUS, C. Challenges and success factors for large-scale agile transformations: A systematic literature review. **Journal of Systems and Software**, Elsevier BV, v. 119, p. 87–108, Sep. 2016. Citation on page 95.

DORI, D. Object-process methodology. In: **Encyclopedia of Knowledge Management**. 2011. p. 1208–1220. Citation on page 28.

DORLING, A. SPICE: Software process improvement and capability determination. **Software Quality Journal**, v. 2, n. 4, p. 209–224, 1993. Citation on page 187.

DYBA, T.; KITCHENHAM, B.; JORGENSEN, M. Evidence-based software engineering for practitioners. **IEEE Software**, v. 22, n. 1, p. 58–65, 2005. Citations on pages 17 and 188.

ELLIOTT, J.; SYNNOT, A.; TURNER, T.; SIMMONDS, M.; AKL, E.; MCDONALD, S.; SALANTI, G.; MEERPOHL, J.; MACLEHOSE, H.; HILTON, J.; SHEMILT, I.; THOMAS, J. Living systematic review: 1. introduction—the why, what, when, and how. **Journal of Clinical Epidemiology**, v. 91, p. 23–30, 2017. Citation on page 81.

ELLIOTT, J. H.; TURNER, T.; CLAVISI, O.; THOMAS, J.; HIGGINS, J. P. T.; MAVERGAMES, C.; GRUEN, R. L. Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. **PLoS Medicine**, Public Library of Science (PLoS), v. 11, n. 2, p. e1001603, Feb. 2014. Citation on page 3.

ERDELYI, K. Special factors of development of green software supporting eco sustainability. In: **11th International Symposium on Intelligent Systems and Informatics (SISY)**. 2013. p. 337–340. Citation on page 171.

FABBRI, S.; SILVA, C.; HERNANDES, E.; OCTAVIANO, F.; THOMMAZO, A.; BELGAMO, A. Improvements in the StArt tool to better support the systematic review process. In: **20th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. 2016. p. 1–5. Citation on page 176.

FABBRI, S. C. P. F.; FELIZARDO, K. R.; FERRARI, F. C.; HERNANDES, E. C. M.; OCTAVIANO, F. R.; NAKAGAWA, E. Y.; MALDONADO, J. C. Externalising tacit knowledge of the systematic review process. **IET Software**, v. 7, n. 6, p. 298–307, December 2013. ISSN 1751-8806. Citations on pages 60, 61, 62, 86, 108, and 164.

FATHI, B.; ANSARI, A.; ANSARI, A. Threats of internet-of-thing on environmental sustainability by e-waste. **Sustainability**, v. 14, n. 16, p. 10161, 2022. Citation on page 63.

FELIZARDO, K.; CARVER, J. Automating systematic literature review. In: **Contemporary Empirical Methods in Software Engineering**. : Springer International Publishing, 2020. p. 327–355. Citations on pages 2, 36, 51, 54, 60, 61, 62, 68, 74, 75, 86, 90, 110, 167, 168, 186, and 191.

FELIZARDO, K.; MENDES, E.; KALINOWSKI, M.; SOUZA, É.; VIJAYKUMAR, N. Using forward snowballing to update systematic reviews in software engineering. In: **10th International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2016. p. 1–6. Citations on pages 73 and 74.

FELIZARDO, K.; NAKAGAWA, E.; FABBRI, S.; FERRARI, F. **Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática**. : Elsevier Brazil, 2017. Citation on page 36.

FELIZARDO, K.; SALLEH, N.; MARTINS, R.; MENDES, E.; MACDONELL, S.; MALDONADO, J. Using visual text mining to support the study selection activity in systematic literature reviews. In: **ESEM**. 2011. p. 1–10. Citation on page 77.

FELIZARDO, K.; SILVA, A. da; SOUZA, E. de; VIJAYKUMAR, N.; NAKAGAWA, E. Evaluating strategies for forward snowballing application to support secondary studies updates: emergent results. In: **32th Brazilian Symposium on Software Engineering**. 2018. p. 184–189. Citations on pages 96 and 105.

FELIZARDO, K.; SOUZA, E.; FALBO, R.; VIJAYKUMAR, N.; MENDES, E.; NAKAGAWA, E. Defining protocols of systematic literature reviews in software engineering: A survey. In: **SEAA**. 2017. p. 202–209. Citations on pages 2, 36, 51, and 108.

FELIZARDO, K.; SOUZA, E.; HESAE, S.; VIJAYKUMAR, N.; NAKAGAWA, E. Analysing the use of graphical abstracts to support study selection in secondary studies. In: **CIbSE**. 2017. p. 497–510. Citations on pages 2 and 36.

FELIZARDO, K.; SOUZA, E.; MALACRIDA, T.; NAPOLEÃO, B.; PETRILLO, F.; HALLE, S.; VIJAYKUMAR, N.; NAKAGAWA, E. Knowledge management for promoting update of systematic literature reviews: An experience report. In: **46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**. 2020. p. 471–478. Citations on pages 67 and 181.

FELIZARDO, K.; SOUZA, É.; NAPOLEÃO, B.; VIJAYKUMAR, N.; BALDASSARRE, M. Secondary studies in the academic context: A systematic mapping and survey. **Journal of Systems and Software**, v. 170, p. 110734, 2020. Citations on pages 2, 50, and 86.

FELIZARDO, K.; SOUZA, E. d.; HESAE, S.; VIJAYKUMAR, N.; NAKAGAWA, E. Analyzing the use of graphical abstracts to support study selection in secondary studies. In: **20th Ibero-American Conference on Software Engineering (CibSE)**. 2017. p. 1–14. Citation on page 77.

FELIZARDO, K.; SOUZA, E. de; FALBO, R.; VIJAYKUMAR, N.; MENDES, E.; NAKAGAWA, E. Defining protocols of systematic literature reviews in software engineering: A survey. In: **43rd Euromicro Conference on Software Engineering and Advanced Applications**. 2017. p. 202–209. Citations on pages 50 and 178.

FERNÁNDEZ-SÁEZ, A.; BOCCO, M.; ROMERO, F. Slr-tool - A tool for performing systematic literature reviews. In: **2nd Int. Conf. on Software and Data Technologie (ICSOFT)**. 2010. p. 157–166. Citation on page 191.

FITZGERALD, B.; STOL, K. Continuous software engineering: A roadmap and agenda. **Journal of Systems and Software**, v. 123, p. 176–189, 2017. Citation on page 81.

FOSTER, E.; DEARDORFF, A. Open science framework (OSF). **Journal of the Medical Library Association**, v. 105, n. 2, p. 203–206, 2017. Citation on page 182.

FU, C.; ZHANG, H.; HUANG, X.; ZHOU, X.; LI, Z. A review of meta-ethnographies in software engineering. In: **23rd International Conference on Evaluation and Assessment in Software Engineering (EASE)**. 2019. p. 68–77. Citations on pages 52, 54, 55, 60, and 80.

GARCÉS, L.; FELIZARDO, K.; OLIVEIRA, L.; NAKAGAWA, E. An experience report on update of systematic literature reviews. In: **29th International Conferences on Software Engineering and Knowledge Engineering (SEKE)**. 2017. p. 91–96. Citations on pages 105 and 179.

GARCIA-MIRELES, G. A. Environmental sustainability in software process improvement: a systematic mapping study. In: MEJIA, J.; MUNOZ, M.; ROCHA, A.; FELIU, T. S.; PENA, A. (Ed.). **Trends and Applications in Software Engineering**. Aguascalientes, Mexico: , 2017. p. 69–78. Citations on pages 15, 18, 59, 61, and 159.

GARCIA-MIRELES, G. A.; MORAGA, M. A.; GARCIA, F.; CALERO, C.; PIATTINI, M. Interactions between environmental sustainability goals and software product quality: A mapping study. **Information and Software Technology**, v. 95, p. 108 – 129, 2018. ISSN 0950-5849. Citations on pages 18, 59, 61, 63, 68, 159, 160, 161, 162, and 163.

GARCÍA-MIRELES, G. A.; VILLA-MARTÍNEZ, H. A. Practices for addressing environmental sustainability through requirements processes. In: **Trends and Applications in Software Engineering**. COZCyT, Zacatecas, Mexico: Springer International Publishing, 2018. p. 61–70. ISBN 978-3-319-69341-5. Citations on pages 18, 59, 62, 159, 161, and 162.

GAROUSI, V.; FELDERER, M.; MäNTYLä, M. V. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. **Information and Software Technology**, v. 106, p. 101–121, 2019. Citations on pages 57 and 92.

GAROUSI, V.; FERNANDES, J. M. Highly-cited papers in software engineering: The top-100. **Information and Software Technology**, Elsevier BV, v. 71, p. 108–128, Mar. 2016. Citations on pages 97, 105, and 107.

GIDDINGS, B.; HOPWOOD, B.; OBRIEN, G. Environment, economy and society: fitting them together into sustainable development. **Sustainable Development**, v. 10, n. 4, p. 187–196, 2002. Citation on page 45.

GIZZI, M. C.; RÄDIKER, S. **The Practice of Qualitative Data Analysis: Research Examples Using MAXQDA**. : BoD–Books on Demand, 2021. Citation on page 104.

GÖHRING, W. The memorandum sustainable information society. In: **18th International Conference on Informatics for Environmental Protection (EnviroInfo)**. Geneva, Switzerland: , 2004. p. 278–286. Citation on page 23.

GUTÉS, M. The concept of weak sustainability. **Ecological Economics**, v. 17, n. 3, p. 147–156, 1996. Citation on page 45.

HANSMANN, R.; MIEG, H.; FRISCHKNECHT, P. Principal sustainability components: Empirical analysis of synergies between the three pillars of sustainability. **International Journal of Sustainable Development & World Ecology**, v. 19, n. 5, p. 451–459, 2012. Citation on page 52.

HASSLER, E.; CARVER, J.; KRAFT, N.; HALE, D. 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. p. 1–10. Citations on pages 51 and 54.

HEAD, L.; GIBSON, C.; GILL, N.; CARR, C.; WAITT, G. A meta-ethnography to synthesise household cultural research for climate change response. **Local Environment**, v. 21, n. 12, p. 1467–1481, 2016. Citation on page 55.

HILL, R.; BOWEN, P. Sustainable construction: principles and a framework for attainment. **Construction Management and Economics**, Informa UK Limited, v. 15, n. 3, p. 223–239, 1997. Citations on pages 38 and 43.

HILTY, L.; LOHMANN, W.; HUANG, E. Sustainability and ict - an overview of the field. **Notizie di POLITEIA**, Tipolito Subalpina, v. 27, n. 104, p. 13–28, 2011. Citations on pages 3, 12, 13, and 43.

HODA, R.; SALLEH, N.; GRUNDY, J.; TEE, H. M. Systematic literature reviews in agile software development: A tertiary study. **Information and Software Technology**, Elsevier BV, v. 85, p. 60–70, May 2017. Citation on page 15.

HUF, C.; RAGGL, A. The normativity of the helping child – meta-ethnographic perspectives on individualised learning in age-mixed classrooms. **Ethnography and Education**, v. 12, n. 2, p. 165–177, 2016. Citation on page 55.

IDRI, A.; HOSNI, M.; ABRAN, A. Systematic literature review of ensemble effort estimation. **Journal of Systems and Software**, Elsevier BV, v. 118, p. 151–175, Aug. 2016. Citation on page 105.

IMTIAZ, S.; BANO, M.; IKRAM, N.; NIAZI, M. A tertiary study: experiences of conducting systematic literature reviews in software engineering. In: **Evaluation and Assessment in Software Engineering (EASE)**. 2013. p. 177–182. Citations on pages 2, 36, 37, 39, 40, 41, 51, 90, 150, 151, 155, 156, and 188.

IWAZAKI, A. Y.; SANTOS, V. dos; FELIZARDO, K. R.; SOUZA, E. F. de; VALENTIM, N. M. C.; NAKAGAWA, E. Y. Benefits and challenges of a graduate course: An experience teaching systematic literature review. In: **2022 IEEE Frontiers in Education Conference (FIE)**. : IEEE, 2022. Citations on pages 2 and 73.

JALALI, S.; WOHLIN, C. Systematic literature studies. In: **ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2012. p. 29–38. Citations on pages 32, 59, and 74.

JOSHI, A.; KALE, S.; CHANDEL, S.; PAL, D. K. Likert scale: Explored and explained. **British journal of applied science & technology**, SCIENCEDOMAIN International, v. 7, n. 4, p. 396, 2015. Citation on page 65.

KAKOS, M.; FRITZSCHE, B. A meta-ethnography of two studies on interactions in schools: reflections on the process of translation. **Ethnography and Education**, v. 12, n. 2, p. 228–242, 2016. Citation on page 55.

KEEBLE, B. The brundtland report:'our common future'. **Medicine and war**, v. 4, n. 1, p. 17–25, 1988. Citations on pages 3, 38, 43, and 45.

KEHOE, R.; JARVIS, A. Overview of software engineering. In: **ISO 9000-3 - A tool for software product and process improvement**. 1996. p. 0–228. Citation on page 28.

KERN, E.; HILTY, L. M.; GULDNER, A.; MAKSIMOV, Y. V.; FILLER, A.; GRöGER, J.; NAUMANN, S. Sustainable software products - towards assessment criteria for resource and energy efficiency. **Future Generation Computer Systems**, v. 86, p. 199–210, 2018. Citation on page 13.

KHAN, R. U.; KHAN, S. U.; KHAN, R. A.; ALI, S. Motivators in green it-outsourcing from vendor's perspective: A systematic literature review. **Pakistan Academy of Sciences**, v. 52, n. 4, p. 345–360, 2015. Citations on pages 18, 59, 62, 159, and 160.

KHATIBSYARBINI, M.; ISA, M. A.; JAWAWI, D. N.; TUMENG, R. Test case prioritization approaches in regression testing: A systematic literature review. **Information and Software Technology**, Elsevier BV, v. 93, p. 74–93, Jan. 2018. Citation on page 95.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004. Citation on page 188.

KITCHENHAM, B.; BRERETON, O. A systematic review of systematic review process research in software engineering. **Information and software technology**, Elsevier, v. 55, n. 12, p. 2049–2075, 2013. Citations on pages 50, 150, 151, 152, 153, 157, 182, and 183.

KITCHENHAM, B.; BRERETON, O.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering–a systematic literature review. **Information and software technology**, Elsevier, v. 51, n. 1, p. 7–15, 2009. Citations on pages 98 and 99.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. **Information and Software Technology**, v. 55, n. 12, p. 2049–2075, 2013. Citations on pages 36, 37, 39, 40, 41, 42, 51, 90, 113, 175, and 188.

KITCHENHAM, B.; BRERETON, P.; LI, Z.; BUDGEN, D.; BURN, A. Repeatability of systematic literature reviews. In: **15th Annual Conf. on Evaluation and Assessment in Software Engineering (EASE)**. 2011. p. 46–55. Citations on pages 2, 36, 42, 50, 60, 61, 62, 86, 165, 166, and 186.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. 2007. Citations on pages 1, 2, 5, 32, 50, 60, 61, 66, 75, 97, 99, 109, 164, 165, 166, 167, 168, 179, 186, and 188.

KITCHENHAM, B.; PRETORIUS, R.; BUDGEN, D.; Pearl Brereton, O.; TURNER, M.; NIAZI, M.; LINKMAN, S. Systematic literature reviews in software engineering – a tertiary study. **Information and Software Technology**, v. 52, n. 8, p. 792 – 805, 2010. ISSN 0950-5849. Citation on page 15.

KITCHENHAM, B. A.; BUDGEN, D.; BRERETON, P. **Evidence-based software engineering and systematic reviews**. : CRC press, 2015. Citations on pages 1, 2, 14, 36, 50, 57, 86, 98, 99, 185, 186, and 196.

KITCHENHAM, B. A.; DYBA, T.; JORGENSEN, M. Evidence-based software engineering. In: **26th International Conference on Software Engineering (ICSE)**. Scotland, United Kingdom: , 2004. p. 273–281. Citations on pages 1, 15, 50, 51, 60, 61, 66, 67, 86, 97, 99, and 166.

KITCHENHAM, B. A.; DYBA, T.; JORGENSEN, M. Evidence-based software engineering. In: **26th International Conference on Software Engineering (ICSE)**. Scotland, United Kingdom: , 2004. p. 273–281. Citation on page 188.

KITCHENHAM, B. A.; MADEYSKI, L.; BUDGEN, D. SEGRESS: Software engineering guidelines for REporting secondary studies. **IEEE Transactions on Software Engineering**, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2022. Citations on pages 99, 100, 109, and 111.

KOLESNICHENKO, O.; MAZELIS, L.; SOTNIK, A.; YAKOVLEVA, D.; AMELKIN, S.; GRIGOREVSKY, I.; KOLESNICHENKO, Y. Sociological modeling of smart city with the implementation of UN sustainable development goals. **Sustainability Science**, Springer Science and Business Media LLC, v. 16, n. 2, p. 581–599, Jan. 2021. Citation on page 31.

KOMEIL, R.; JAMAIAH, Y.; AZIZ, D. Current challenges and conceptual model of green and sustainable software engineering. **Journal of Theoretical and Applied Information Technology**, Little Lion Scientific, 2016. Citations on pages 18, 24, and 163.

KOZIOLEK, H. Sustainability evaluation of software architectures: A systematic review. In: **Joint ACM SIGSOFT Conference – QoSA and ACM SIGSOFT Symposium – ISARCS on Quality of Software Architectures – QoSA and Architecting Critical Systems – ISARCS**. Boulder, Colorado, USA: Association for Computing Machinery, 2011. p. 3–12. Citation on page 18.

KRIPPENDORFF, K. **Content analysis: An introduction to its methodology**. : Sage publications, 2018. Citation on page 102.

KRUCHTEN, P. **The Rational Unified Process: An Introduction**. 3. ed. Boston, MA: Addison-Wesley, 2003. ISBN 0201707101. Citation on page 66.

KUCKARTZ, U.; RäDIKER, S. **Analyzing Qualitative Data with MAXQDA**. Springer International Publishing, 2019. Available: <https://doi.org/10.1007/978-3-030-15671-8>. Citation on page 104.

KUHLMAN, T.; FARRINGTON, J. What is sustainability? **Sustainability**, MDPI AG, v. 2, n. 11, p. 3436–3448, Nov. 2010. Citation on page 3.

KUHRMANN, M.; FERNÁNDEZ, D.; DANEVA, M. On the pragmatic design of literature studies in software engineering: an experience-based guideline. **Empirical Software Engineering**, v. 22, n. 6, p. 2852–2891, 2017. Citations on pages 2, 36, 50, and 86.

LAGO, P.; KOcAK, S. A.; CRNKOVIC, I.; PENZENSTADLER, B. Framing sustainability as a property of software quality. **ACM communications**, v. 58, n. 10, p. 70–78, 2015. ISSN 0001-0782. Citations on pages 12 and 13.

LAGO, P.; KOcAK, S. A.; CRNKOVIC, I.; PENZENSTADLER, B. Framing sustainability as a property of software quality. **Communication of the ACM**, ACM, New York, NY, USA, v. 58, n. 10, p. 70–78, Sep. 2015. ISSN 0001-0782. Citations on pages 53 and 89.

LAGO, P.; MEYER, N.; MORISIO, M.; MÜLLER, H. A.; SCANNIELLO, G. Leveraging energy efficiency to software users: summary of the second greens workshop, at icse 2013. **ACM SIGSOFT Software Engineering Notes**, ACM, v. 39, n. 1, p. 36–38, 2014. Citation on page 12.

LARMAN, C.; BASILI, V. Iterative and incremental developments. A brief history. **Computer**, v. 36, n. 6, p. 47–56, 2003. Citation on page 66.

LAVALLEE, M.; ROBILLARD, P.-N.; MIRSALARI, R. Performing systematic literature reviews with novices: An iterative approach. **IEEE Transactions on Education**, v. 57, n. 3, p. 175–181, 2014. Citation on page 67.

LEE, J.; LEE, D.; KANG, S. An overview of the business process maturity model (BPMM). In: **Int. Conf. on Web-Age Information Management (WAIM), LNCS**. 2007. p. 384–395. Citation on page 187.

LEVY, M.; HAZZAN, O. Knowledge management in practice: The case of agile software development. In: **Workshop on Cooperative and Human Aspects on Software Engineering (CHASE)**. 2009. p. 60–65. Citation on page 67.

LI, L.; BISSYANDÉ, T. F.; PAPADAKIS, M.; RASTHOFER, S.; BARTEL, A.; OCTEAU, D.; KLEIN, J.; TRAON, L. Static analysis of android apps: A systematic literature review. **Information and Software Technology**, Elsevier BV, v. 88, p. 67–95, 2017. Citation on page 95.

LI, N.; SHEPPERD, M.; GUO, Y. A systematic review of unsupervised learning techniques for software defect prediction. **Information and Software Technology**, Elsevier BV, v. 122, p. 106287, Jun. 2020. Citation on page 95.

MACDONELL, S.; SHEPPERD, M.; KITCHENHAM, B.; MENDES, E. How reliable are systematic reviews in empirical software engineering? **IEEE Transactions on Software Engineering**, v. 36, n. 5, p. 676–687, 2010. Citations on pages 60, 61, 62, 68, 69, and 164.

MACDONELL, S.; SHEPPERD, M.; KITCHENHAM, B.; MENDES, E. How reliable are systematic reviews in empirical software engineering? **IEEE Transactions on Software Engineering**, v. 36, n. 5, p. 676–687, 2010. Citation on page 95.

MARIMUTHU, C.; CHANDRASEKARAN, K. Software engineering aspects of green and sustainable software: A systematic mapping study. In: **10th Innovations in Software Engineering Conference**. Jaipur, India: Association for Computing Machinery, 2017. p. 34–44. Citations on pages 13, 18, and 51.

MARSHALL, C.; BRERETON, P. Tools to support systematic literature reviews in software engineering: A mapping study. In: **International Symposium on Empirical Software Engineering and Measurement**. Baltimore, USA: , 2013. p. 296–299. Citations on pages 2, 36, 60, 61, 186, and 191.

MARSHALL, C.; BRERETON, P. Systematic review toolbox. In: **19th International Conference on Evaluation and Assessment in Software Engineering**. Nanjing, China: ACM, 2015. p. 1–6. Citations on pages 2, 40, 104, and 169.

MARSHALL, C.; KITCHENHAM, B.; BRERETON, P. Tool features to support systematic reviews in software engineering - A cross domain study. **e-Informatica**, v. 12, n. 1, p. 79–115, 2018. Citations on pages 74 and 177.

MENDES, E.; FELIZARDO, K.; WOHLIN, C.; KALINOWSKI, M. Search strategy to update systematic literature reviews in software engineering. In: **45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**. 2019. p. 355–362. Citation on page 74.

MENDES, E.; WOHLIN, C.; FELIZARDO, K.; KALINOWSKI, M. When to update systematic literature reviews in software engineering. **Journal of Systems and Software**, v. 167, p. 110607, 2020. Citations on pages 14, 36, and 185.

MENDES, E.; WOHLIN, C.; FELIZARDO, K.; KALINOWSKI, M. When to update systematic literature reviews in software engineering. **Journal of Systems and Software**, v. 167, p. 110607, 2020. Citations on pages 54, 60, 61, 62, 97, 105, 168, 179, and 186.

MENDEZ, D.; GRAZIOTIN, D.; WAGNER, S.; SEIBOLD, H. Open science in software engineering. In: **Contemporary Empirical Methods in Software Engineering**. 2020. p. 477–501. Citations on pages 67, 165, and 176.

MERGEL, G.; SILVEIRA, M.; SILVA, T. A method to support search string building in systematic literature reviews through visual text mining. In: **30th Annual ACM Symposium on Applied Computing (SAC)**. 2015. p. 1594–1601. Citation on page 77.

MESHKAT, B.; COWMAN, S.; GETHIN, G.; RYAN, K.; WILEY, M.; BRICK, A.; CLARKE, E.; MULLIGAN, E. Using an e-Delphi technique in achieving consensus across disciplines for developing best practice in day surgery in Ireland. **Journal of Hospital Administration**, v. 3, n. 4, p. 1–8, 2014. Citation on page 65.

MOISES, A. C.; MALUCELLI, A.; REINEHR, S. Practices of energy consumption for sustainable software engineering. In: **9th International Green and Sustainable Computing Conference**. Pittsburgh, PA, USA: , 2018. p. 1–6. Citations on pages 18, 59, 61, 160, 162, and 163.

MORANDÍN-AHUERMA, I.; CONTRERAS-HERNÁNDEZ, A.; AYALA-ORTIZ, D.; PÉREZ-MAQUEO, O. Socio–ecosystemic sustainability. **Sustainability**, v. 11, n. 12, 2019. Citation on page 45.

MOURÃO, B.; KARITA, L.; MACHADO, I. Green and sustainable software engineering - a systematic mapping study. In: **17th Brazilian Symposium on Software Quality (SBQS)**. 2018. p. 121–130. Citations on pages 13, 18, 51, 59, 61, 68, and 159.

MOURÃO, E.; PIMENTEL, J.; MURTA, L.; KALINOWSKI, M.; MENDES, E.; WOHLIN, C. On the performance of hybrid search strategies for systematic literature reviews in software engineering. **Information and Software Technology**, v. 123, p. 106294, 2020. Citations on pages 51 and 96.

NAPOLEÃO, B.; FELIZARDO, K.; SOUZA, E.; PETRILLO, F.; HALLE, S.; VIJAYKUMAR, N.; NAKAGAWA, E. Establishing a search string to detect secondary studies in software engineering. In: **47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**. 2021. p. 9–16. Citation on page 78.

NAPOLEÃO, B.; PETRILLO, F.; HALLE, S.; KALINOWSKI, M. Towards continuous systematic literature review in software engineering. In: **48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**. Gran Canaria - Spain: IEEE, 2022. p. 467–474. Citations on pages 81, 112, and 122.

NAPOLEÃO, B. M.; de Souza Érica F.; RUIZ, G. A.; FELIZARDO, K. R.; MEINERZ, G. V.; VIJAYKUMAR, N. L. Synthesizing researches on knowledge management and agile software development using the meta-ethnography method. **Journal of Systems and Software**, v. 178, p. 110973, 2021. ISSN 0164-1212. Citations on pages 17 and 176.

NATURE, I. U. for Conservation of; FUND, W. W. **World conservation strategy: Living resource conservation for sustainable development**. : Gland, Switzerland: IUCN, 1980. Citation on page 88.

NAUMANN, S.; DICK, M.; KERN, E.; JOHANN, T. The GREENSOFT model: A reference model for green and sustainable software and its engineering. **Sustainable Computing: Informatics and Systems**, v. 1, n. 4, p. 294–304, 2011. Citation on page 68.

NAZIR, S.; FATIMA, N.; CHUPRAT, S. Situational factors for modern code review to support software engineers' sustainability. **International Journal of Advanced Computer Science and Applications**, The Science and Information Organization, v. 11, n. 1, 2020. Citations on pages 18, 59, 62, 67, 160, 161, 162, and 163.

NAZIR, S.; FATIMA, N.; CHUPRAT, S.; SARKAN, H.; F, N.; SJARIF, N. N. Sustainable software engineering:a perspective of individual sustainability. **International Journal on Advanced Science, Engineering and Information Technology**, Insight Society, v. 10, n. 2, p. 676, Apr. 2020. Citations on pages 13, 18, 24, 59, 62, 67, 160, 161, and 162.

NEPOMUCENO, V.; SOARES, S. Maintaining systematic literature reviews: Benefits and drawbacks. In: **12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2018. p. 1–4. ISBN 9781450358231. Citation on page 179.

NEPOMUCENO, V.; SOARES, S. On the need to update systematic literature reviews. **Information and Software Technology**, v. 109, p. 40–42, 2019. Citation on page 105.

NEUMAYER, E. **Weak versus Strong Sustainability**. : Edward Elgar Publishing, 2013. Citation on page 45.

NIAZI, M. Do systematic literature reviews outperform informal literature reviews in the software engineering domain? an initial case study. **Arabian Journal for Science and Engineering**, v. 40, n. 3, p. 845–855, 2015. Citations on pages 2, 36, 50, 86, and 186.

NOBLIT, G.; HARE, R. **Meta-ethnography: Synthesizing qualitative studies**. : Sage, 1988. Citations on pages 6, 52, 54, 55, 60, 63, and 80.

NUSSBAUMER-STREIT, B.; ELLEN, M.; KLERINGS, I.; SFETCU, R.; RIVA, N.; MAHMIĆ-KAKNJO, M.; POULENTZAS, G.; MARTINEZ, P.; BALADIA, E.; ZIGANSHINA, L.; MARQUÉS, M.; AGUILAR, L.; KASSIANOS, A.; FRAMPTON, G.; SILVA, A.; AFFENGRUBER, L.; SPJKER, R.; THOMAS, J.; BERG, R.; KONTOGIANI, M.; SOUSA, M.; KONTOGIORGIS, C.; GARTLEHNER, G. Resource use during systematic review production varies widely: a scoping review. **Journal of Clinical Epidemiology**, v. 139, p. 287–296, 2021. Citations on pages 68 and 178.

OJAMERUAYE, B.; BAHSOON, R.; DUBOC, L. Sustainability debt: A portfolio-based approach for evaluating sustainability requirements in architectures. In: **38th International Conference on Software Engineering Companion (ICSE)**. Austin, USA: , 2016. p. 543–552. Citation on page 12.

OJAMERUAYE, B.; BAHSOON, R.; DUBOC, L. Sustainability debt: a portfolio-based approach for evaluating sustainability requirements in architectures. In: ACM. **38th International Conference on Software Engineering Companion (ICSE'16)**. 2016. p. 543–552. Citations on pages 53 and 89.

OKOLI, C.; PAWLOWSKI, S. D. The delphi method as a research tool: an example, design considerations and applications. **Information & Management**, Elsevier BV, v. 42, n. 1, p. 15–29, Dec. 2004. Citation on page 64.

OMANN, I.; SPANGENBERG, J. H. Assessing social sustainability the social dimension of sustainability in a socio-economic scenario. In: **Biennial Conference of the International Society for Ecological Economics**. Sousse, Tunisia: , 2002. p. 1–22. Citations on pages 12 and 13.

PAN, R.; BAGHERZADEH, M.; GHALEB, T. A.; BRIAND, L. Test case selection and prioritization using machine learning: a systematic literature review. **Empirical Software Engineering**, Springer Science and Business Media LLC, v. 27, n. 2, Dec. 2021. Citation on page 105.

PAPAVASILEIOU, E.; CORNELIS, J.; JANSEN, B. A systematic literature review of the successors of "NeuroEvolution of augmenting topologies". **Evolutionary Computation**, MIT Press - Journals, v. 29, n. 1, p. 1–73, Mar. 2021. Citation on page 105.

PAULK, M. C. A history of the capability maturity model for software. **ASQ Software Quality Professional**, Citeseer, v. 12, n. 1, p. 5–19, 2009. Citations on pages 187 and 188.

PEJCINOVIC, B. Development and uses of iterative systematic literature reviews in electrical engineering education. **Electrical and Computer Engineering Faculty Publications and Presentations**, v. 327, n. 1, p. 1–10, 2015. Citations on pages 50 and 86.

PENZENSTADLER, B.; BAUER, V.; CALERO, C.; FRANCH, X. Sustainability in software engineering: A systematic literature review. In: **16th International Conference on Evaluation Assessment in Software Engineering (EASE)**. Ciudad Real, Spain: , 2012. p. 32–41. Citations on pages 12, 15, 18, 59, 61, 159, and 160.

PENZENSTADLER, B.; DUBOC, L.; VENTERS, C. C.; BETZ, S.; SEYFF, N.; WNUK, K.; CHITCHYAN, R.; EASTERBROOK, S. M.; BECKER, C. Software engineering for sustainability: Find the leverage points! **IEEE Software**, v. 35, n. 4, p. 22–33, 2018. Citations on pages 54 and 89.

PENZENSTADLER, B.; FEMMER, H. Towards a definition of sustainability in and for software engineering. In: **28th ACM Symposium on Applied Computing (SAC'13)**. Coimbra, Portugal: ACM, 2013. v. 13, p. 18–22. Citation on page 12.

PENZENSTADLER, B.; FLEISCHMANN, A. Teach sustainability in software engineering? In: **Conference on Software Engineering Education and Training (CSEE&T)**. Honolulu, Hawai: , 2011. p. 454–458. Citation on page 12.

PENZENSTADLER, B.; RATURI, A.; RICHARDSON, D.; CALERO, C.; FEMMER, H.; FRANCH, X. Systematic mapping study on software engineering for sustainability (se4s). In: **18th International Conference on Evaluation and Assessment in Software Engineering**. New York, NY, USA: , 2014. p. 1–14. Citations on pages 4, 15, 18, 43, 59, 61, 88, and 159.

PÉREZ, J.; DÍAZ, J.; GARCIA-MARTIN, J.; TABUENCA, B. Systematic literature reviews in software engineering—enhancement of the study selection process using cohen's kappa statistic. **Journal of Systems and Software**, Elsevier BV, v. 168, p. 110657, Oct. 2020. Citation on page 102.

PETERSEN, K.; ALI, N. Identifying strategies for study selection in systematic reviews and maps. In: **1st International Symposium on Empirical Software Engineering and Measurement**. 2011. p. 351–354. Citations on pages 2, 57, 75, and 92.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: **12th International Conference on Evaluation and Assessment in Software Engineering**. Swindon, UK: , 2008. p. 68–77. Citation on page 188.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology**, v. 64, p. 1–18, 2015. Citation on page 188.

PETERSEN, K.; WOHLIN, C.; BACA, D. The waterfall model in large-scale development. In: **Lecture Notes in Business Information Processing**. 2009. p. 386–400. Citation on page 66.

PURVIS, B.; MAO, Y.; ROBINSON, D. Three pillars of sustainability: in search of conceptual origins. **Sustainability Science**, v. 14, n. 3, p. 681–695, 2018. Citations on pages 3, 38, 43, and 45.

PURVIS, B.; MAO, Y.; ROBINSON, D. Three pillars of sustainability: in search of conceptual origins. **Sustainability Science**, v. 14, n. 3, p. 681–695, 2018. Citations on pages 3, 51, 52, and 88.

RADJENOVIĆ, D.; HERIČKO, M.; TORKAR, R.; ŽIVKOVIČ, A. Software fault prediction metrics: A systematic literature review. **Information and Software Technology**, v. 55, n. 8, p. 1397–1418, 2013. Citation on page 95.

RAISIAN, K.; YAHAYA, J.; DERAMAN, A. Current challenges and conceptual model of green and sustainable software engineering. **Journal of Theoretical and Applied Information Technology**, v. 94, p. 428–443, 2016. Citation on page 43.

RASHID, N.; KHAN, S. Developing green and sustainable software using agile methods in global software development: Risk factors for vendors. In: **11th International Conference on Evaluation of Novel Software Approaches to Software Engineering (ENASE)**. 2016. p. 247–253. ISBN 9789897581892. Citations on pages 159 and 161.

RASHID, N.; KHAN, S. U. Using agile methods for the development of green and sustainable software: Success factors for GSD vendors. **Journal of Software: Evolution and Process**, Wiley, v. 30, n. 8, p. e1927, Dec. 2017. Citations on pages 13 and 18.

RASHID, N.; KHAN, S. U. Agile practices for global software development vendors in the development of green and sustainable software. **Journal of Software: Evolution and Process**, Wiley Online Library, v. 30, n. 10, p. e1964, 2018. Citations on pages 13, 15, 18, 59, 61, 67, 68, 159, 160, 161, 162, and 163.

RAZAVIAN, M.; PROCACCIANTI, G.; TAMBURRI, D. A. Four-dimensional sustainable e-services. In: **Informatics for Environmental Protection, Sustainable Development and Risk Management**. Oldenburg, Germany: , 2014. Citation on page 13.

RAZAVIAN, M.; PROCACCIANTI, G.; TAMBURRI, D. A. Four-dimensional sustainable e-services. In: **28th International Conference on Informatics for Environmental Protection (EnviroInfo'14)**. 2014. p. 221–228. Citations on pages 53 and 89.

REVIEWS, N. C. for; DISSEMINATION. The database of abstracts of reviews of effects (dare). The University of York, v. 6, n. 2, 2002. Citation on page 188.

RIAZ, M.; SULAYMAN, M.; SALLEH, N.; MENDES, E. Experiences conducting systematic reviews from novices' perspective. In: **14th International Conference on Evaluation and Assessment in Software Engineering (EASE'10)**. 2010. p. 44–53. Citations on pages 2, 150, 151, 152, 153, 154, 155, 156, and 157.

RIAZ, M.; SULAYMAN, M.; SALLEH, N.; MENDES, E. Experiences conducting systematic reviews from novices' perspective. In: **14th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. 2010. p. 44–53. Citations on pages 36, 37, 39, 40, 41, 42, 51, and 90.

ROBERTS, I.; KER, K. How systematic reviews cause research waste. **The Lancet**, v. 386, n. 10003, p. 1536, 2015. Citation on page 63.

ROYCE, W. Managing the development of large software systems: concepts and techniques. In: **9th International Conference on Software Engineering (ICSE)**. 1987. p. 328–338. Citation on page 66.

SACKETT, D. L. Evidence-based medicine. **Seminars in Perinatology**, Elsevier BV, v. 21, n. 1, p. 3–5, Feb. 1997. Citation on page 1.

SAILLARD, E. K. Systematic versus interpretive analysis with two caqdas packages: Nvivo and maxqda. **Forum Qualitative Sozialforschung / Forum: Qualitative Social Research**, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Vol 12, p. No 1 (2011): The KWALON Experiment: Discussions on Qualitative Data Analysis Software by Developers and Users, 2011. Citation on page 104.

SAKUMA, N.; TRENCHER, G.; YARIME, M.; ONUKI, M. A comparison of smart city research and practice in sweden and japan: trends and opportunities identified from a literature review and co-occurrence network analysis. **Sustainability Science**, Springer Science and Business Media LLC, Jul. 2021. Citation on page 31.

SALAM, M.; KHAN, S. U. Systematic literature review protocol for green software multi-sourcing with preliminary results. **Pakistan Academy of sciences**, v. 52, p. 285–300, 2015. Citation on page 18.

SALAM, M.; KHAN, S. U. Developing green and sustainable software: Success factors for vendors. In: **7th IEEE International Conference on Software Engineering and Service Science**. Beijing,China: , 2016. p. 1059–1062. Citations on pages 18, 59, 61, 159, and 161.

SANTOS, R.; SILVA, F. Motivation to perform systematic reviews and their impact on software engineering practice. In: **7th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**. 2013. p. 292–295. Citation on page 54.

SANTOS, V.; IWAZAKI, A.; FELIZARDO, K.; SOUZA, E. de; NAKAGAWA, E. Towards sustainability of systematic literature reviews. In: ____. **15th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Palermo, Italy: ACM, 2021. Citations on pages 51, 54, 87, and 90.

SANTOS, V. dos; ANDERSON, I. Y.; FELIZARDO, K. R.; SOUZA, E. F.; NAKAGAWA, E. Y. Sustainable systematic literature reviews. Elsevier BV, 2023. Citations on pages 87, 90, and 92.

SCHIAVO, J. H. PROSPERO: An international register of systematic review protocols. **Medical Reference Services Quarterly**, v. 38, n. 2, p. 171–180, 2019. Citation on page 182.

SHEN, L.; DU, X.; CHENG, G.; WEI, X. Capability maturity model (CMM) method for assessing the performance of low-carbon city practice. **Environmental Impact Assessment Review**, v. 87, p. 106549, 2021. Citation on page 188.

SHINGARI, D.; ARUNKUMAR, A.; GAUDETTE, B.; VRUDHULA, S.; WU, C.-J. DORA: Optimizing smartphone energy efficiency and web browser performance under interference. In: **International Symposium on Performance Analysis of Systems and Software (ISPASS)**. Belfast, United Kingdom: , 2018. Citations on pages 21 and 23.

SILVA, F. Q. da; SANTOS, A. L.; SOARES, S.; FRANcA, A. C. C.; MONTEIRO, C. V.; MACIEL, F. F. Six years of systematic literature reviews in software engineering: An updated tertiary study. **Information and Software Technology**, v. 53, n. 9, p. 899–913, 2011. Citation on page 188.

SKULMOSKI, G. J.; HARTMAN, F. T.; KRAHN, J. The delphi method for graduate research. **Journal of Information Technology Education: Research**, Informing Science Institute, v. 6, p. 001–021, 2007. Citation on page 64.

SOINI, K.; BIRKELAND, I. Exploring the scientific discourse on cultural sustainability. **Geoforum**, v. 51, p. 213–223, 2014. Citations on pages 38 and 43.

SPANGENBERG, J. Towards indicators for institutional sustainability: lessons from an analysis of agenda 21. **Ecological Indicators**, v. 2, n. 1-2, p. 61–77, 2002. Citations on pages 38 and 43.

STAPLES, M.; NIAZI, M. Experiences using systematic review guidelines. **Journal of Systems and Software**, Elsevier BV, v. 80, n. 9, p. 1425–1437, Sep. 2007. Citation on page 98.

STRAUSS, A.; CORBIN, J. **Basics of qualitative research: Techniques and procedures for developing grounded theory**. : Sage Publications, Inc., 1998. Citation on page 188.

TEAM, C. P. **CMMI for Development, Version 1.3**. Pittsburgh, PA, 2010. Citation on page 187.

TIEPPO, E.; SANTOS, R. R. dos; BARDDAL, J. P.; NIEVOLA, J. C. Hierarchical classification of data streams: a systematic literature review. **Artificial Intelligence Review**, Springer Science and Business Media LLC, v. 55, n. 4, p. 3243–3282, Oct. 2021. Citations on pages 97 and 105.

VACCA, A.; SORBO, A. D.; VISAGGIO, C. A.; CANFORA, G. A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges. **Journal of Systems and Software**, Elsevier BV, v. 174, p. 110891, Apr. 2021. Citation on page 95.

VENTERS, C.; BECKER, C.; BETZ, S.; CHITCHYAN, R.; DUBOC, L.; EASTERBROOK, S.; PENZENSTADLER, B.; RODRIGUEZ-NAVAS, G.; SEYFF, N. Mind the gap: bridging the sustainable software systems research divide. **Journal of Systems and Software**, v. 138, p. 174–188, 2015. Citation on page 53.

VENTERS, C.; SEYFF, N.; BECKER, C.; BETZ, S.; CHITCHYAN, R.; DUBOC, L.; MCINTYRE, D.; PENZENSTADLER, B. Characterising sustainability requirements: A new species red herring or just an odd fish? In: **39th International Conference on Software Engineering: Software Engineering in Society Track (ICSE-SEIS)**. 2017. p. 3–12. Citation on page 12.

VENTERS, C. C.; LAU, L.; GRIFFITHS, M. K.; HOLMES, V.; WARD, R. R.; JAY, C.; DIBSDALE, C. E.; XU, J. The blind men and the elephant: Towards an empirical evaluation framework for software sustainability. **Journal of Open Research Software**, v. 2, n. 1, 2014. Citation on page 13.

WANG, Y.-F.; HSU, Y.-F.; FANG, K. The key elements of gamification in corporate training – The Delphi method. **Entertainment Computing**, v. 40, p. 100463, 2022. Citation on page 64.

WATANABE, W.; FELIZARDO, K.; CANDIDO, A.; SOUZA, É.; NETO, J. C.; VIJAYKUMAR, N. Reducing efforts of software engineering systematic literature reviews updates using text classification. **Information and Software Technology**, v. 128, p. 106395, 2020. Citations on pages 51, 54, 75, 77, 178, and 186.

WEBSTER, J.; WATSON, R. T. Analyzing the past to prepare for the future: Writing a literature review. **MIS quarterly**, JSTOR, p. xiii–xxiii, 2002. Citation on page 98.

WEIDE, B. W.; OGDEN, W. F.; ZWEBEN, S. H. Reusable software components. In: **Advances in Computers**. : Elsevier, 1991. p. 1–65. Citation on page 187.

WEN, J.; LI, S.; LIN, Z.; HU, Y.; HUANG, C. Systematic literature review of machine learning based software development effort estimation models. **Information and Software Technology**, Elsevier BV, v. 54, n. 1, p. 41–59, Jan. 2012. Citations on pages 95 and 104.

WENDLER, R. The maturity of maturity model research: A systematic mapping study. **Information and Software Technology**, Elsevier BV, v. 54, n. 12, p. 1317–1339, 2012. Citation on page 187.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: **18th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. London, United Kingdom: , 2014. p. 1–10. Citation on page 17.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: **18th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. London, United Kingdom: , 2014. p. 1–10. Citations on pages 98 and 188.

WOHLIN, C. Second-generation systematic literature studies using snowballing. In: **20th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. 2016. p. 1–6. Citations on pages 54, 74, 178, and 179.

WOHLIN, C.; KALINOWSKI, M.; FELIZARDO, K.; MENDES, E. Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. **Information and Software Technology**, v. 147, p. 106908, 2022. Citation on page 75.

WOHLIN, C.; MENDES, E.; FELIZARDO, K. R.; KALINOWSKI, M. Guidelines for the search strategy to update systematic literature reviews in software engineering. **Information and Software Technology**, Elsevier BV, v. 127, p. 106366, Nov. 2020. Citations on pages 60, 61, 62, 67, 75, 105, 165, 166, 167, and 168.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M.; REGNELL, B.; WESSLÉN, A. **Experimentation in software engineering**. Berlim: Springer Science & Business Media, 2012. Citations on pages 65, 82, and 112.

WOHLIN, C.; RUNESON, P.; NETO, P. S.; ENGSTRöM, E.; MACHADO, I.; ALMEIDA, E. On the reliability of mapping studies in software engineering. **Journal of Systems and Software**, v. 86, n. 10, p. 2594–2610, 2013. Citation on page 69.

WOLFRAM, N.; LAGO, P.; OSBORNE, F. Sustainability in software engineering. In: **5th Sustainable Internet and ICT for Sustainability**. Funchal, Portugal: , 2017. p. 1–7. Citations on pages 13 and 18.

ZHANG, H.; BABAR, M. A. On searching relevant studies in software engineering. In: **Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering**. Swindon, GBR: , 2010. (EASE'10), p. 111–120. Citation on page 195.

ZHANG, H.; MUHAMMAD, A. Systematic reviews in software engineering: An empirical investigation. **Information and Software Technology**, v. 55, n. 7, p. 1341–1354, 2012. Citations on pages 50 and 86.

ZHOU, X.; YUQIN, J.; ZHANG, H.; LI, S.; HUANG, X. A map of threats to validity of systematic literature reviews in software engineering. In: **APSEC**. 2016. p. 153–160. Citations on pages 36, 37, 38, 39, 40, 41, 42, 50, 51, 78, 86, 90, 150, 151, 154, 155, 156, 157, 186, and 188.

ZHOU, Y.; ZHANG, H.; HUANG, X.; YANG, S.; BABAR, M. A.; TANG, H. Quality assessment of systematic reviews in software engineering. In: **19th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. 2015. p. 1–14. Citations on pages 18, 59, and 78.

APPENDIX

A

Supplementary materials for Chapter 2

Table 21 – Understandings and characteristics of sustainable software systems

ID	Understanding	Characteristic
S1, S3, S4, S5, S7, S11, S14, S18, S19, S20	Sustainable software is that whose production has minimal direct or indirect negative effect on a country's economy, people, society, and environment, while it also impacts positively the sustainable software production (i.e., the process of creating software aligned with sustainable software engineering best practices)	• Minimal negative impacts on economy, society, human beings, and environment • Positive impact on sustainable software production
S4, S5	Green software generates less e-waste as possible during its development and operation.	• Less e-waste
S2, S12, S16	Sustainable software handles economic, environmental, social, and individual sustainability. Sustainable software does not impact the capacity of the future human generation at the expense of the fulfillment of their particular needs.	• Alignment to the economic, environmental, social, and individual sustainability • Less impact on future human generations
S6, S15	Green software engineering aims to create reliable and durable software that meets users' needs while reducing environmental impacts.	• Reliable software • Long-living software • Less environmental impact
S12	A sustainable software make a responsible use of ecological, human, and financial resources.	• Responsible use of resources
S18	25010+S is presented adapting ISO/IEC 25010 to comply with sustainability. ISO/IEC 25010 states reliability, functionality, usability, and efficiency as quality criteria. ISO/IEC 25010+S includes resource usage and energy impact as environmental quality criteria. Other characteristics are considered sustainable sub-characteristics, such as time behavior, resource utilization, learnability, accessibility, reusability, modifiability, adaptability.	• Easy-to-learn software (learnability) • Accessible software • Reusable software • Modifiable software • Adaptable software
S9, S16	Sustainable software is long-living and energy efficient, while it reduces the negative impact on the environment.	• Long-living software • Energy-efficiency software • Less environmental impact

Continues on next page

Table 21 – continued from previous page

ID	Understanding	Characteristic
S22, S23	Sustainable software refers to its capacity to endure over time and can be continuously assessed, documented, and maintained.	• Long-living software • Software continuously assessed • Software continuously documented • Software continuously maintained
S1	Sustainable software avoids extensive documentation and non-essential deliverable during its development. Sustainable software is polymorphic in the sense it can be adjusted to meet both its present and long-term goals. Sustainable software requires less effort for its development and consumes less energy.	• Non-extensive software documentation • Essential software deliverable production • Polymorphic design • Satisfaction of current and long-term software goals • Less development effort • Less energy consumption
S2	Green or sustainable software positively affects the sustainable development and/or whose immediate/backhanded negative effects on the economy, society, individuals, and environment that come from the use of the software are negligible. Sustainable software is a software that brings profits.	• Positive effect on sustainable development • Negligible effect on the economy, society, individuals, and environment • Profitable software
S8	Green or sustainable software follows development process with practical application of sustainability aspects (e.g., usage of refactoring techniques and energy-saving coding).	• Sustainability-driven software development
S11	Green software is continuously assessed and documented over its whole life cycle and also continuously reused.	• Software continuously assessed • Software continuously documented • Software continuously reused
S12	Sustainable software aims to maintain its viability during a maintenance period or replacement by a new system (with continuous monitoring of quality, and knowledge management). In addition, sustainability can also be considered in software usage, whose processes in the application domain are triggered by the software system as product.	• Software viability during maintenance or replacement • Software that leverages sustainability during its usage
S19	Sustainable software is an environment-friendly software that promotes the energy efficiency, minimizes the environmental impact of the processes it supports, and has a positive impact on social and/or economic sustainability. Green software refers to that developed for domains focusing on the preservation of the environment and manages energy-intensive applications. Sustainable software is one that its source code is sustainable (e.g., reusable, energy-efficient, and maintainable) and has the purpose of achieving sustainability goals.	• Energy efficiency • Minimal environmental impact • Positive impact on social sustainability • Positive impact on economic sustainability • Management of energy-intensive applications • Alignment to the sustainability goals
S20	Green software uses more efficiently computing resources while it maintains or increases the overall performance.	• Software performance • Efficient computing resource consumption

Continues on next page

Table 21 – continued from previous page

ID	Understanding	Characteristic
S21	Long-living software systems are sustainable if they can be cost-efficiently maintained and evolved over their entire life-cycle. The quality of software architectures determines sustainability to a large extent. [...] A software-intensive system is long-living if it must be operated for more than 15 years. [...] A long-living software system is sustainable if it can be cost-efficiently maintained and evolved over its entire life-cycle [...]. [...] The opposite of a sustainable software system is a long-living system that cannot be adapted to changing requirements and environments due to unjustifiable costs or even technical infeasibility. [...] The architecture of a sustainable system may evolve during its life-cycle, but the fulfillment of customer requirements within timing, budget, and quality constraints must be assured.	• Long-living software systems • Cost-efficient maintenance • Continuous software evolution • Quality software architecture

Table 22 – Resources mentioned in sustainability of software systems

ID	Text Passages	Resource mentioned
S1, S3	Global software engineering aims to design, develop, and use a software with limited **energy** and **computing resources**. [...] Software industries have now realized the competitive advantages of the integrated approach for producing high-quality software with **accelerated delivery**, **minimal cost**, and flexibility to manage the requirements even late in the development process" [...] Efficient utilization of **time** and **computing resources** [...]	• Electric energy • Production time • Minimal cost • Computing resources
S2	Then, a conceptual model is rendered demonstrating the consolidated life cycles of sustainable product and principle sustainable measurement dimensions, such as **energy** or information efficiency, low **cost** and **human health**.	• Electric energy • Cost • Human Health
S3	Several studies have been conducted to shift the focus in applying agile principles and practices in GSD to get several benefits, such as **lower production cost**, around-the-clock development, **faster time to market**, and the liberty of **involving the most talented developers** across the globe.	• Production cost • Production time • Human resources
S4	There are a number of design activities that support greenness in various phases of software design such as conceptual design, logical design, physical design, data structure and algorithm design. Our findings also show that '**power-saving software** strategies' (70%)' is the second critical factor for green and sustainable software development. [...] Sustainable development refers to resource use for meeting the needs of humans while taking into account the ecological, economic, and societal impacts [...]	• Electric energy
S5	Sustainability is a main concern in our current society. One of the aspects that play an important role in supporting sustainable development is Information Technology (IT). Both software behavior and the way it is developed impact the amount of **energy consumption**.	• Electric energy
S6	This may indicate that researchers expect these three contribution types would promote gains in **energy efficiency** and, consequently, to obtain more sustainable software products	• Electric energy
S7	Information Technology and, particularly, software can be a means to support the challenges that face sustainable development. It can be used to optimize industrial processes as well as to reduce **energy** and resource consumption.	• Electric energy
S8	The energy consumption during the software processing is considered as a first order impact because it directly leads to high **costs** on **energy bills** and consequently on the environment.	• Electric energy • Monetary cost

Continues on next page

Table 22 – continued from previous page

ID	Text Passages	Resource mentioned
S9	The most common way to address sustainability in software is through energy efficiency, but other effects can be considered, such as **energy usage**, E-waste production, emissions caused by required infrastructure (first order effect) [...] The interaction between performance efficiency and **energy efficiency** is what is reported most and there is a fairly positive interaction.	• Electric energy
S10	**Energy-efficiency** optimization would manifest during software runtime, while optimization in the software architecture would make the maintenance phase of software more sustainable." [...] Maintenance process: papers that aim to make the maintenance of a finished product more sustainable by, for example, **reducing the personal, monetary and environmental cost.**	• Monetary cost • Personal cost • Environmental cost • Electric energy
S11	This expresses a need for a more sustainable development (SD) in order to **reduce energy consumption** and greenhouse emissions (GHG)."	• Electric energy
S12	Tool for estimating the **energy consumption** of software" [...] "Development process aspect: Sustainability in the initial system development process (with responsible use of **ecological, human, and financial resources**)	• Electric energy • Financial Resources • Human resources
S13	Creating **energy efficient** and green software is becoming popular subject rapidly" [...] Software development multi-sourcing is emerging global software engineering (GSE) paradigm for producing high quality software at **minimum cost** and **time**.	• Electric energy • Cost • Production time
S14	**Energy-efficient**, minimizes the environmental impact of the processes it supports and has a positive impact on **social and/or economic sustainability**	• Electric energy • Human resources • Monetary cost
S15	In order to make greener software products, **software practitioners need actionable timely information**, to make useful trade-offs between **energy efficiency** and other quality attributes, like performance, during development	• Electric energy • Practitioners need information
S16	Then, Penzenstadler et al. (2013) attached the elements of responsible use of **ecological, human and financial resources** towards sustainability. Next, Calero et al. (2013) contributed that the consumption of resources can improve the performance of **energy** used in producing the software product towards sustainability.	• Ecological • Financial resources • Human resources
S17	Reducing **energy consumption** and carbon footprint, in order to achieve high levels of sustainability" [...] the use of ICT to improve **energy efficiency** and **reduce costs** is the subject of a number of papers in this special theme [...] **Costs** of **human factors** affecting the software life-cycle	• Electric energy • Monetary cost • Human Factors
S18	In addition we would like to remark the following results. **Reusability and Time behavior** are two sub characteristics that we have consider closely related to sustainability however no measures have been founded in this SLR to assess them.	• Production time
S19	Software developers need practical guidance to support sustainability during software development. Although there are scarce results in industrial settings, **we can rely on the expert opinion of researchers** as a basis for providing some general suggestions. They must be adapted considering the particular settings of the project or organization. But other effects can be considered, such as **energy usage**, e-waste production, emissions caused by required infrastructure (first order effect); in addition, there are changes in user behavior caused by software (second order effects); and changes in social behaviors induced by software systems that erode the benefits of **optimizing energy** efficiency (third order effect).	• Human resources • Electric energy
S20	Where the emphasis has been primarily **minimizing power usage** for data-centers and technical equipment's (such as desktops, projectors)	• Electric energy
S21	A long-living software system is sustainable if it can be **cost-efficiently** maintained and evolved over its entire life-cycle.	• Monetary cost
S22	Long term investment and high return on **investment**.	• Monetary Cost
S24	Protect the natural capitals such as **energy**, air, water and optimal utilization of resources	• Electric energy

APPENDIX

B

Supplementary materials for Chapter 3

Table 23 – Data synthesis on current barriers of SLR process

Dim.	Coding	Barrier	Reference
Social	SLR does not present generalizable results which can be used in industry or other areas	The relevant information is apt to be spread among the later sections of a paper, necessitating thorough scrutiny of 'discussion' sections as well as 'conclusions', and sometimes the 'results' sections too. **Only a few papers provided tabulated presentations of results that showed which primary studies supported or refuted a particular conclusion.** **Provenance in the form of a link between a conclusion (or recommendation) and the supporting data is often unclear.** Since we were only prepared to include those conclusions or recommendations that were explicitly supported by the primary studies, this lack of clear links often made it quite difficult to identify where such support existed. **We noted a relative lack of recommendations. Since identifying these is essentially the task of knowledge translation, and this process is still ill-defined even for those disciplines with a longer tradition of using systematic reviews, this should not be that surprising.** Identifying recommendations does also require domain expertise, and it may well be that many of the systematic review teams did not feel confident to do so. (It might be argued that it is better not to do so than to do it badly!)	(BUDGEN *et al.*, 2018b)
		Generalizability (27 studies) refers to the possibility of not being able to generalize the results of the secondary study (for example due to the identification of only a portion of existing primary studies). A special case of this threat that is quite frequently reported is Results not applicable to other organizations or domains (12 studies).	(AMPATZOGLOU *et al.*, 2020)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
Social	Inappropriate research questions are defined	**Inappropriate Research Questions** The researchers are not familiar with this research field, so they may put forward some unsuitable research questions.	(ZHOU *et al.*, 2016)
		Defining Research Questions	(RIAZ *et al.*, 2010a)
		Defining research questions is difficult	(KITCHENHAM; BRERETON, 2013)
		Structured questions not appropriate	(KITCHENHAM; BRERETON, 2013)
		Coverage of research questions	(AMPATZOGLOU *et al.*, 2020)
		Search Strategy depends on Research Question	(IMTIAZ *et al.*, 2013)
Social	Cultural differences may impact the validity results	**Culture Bias** Due to the authors of the cultural differences, the validity of the results is not assured. (e.g., preferences for the studies of some researchers' nationality)	(ZHOU *et al.*, 2016)
		distributed nature of team complicates the process of protocol development	(RIAZ *et al.*, 2010a)
Soc./ Eco.	Data synthesis, interpretation, and presentation of results are damaged when researchers have different interpretations of data extracted or these processes are conducted by a made by a single researcher.	**Subjective Interpretation About the Extracted Data** Researchers may have different interpretation of the extracted data and also different opinions on how to deal with the data.	(ZHOU *et al.*, 2016)
		Researcher bias (40 studies) refers to potential bias the authors of the secondary studies may have, while interpreting or synthesizing the extracted results. This can be a bias towards a certain topic, or because only one author worked on data synthesis.	(AMPATZOGLOU *et al.*, 2020)
		no standard method for synthesizing data from qualitative or mixed methods studies	(RIAZ *et al.*, 2010a)
		data extractors not as familiar with the statistical terms as members who defined data extraction form	(RIAZ *et al.*, 2010a)
		Unsatisfactory Data Synthesis Synthesis may be complete and it merely presents our preliminary synthesis.	(ZHOU *et al.*, 2016)
Economic	Lack of standatization and Incorrect/incomplete search terms hinder the search string construction	**Incorrect or Incomplete Search Terms in Automatic Search** The search string in the searching process may include inadequate search terms related to research topic. Sonia also applied terms patterns and adapted the search string to each digital library with the purpose of making easier the replication of the process [15].	(ZHOU *et al.*, 2016)
		Construction of the search string (92 studies) refers to problems that might occur when the researchers are building the search string. As a consequence, the search might return a large number of primary studies (including many irrelevant ones) or miss some relevant studies.	(AMPATZOGLOU *et al.*, 2020)
		SE keywords are not standardized	(KITCHENHAM; BRERETON, 2013)
		Non-standardized keywords and difficulty in identifying synonyms	(RIAZ *et al.*, 2010a)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
		Careful selection of keywords, synonyms, alternative terms and avoiding homonyms	(IMTIAZ *et al.*, 2013)
		Lack of Standard Languages and Terminologies It means SLRs used different terms for similar concepts. For example, if the basic concept of coupling is not well understood, then the metric that captures coupling may be inaccurate or incorrect.	(ZHOU *et al.*, 2016)
Economic	Inadequate number of studies are evaluated hindering the validity of results	**Inadequate Size and Number of Samples** If the sample size is inadequate or too small to be reasonable, then the validity of the results is not assured.	(ZHOU *et al.*, 2016)
		Not knowing how the threats to the validity were controlled and if including a large number of primary studies is better than conducting high-quality review with more selective quality assessment criteria	(RIAZ *et al.*, 2010a)
Economic	Use of not credible or very specific/broad digital databases may return many irrelevant studies or miss relevant studies	**Incomprehensive Venues or Databases** The library resources used to search for primary studies do not include some important resources databases.	(ZHOU *et al.*, 2016)
		Selection of Digital Libraries (DL) (70 studies) refers to problems that can arise from using very specific, too broad, or not credible search engines. The consequence of this threat can be either the return of a lot irrelevant or the miss of relevant studies. In addition to that Search Engine Inefficiencies (10 studies) pointed out cases when the search engine interface cannot accommodate complex queries.	(AMPATZOGLOU *et al.*, 2020)
		Choosing appropriate digital libraries	(KITCHENHAM; BRERETON, 2013)
		Retrieved results depend on searched Conferences/Journals, databases and indexing systems.	(IMTIAZ *et al.*, 2013)
		not finding many studies	(RIAZ *et al.*, 2010a)
Economic	Primary studies may be duplicated in different databases	**Primary Study Duplication** The duplication of papers is a potential threat (e.g., the same paper included in more than one database or in more than one journal). Researchers need to identify and remove the duplication [16]. A study was reported by more than one paper.	(ZHOU *et al.*, 2016)
Economic	Inaccessibility of resources (papers/databases)	**Paper/database Inaccessible** Some researchers could not download the papers and contact the author(s) of the exact article/some digital libraries which they do not have the authority to access.	(ZHOU *et al.*, 2016)
Eco./Tech.	Database limitations, inefficiencies (e.g., interface, search string syntax)	Limitations of online databases create problems during SLR	(IMTIAZ *et al.*, 2013)
		Databases (DBs) have limitations on Boolean expressions and number of characters	(RIAZ *et al.*, 2010a)
		Digital library interfaces functionality inappropriate for SRs	(KITCHENHAM; BRERETON, 2013)
Eco./Tech.	Need tool to support SLR	Tools that help locate, organize and summarize SLR information required | Digital library interfaces functionality inappropriate for SRs	(KITCHENHAM; BRERETON, 2013)
		Need tool to support SLR	(IMTIAZ *et al.*, 2013)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
		Databases (DBs) have limitations on Boolean expressions and number of characters	(RIAZ *et al.*, 2010a)
Economic	Lack of information about the SLR conduction process (e.g, initial date, duplicated studies).	A noticeable feature in Table 4 is the wide variation in the way that the period covered by a systematic review was reported. In a few cases, we were completely unable to determine what this was, and for many others the information was incomplete, particularly regarding the start date. An open start date is of course acceptable, but this should be stated explicitly. **Very few studies reported the date when searching took place, although knowing this can be important for anyone wanting to extend a review, or replicate it in some form.** Indeed, knowing both the complete value for the end date (e.g. 31st December 2015) and the date when the search was performed gives some idea about completeness. Digital library indexing is not always up to date (nor is the indexing of journals), so to be fairly sure of including all relevant studies published within a given period, it is prudent to conduct the search some time after the end date. Three months would seem to be a reasonable period to allow for this. This issue is relevant for manual searches as well of course.	(BUDGEN *et al.*, 2018b)
		Papers omit information	(KITCHENHAM; BRERETON, 2013)
		limited quantity and quality of record keeping	(RIAZ *et al.*, 2010a)
		Limited space while publishing resulting in assessment of reporting quality rather than research quality	(RIAZ *et al.*, 2010a)
		Not clear if the final protocol should be shown or its evolution process	(RIAZ *et al.*, 2010a)
		Space constraints for papers	(KITCHENHAM; BRERETON, 2013)
Economic	Lack information about the use of inclusion/exclusion criterias.	**The process involved in applying the inclusion/exclusion criteria is not always very clearly reported.** In particular, it was not always clear how many people were involved in assessing each candidate paper, or even how this was organised. Reporting this is important, as from a quality perspective, the reader needs to know how reliable the assessments are likely to be.	(BUDGEN *et al.*, 2018b)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
		A reporting issue that was encountered with a number of the systematic reviews was that **the authors failed to make clear when they were counting papers and when they were counting studies.** Empirical papers quite commonly report the results of more than one study, and for a secondary study it is usually appropriate to treat these as individual inputs. **This complicates the reporting of counts, since for searching and inclusion/exclusion the relevant operational unit is the paper, while for analysis it is the study.** (And of course, there is the added complication that conference papers may be extended for journal publication, making the relationship between papers and studies to be many-to-many.) And as a further complication, a review might include one study from a paper, while excluding another.	(BUDGEN *et al.*, 2018b)
		Paper selection/Inclusion exclusion	(KITCHENHAM; BRERETON, 2013)
Economic	Insufficient details/report about the quality evaluation	**Few of the reviews provided much detail about the quality scores for the primary studies, and there were relatively few examples of the quality score being used in any way during synthesis (or used at all).** In some cases very little detail about this was provided and there were sometimes statements about use that could not be substantiated from the available data. Again, **the associated processes were rarely described adequately, if at all. In this case there are two relevant processes that should be reported**. The process by which the quality questions were derived, such as whether or not they had been used in other studies, or were derived from other sets of questions. Quality questions do need to be relevant to the issues being addressed in the systematic review, and so need to be justified in some way. The process used to derive quality scores, including how many people performed each assessment and what mechanism was used to resolve any differences where there was more than one person performing the task. Table 11 shows how the task of making a quality assessment was organised for the 27 papers that did perform a quality assessment. It is notable that this was less well reported than the procedures used to determine inclusion/exclusion.	(BUDGEN *et al.*, 2018b)
		Quality assessment depends on study type	(KITCHENHAM; BRERETON, 2013)
		Managing quality evaluation of mixed study types	(KITCHENHAM; BRERETON, 2013)
		Problems assessing quality of primary studies and that of SRs	(RIAZ *et al.*, 2010a)
Continues in the next page			

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
Economic	Additional search is not being reported correctly	Many studies report the number of papers that were found per search engine, and a number also report the degree to which later searches found duplicates of those candidates already identified. In general, electronic searching was well reported. **However, additional searching activities were less thoroughly reported, particularly where the use of manual searches of journals or conferences were concerned, and in particular, where any form of snowballing was employed**. We noted one study where, because snowballing had not returned any additional papers, it was not included in the report on searching, and we were only able to identify that snowballing had been used because of a passing mention elsewhere in the report.	(BUDGEN *et al.*, 2018b)
Economic	Lack of details about the primary studies selected (e.g. study context, characteristics of participants, source material, among others)	There are examples of this that range from **providing hardly any information about the primary studies up to the provision of quite detailed information using tables**. Some use one table effectively, others use multiple tables effectively and there is probably no one clear lesson here. Most give bibliographic information. However, **many studies did score zero for this one. This was largely because systematic reviews often provided little information about the primary studies and their characteristics (providing only bibliographic information would lead to a score of 0).**	(BUDGEN *et al.*, 2018b)
Economic	Studies do not specify important SLR details, causing problems to repeatability and replicability	**Non-Specification of SLR Setting and Sufficient Details**. If the SLR's setting (venues, search string, etc.) or other important details are not clearly stated, then this may pose a **threat to repeatability and replicability of the study**.	(ZHOU *et al.*, 2016)
		Repeatability (20 studies) refers to threats that deal with the replication of a secondary study. The most common reason for the existence of such threats is the lack of a detailed protocol, or the existence of researcher and data extraction bias.	(AMPATZOGLOU *et al.*, 2020)
		mechanism to undertake meta-analysis are not clear	(RIAZ *et al.*, 2010a)
Soc. / Eco.	The synthesis process is not clear	Many papers **don't make their synthesis method clear** (including misleading titles about mapping studies) Our analysis of synthesis was complicated by a number of issues. One is that many **authors do not describe the form of synthesis employed**, or if they do, they may have used terms taken from other sources (and the descriptions are not always even consistent across different sections). It is significant that only 13 of the 37 papers described the form(s) of synthesis employed. Inevitably perhaps, there is an overlap between these and the set of studies that used the quality scores during synthesis (130, 134, 138, 157, 205).	(BUDGEN *et al.*, 2018b)
		Not clear how much categorization is done during extraction and how much during data synthesis	(RIAZ *et al.*, 2010a)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
Soc. / Eco.	Lack information about the reviewers' participation during the SLR conduction process	Table 7 shows that many studies did use two reviewers, who then resolved any differences, but **a substantial number still used a single reviewer with a checker**. For many of these, the checker only checked a percentage of the selections (the lowest proportion of checks observed was 5% which is very weak). Five papers didn't report how this was done at all. One of the two papers described as 'other' had a quite complex multiple reviewer structure that was reported very clearly as a table [20], while in the other, the process of inclusion/exclusion appeared to be performed by two people working together, rather as in pair programming [41].	(BUDGEN *et al.*, 2018b)
		lack of consistency among researchers for defining conceptual structure during SR	(RIAZ *et al.*, 2010a)
		Two phases strategy ensures representative set of empirical evidence	(IMTIAZ *et al.*, 2013)
Economic	Many studies do not perform the quality evaluation and others studies do not use it correctly in selection process.	Table 10 shows little to indicate that this is common practice in software engineering, **with only eight studies from 37 using the quality analysis in this way, and half of the studies (19) either performing no scoring of quality or performing one and not using it.** This leaves the question open as to why researchers performed a quality analysis and then made little use of it, other than because it was recommended in the guidelines.**A number of studies (10) used the quality scores as part of the selection process, usually by omitting those primary studies that had scores below some (arbitrary) threshold. Not all of them reported much about the studies that were discarded or about the reasons for choosing a particular threshold value.** One concern about this practice is that the choice of a threshold value introduces an non-systematic element into the selection process. It also muddles the issue of quality assessment with inclusion/exclusion rules. Overall, it seems undesirable to conflate quality assessment with selection in this way. One reason for the profile shown in Table 10 may be that it is linked to a lack of confidence about the process of synthesis, a point that we will return to later.	(BUDGEN *et al.*, 2018b)
		Problems assessing quality of primary studies and that of SRs	(RIAZ *et al.*, 2010a)
		Unavailability of scoring methods for diverse study types	(RIAZ *et al.*, 2010a)
Economic	Few studies had a good search coverage.	Table 6 shows that slightly fewer than two thirds of the studies were considered to have performed a search with good coverage as defined in Table 3.	(BUDGEN *et al.*, 2018b)
Economic	Inappropriate search methods leading to problems in SLR coverage	**Inappropriate Search Method** Researchers apply search methods to search for relevant papers to get evidence for their studies. But imperfect methods may lead to missing relevant papers (e.g. using automatic search only).	(ZHOU *et al.*, 2016)
		Combination of automatic and manual search can ensure coverage	(IMTIAZ *et al.*, 2013)
Continues in the next page			

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
		Including grey literature broadens research	(IMTIAZ *et al.*, 2013)
Economic	Selection process could be difficult when the inclusion/exclusion criteria are generic or inappropriate.	**Inappropriate Inclusion & Exclusion Criteria** The researchers are not familiar with this research field, so they may put forward some unsuitable inclusion & exclusion criteria.	(ZHOU *et al.*, 2016)
		Study inclusion/exclusion bias (100 studies) refers to problems that might occur in the study filtering phase, i.e., when applying the inclusion/exclusion criteria. Such threats are usually found in studies, in which there are conflicting inclusion/exclusion criteria or very generic ones.	(AMPATZOGLOU *et al.*, 2020)
Economic	The wrong classification of the primary studies may cause the secondary study to lack robustness.	**Identification Error of Primary Studies in the Searching Process** It means some errors (e.g., related studies are not chosen or irrelevant studies are chosen), which may be found in the search process.	(ZHOU *et al.*, 2016)
		Robustness of initial classification (35 studies) is applicable to secondary studies, whose data collection relies upon a classification schema. A common practice while performing such an activity is to identify an existing classification schema that (if needed) is tailored to fit the needs of the secondary study. The selection of this initial classification schema poses a threat to validity, since it might not be fitting for the domain, and its tailoring is not efficient.	(AMPATZOGLOU *et al.*, 2020)
		Bias in selection and no inter-rater reliability checks	(RIAZ *et al.*, 2010a)
		Primary studies may be missclassified by researchers	(ZHOU *et al.*, 2016)
Economic	Delimiting a time span affect the coverage of SLR	**Restricted Time Span** Researchers cannot anticipate other relevant studies outside the time span, constructively by limited effort.	(ZHOU *et al.*, 2016)
Economic	Lack of unpartiality of researchers resulting in bias in study selection	**Bias in Study Selection** In the process of search. reviewers have own subjective conjecture, and they do not completely use the inclusion and exclusion criteria for judgment.	(ZHOU *et al.*, 2016)
Economic	Extraction process is difficult when the quality assessment subjectivity is a bias or reviewers do not completely understand the data extraction items	**Bias in Data Extraction** In the process of data extraction, reviewers do not completely understand the definition of data extraction item and the relationship with research questions.	(ZHOU *et al.*, 2016)
		Subjective Quality Assessment It refers to authors' assessment of the criteria based on their own judgment, which may lead to bias and can be a threat. Heckman and Williams stated that their first author assessed the quality of the papers based on 10 questions, which may lead to measurement bias [18].	(ZHOU *et al.*, 2016)

Continues in the next page

Table 23 – continued from previous page			
Dim.	**Coding**	**Barrier**	**Reference**
		Data extraction bias (91 studies) refers to problems that can arise in the data extraction phase. Such problems might be caused from the use of open questions in the collected variables, whose handling is not explicitly discussed in the protocol. A special type of data extraction bias is the Quality assessment subjectivity (13 studies), i.e., the process during which the quality of the primary studies is evaluated by the authors of the secondary study. This threat is relevant only for SLRs that report the evaluation of primary studies' quality.	(AMPATZOGLOU *et al.*, 2020)
Social	Lack of expert evaluation of the results	Need domain knowledge	(KITCHENHAM; BRERETON, 2013)
		Difficulty in initial learning, lack of domain knowledge, and guidance	(RIAZ *et al.*, 2010a)
		education of the supervisors	(RIAZ *et al.*, 2010a)
		Lack of Expert Evaluation The conclusions or results should be evaluated by an expert to understand and interpret their true meaning and significance [19]. Without expert assessment there might be erroneous conclusions reported.	(ZHOU *et al.*, 2016)
		Publication problems due to lack of reviewer's knowledge of SRs	(RIAZ *et al.*, 2010a)
Economic	Data model and data extraction forms may change during extraction	Data model and data extraction forms may change during extraction	(KITCHENHAM; BRERETON, 2013)
Economic	Difficulties in deciding when to stop the piloting process	not clear when to stop the piloting process	(RIAZ *et al.*, 2010a)

APPENDIX C

Supplementary materials for Chapter 4

Table 24 – Text passages on sustainable software systems

Cat	ID	Text Passages	Reference	Codes
Definitions	DF1	Sustainable software is software whose direct and indirect negative impacts on economy, society, human beings, and the environment resulting from development, and usage of the software is minimal and/or has positive effect on sustainable development and post-development phases are negligible and/or which have a positive impact on sustainable production.	(RASHID; KHAN, 2018) (RASHID; KHAN, 2016) (SALAM; KHAN, 2016) (GARCIA-MIRELES, 2017) (GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018) (BERNTSEN *et al.*, 2017) (PENZENSTADLER *et al.*, 2014) (CALERO *et al.*, 2013) (GARCIA-MIRELES *et al.*, 2018) (KHAN *et al.*, 2015)	• Improve production • Reduce negative effects on economy, society, individuals, and environment]
	DF2	Green software is a software that generates as tiny e-waste as possible during its operation and development.	(SALAM; KHAN, 2016) (GARCIA-MIRELES, 2017) (KHAN *et al.*, 2015)	• E-waste
	DF3	Sustainable software is one that satisfy the that satisfy the requirements of the three dimensions of society, economy, environment with responsible use of ecological, human, and financial resources. In addition, a fourth dimension "human sustainability" should be also included. Most cited definition for sustainable software is one that "meet the needs of the present without compromising the ability of future generations to satisfy their own needs"	(PENZENSTADLER *et al.*, 2012) (AHMAD *et al.*, 2014)	• Responsible use of resources • Consider human sustainability • Capacity of future eras to fulfil their own particular needs
	DF4	Green SE aims to create reliable and durable software that meets users' needs while reducing environmental impacts.	(MOURÃO *et al.*, 2018) (ANWAR; PFAHL, 2017)	• Improve Lifetime • Increase Reliability • Reduce Negative effects on environment
	DF5	A sustainable software make responsible use of ecological, human, and financial resources	(PENZENSTADLER *et al.*, 2012) (CALERO *et al.*, 2013)	• Improve Learnability and accessibility, reusability, modifiability and adaptability

Continued on next page

Table 24 – continued from previous page				
Cat	**ID**	**Text Passages**	**Reference**	**Codes**
	DF6	Is a software that avoids non essential deliverables during software development. Also is a project that avoid extensive documentations. Is a software that is polymorphic and fits to meet the long-term goals of project. Is a software that requires less effort for development and consumes less energy.	(RASHID; KHAN, 2018)	• Improve polymorphic Design • Lifetime • Reduce non-essential deliverable • Reduce Extensive documentation • Reduce Effort and Energy consumed
	DF7	Green or Sustainable Software is a software that follow software development process where practical application of sustainability aspects takes place. For instance, software that complies with customer software requirements causing less changes during it scope.	(MOISES *et al.*, 2018)	• Reduce change in software requirements
	DF8	Green software is a software product that over its whole life cycle are continuously assessed, documented, and used for further optimization of the software product.	(CALERO *et al.*, 2013; BERNTSEN *et al.*, 2017)	• Improve continuous evaluation • Improve documentation quality
	DF9	Sustainable software aims to maintain its viability during a maintenance period or replacement by a new system (with continuous monitoring of quality, and knowledge management). In addition, sustainability can also be considered in software usage, whose processes in the application domain are triggered by the software system as product.	(PENZENSTADLER *et al.*, 2012)	• Improve viability of maintenance or replacement • Software that leverages sustainability during its usage
	DF10	In enviromental dimension, a sustainable software promotes energy efficiency, minimizes the environmental impact of the processes it supports, and has a positive impact on social and/or economic sustainability. A sustainable software is a software that follow the guidelines to produce more environment-friendly software. On the other hand, green by software refers to software developed for domains focusing on the preservation of the environment, as well as to software that helps to manage energy-intensive applications.	(GARCIA-MIRELES *et al.*, 2018)	• Improve Energetic efficiency • Reduce Negative effects on economy, society and environment
	DF11	Sustainable software in two ways: as software code that is sustainable, agnostic of purpose; and also as a software purpose directed at achieving sustainability goals.	(GARCIA-MIRELES *et al.*, 2018)	• Improve Polymorphic design
	DF12	Green software uses computing resources more efficiently while mantaining or increasing overall performance.	(KHAN *et al.*, 2015)	• Improve Software performance • Improve Resources consumption efficiency]
	DF13	Sustainability refers 'to capacity to endure' while sustainable software refers to the software that can be continuously assessable, documented and maintainable.	(NAZIR *et al.*, 2020a) (NAZIR *et al.*, 2020b)	• Improve Assessability, documentation quality and maintainability

Continued on next page

Table 24 – continued from previous page				
Cat	**ID**	**Text Passages**	**Reference**	**Codes**
	DF14	A summary of the 25010+S (product quality model and quality in use model) is presented and characteristics from software quality models adapted to complies with sustainability issues are presented. Characteristics from type 1) Time behavior, Resource utilization, Learnability, Accessibility, Reusability, Modifiability, Adaptability, Usefulness, Freedom from risk, Environmental risk mitigation, Context coverage, Context completeness, Flexibility. Characteristics from type 2: Sustainability Functional appropriateness, Sustainability Capacity, Sustainability Appropriateness recognizability, Sustainability Operability, Sustainability Effectiveness, Sustainability Efficiency.	(CALERO *et al.*, 2013)	• Sustainable software characteristics
Critical Factors	F1	Rich communication and collaboration	(RASHID; KHAN, 2018) (RASHID; KHAN, 2016)	• Communication and collaboration
	F2	Power Saving Strategies	(SALAM; KHAN, 2016)	• Energy consume
	F3	Efficient estimation strategies	(SALAM; KHAN, 2016)	• Efficient estimation strategies
	F4	Efficient resources utilization	(RASHID; KHAN, 2016) (SALAM; KHAN, 2016) (NAZIR *et al.*, 2020b)	• Resource utilization
	F5	Software reusability	(SALAM; KHAN, 2016)	• Software reusability
	CF6	Use of Refactoring techniques	(GARCIA-MIRELES *et al.*, 2018)	• Refactoring Techniques
	F7	Run feasibility studies that assess the impact of the project on the environment	(GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018)	• Environmental Feasibility Studies
	F8	Sustainable maintenance of the software	(SALAM; KHAN, 2016)	• Maintenance
	F9	Efficient utilization of time and computing resources	(RASHID; KHAN, 2018)	• Resource utilization
	F10	Lack of customer's presence	(RASHID; KHAN, 2016)	• Customer Presence and Knowledge
	F11	Insufficient knowledge of the customer	(RASHID; KHAN, 2016)	• Customer Presence and Knowledge
	F12	Iterative development	(RASHID; KHAN, 2016)	• Iterative Development
	F13	Lack of methodologies and tool support	(NAZIR *et al.*, 2020a)	• Support Tools
	F14	Lack of experience	(NAZIR *et al.*, 2020a)	• Experience
	F15	Efficient Knowledge sharing	(NAZIR *et al.*, 2020b) (NAZIR *et al.*, 2020a)	• knowledge Management
	F16	Technology Maturity	(NAZIR *et al.*, 2020b)	• Technology Maturity and Accessibility
	F17	Technology Accessibility	(NAZIR *et al.*, 2020b)	• Technology Maturity and Accessibility
	GD1	Schedule daily stand-up meeting to encourage the communication of team members.	(RASHID; KHAN, 2018)	• Communication among team members
	GD2	Arrange face to face meetings and use web technologies such as online chats, webinars, email, and discussion forums for strong interactions between customer and developers.	(RASHID; KHAN, 2018)	• Use technologies to improve team members communication

Continued on next page

Table 24 – continued from previous page

Cat	ID	Text Passages	Reference	Codes
	GD3	Arrange social activities for providing opportunities to the team members for informal communication.	(RASHID; KHAN, 2018)	• Improve communication among team members with social activities
	GD4	Plan interaction sessions among developers and customers to understand the complexity of software project.	(RASHID; KHAN, 2018) (NAZIR *et al.*, 2020a)	• Plan interaction session among stakeholders
	GD5	Encourage informal communication across the team members.	(RASHID; KHAN, 2018)	• Encourage informal communication
	GD6	Feedback from customers during retrospective meeting should be recorded to improve the overall quality of software.	(RASHID; KHAN, 2018)	• Receive feedback from customers to improve quality
	GD7	Perform regular team sessions for effective software development plans (schedule, iteration, quality, and sustainability).	(RASHID; KHAN, 2018)	• Perform regular sessions to discuss software plans
	GD8	Effective knowledge management	(NAZIR *et al.*, 2020b)	• Perform effective knowledge management
	GD9	Identify a sustainability expert within software project	(GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018)	• Identify sustainability expert to support the project
	GD10	Consult relevant experts in any of the sustainability dimensions	(GARCÍA-MIRELES; VILLA-MARTÍNEZ, 2018)	• Identify sustainability expert for support
	GD11	Support from senior managers.	(MOISES *et al.*, 2018)	• Support from senior managers
	GD12	Demonstrate the current tools and methodologies applicability to software engineers.	(NAZIR *et al.*, 2020a)	• Demonstrate tools and methodologies applicability to software engineers
	GD13	Provide knowledge sharing strategies	(NAZIR *et al.*, 2020a)	• Create knowledge sharing strategies
	GD14	Create knowledge base	(NAZIR *et al.*, 2020a)	• Create knowledge base
	GD15	Acquire knowledge by hiring skillful and knowledgeable software engineers	(NAZIR *et al.*, 2020a)	• Hire skillful software engineers to acquire knowledge
	GD16	Avoid knowledge loss by effective knowledge sharing	(NAZIR *et al.*, 2020a)	• Avoid knowledge loss sharing knowledge
	GD17	Develop a framework for individual or group discussions in collective learning and problem solving	(NAZIR *et al.*, 2020a)	• Develop a framework to solve problems and improve group discussions
	GD18	Provide effective knowledge sharing platforms, for example, social media	(NAZIR *et al.*, 2020a)	• Provide knowledge sharing strategies
	GD19	Track the velocity of development to deliver software in time.	(RASHID; KHAN, 2018)	• Track the velocity of development
	GD20	Develop reusable prototype instead of throw-away prototype.	(RASHID; KHAN, 2018)	• Develop reusable prototypes
	GD21	Refactor the software system whenever required	(RASHID; KHAN, 2018) (GARCIA-MIRELES *et al.*, 2018)	
	GD22	Adopt component-based development strategy for reuse of the existing components in the development of a new system to minimize time, cost, and efforts.	(RASHID; KHAN, 2018)	• Adopt component development to improve reuse and minimize time, cost and efforts
	GD23	Highly reusable assets are prone to be optimized, as is their greenability	(GARCIA-MIRELES *et al.*, 2018)	• Develop reusable assets

Continued on next page

Table 24 – continued from previous page				
Cat	**ID**	**Text Passages**	**Reference**	**Codes**
	GD24	If an asset is easy to modify, it is likely to keep (and not worsen) its greenability	(GARCIA-MIRELES *et al.*, 2018)	• Develop modifiable assets to improve sustainability
	GD25	Identify sustainability stakeholders who issues objectives, constraints and consideration about the system under development.	(MOISES *et al.*, 2018)	• Identify sustainability stakeholders
	GD26	Consider business process when Construction a Green System.	(MOISES *et al.*, 2018)	• Consider business process when developing a system
	GD27	Always code the software in pairs (pair programming).	(RASHID; KHAN, 2018) (KOMEIL *et al.*, 2016)	• Always code in pairs (pair programming)
	GD28	Design simple and user friendly software according to customer's needs.	(RASHID; KHAN, 2018) (KOMEIL *et al.*, 2016)	• Design simple user friendly interface
	GD29	Update the design as per the changes in customer's requirements.	(RASHID; KHAN, 2018)	• Update the design according to customers requirements
	GD30	Manage to redesign software system according to the volatile customer's requirements.	(RASHID; KHAN, 2018) (MOISES *et al.*, 2018)	• Manage the redesign according to volatile customer requirements
	GD31	Follow short development cycles/iterations	(RASHID; KHAN, 2018)	• Use short iterations
	GD32	Involve the customer throughout software development to validate the work in iteration.	(RASHID; KHAN, 2018)	• Involve customer in software development
	GD33	Use acceptance testing to validate the requirements of customer.	(RASHID; KHAN, 2018)	• Use acceptance testing to validate requirements of customer
	GD34	Sustainability is more than only maintainability of software. A sustainable product should have a lower impact when introducing changes (reusability) or should improve usability that can make the product life-cycle longer	(GARCIA-MIRELES *et al.*, 2018)	• Sustainable products are reusable and longer lyfe cycle
	GD35	Reliability has a potentially positive interaction with sustainability	(GARCIA-MIRELES *et al.*, 2018)	• Reliability interacts positively with sustainability
	GD36	Usability can have a positive interaction with sustainability when addressing indirect sustainability impacts. Direct sustainability impact can have a negative interaction with usability	(GARCIA-MIRELES *et al.*, 2018)	• Usability interacts positively with sustainability
	GD37	Quality attributes such as maintainability, portability and usability promote software perdurability	(GARCIA-MIRELES *et al.*, 2018)	• Quality attributes (e.g., maintainability, portability & usability) promote perdurability
	GD38	Avoid tools and methods that affect software engineers in their work negatively	(NAZIR *et al.*, 2020a)	• Avoid tools and methods that impacts negatively developers
	GD39	Provide sustainable software engineering tools and methods to software engineers	(NAZIR *et al.*, 2020a)	• Provide tools and methods to software engineers
	GD40	Use the existing tools for software development such as Jira, Axosoft and LeanKit.	(RASHID; KHAN, 2018)	• Support Tools

Table 25 – Text passages on SLR

Cat.	ID	Text Passages	Reference	Codes
Characteristic	CH1	A form of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable.	(KITCHENHAM; CHARTERS, 2007)	• Well defined methodology • Repeatability
	CH2	In this regard, the systematic review approach proved to be reasonably robust and could be considered as a reliable research method.	(MACDONELL *et al.*, 2010a)	• Reliability of SLR
	CH3	It is necessary to determine and follow a search strategy. This should be developed in consultation with librarians or others with relevant experience	(KITCHENHAM; CHARTERS, 2007)	• Determine and Follow a search strategy
	CH4	The search strategy must allow the completeness of the search to be assessed. In particular, researchers performing a systematic review must make every effort to identify and report research that does not support their preferred research hypothesis as well as identifying and reporting research that supports it.	(KITCHENHAM; CHARTERS, 2007)	• Assessability • Careful report]
	CH5	The well-defined methodology makes it less likely that the results of the literature are biased, although it does not protect against publication bias in the primary studies.	(KITCHENHAM; CHARTERS, 2007)	• Well defined methodology
	CH6	The stages listed above may appear to be sequential, but it is important to recognise that many of the stages involve iteration	(KITCHENHAM; CHARTERS, 2007)	• Iterative process
	CH7	Previous findings reinforce the cyclic characteristics of some phases or subsets of activities within a given phase in the SR process. For example, in Zhang and Muhammad's work, many interviewees mentioned the study selection and data extraction are done in an iterative way, specially when novices are involved in the SR. This may even lead to a revision of the inclusion and exclusion criteria and the data extraction form as well.	(FABBRI *et al.*, 2013)	• Iterative process
Communication and Collaboration	CC1	Academia and industry collaborations are important for better utilization of research results and for guiding future research directions	(BADAMPUDI *et al.*, 2019)	• Academia and industry collaboration
	CC2	All systematic review team members need to take an active part in developing the review protocol.	(BRERETON *et al.*, 2007)	• Team members collaboration
	CC3	Choosing practical problems and collaborating with industry are some of the possible solutions	(CARTAXO *et al.*, 2018)	• Industry collaborations
	CC4	Consult target audience	(AMPATZOGLOU *et al.*, 2020)	• Consult target audience
	CC5	Different approaches to provide knowledge to practice have been proposed in the literature are discussed	(BADAMPUDI *et al.*, 2019)	• Provide knowledge to practitioners
	CC6	Discuss among authors	(AMPATZOGLOU *et al.*, 2020)	• Author discussion

Continued on next page

Cat.	ID	Text Passages	Reference	Codes
		Table 25 – continued from previous page		
	CC7	Generalizability threats refer to the possibility of not being able to generalize the results of the secondary study (e.g., due to the identification of only a portion of existing primary studies). A special case of this threat that is quite frequently reported is results not applicable to other organizations or domains.	(AMPATZOGLOU *et al.*, 2020)	• Produce generalizable results]
	CC8	Import expert evaluation	(KITCHENHAM; CHARTERS, 2007)	• Expert support
	CC9	Include more than one researcher in the initial screening to minimize the risk of removing studies that should be included (false negatives)	(WOHLIN *et al.*, 2020)	• Include researchers from initial study
	CC10	Involve more than one researchers	(AMPATZOGLOU *et al.*, 2020)	• Involve more than one researcher
	CC11	More than two reviewers perform the data extraction independently and check the results of each review	(AMPATZOGLOU *et al.*, 2020)	• Envolve more thant two researchers for data extraction
	CC12	One of the anticipated benefits of systematic literature reviews (SLRs) is that they can be conducted in an auditable way to produce repeatable results.	(KITCHENHAM *et al.*, 2011)	• Auditable and repeatable results
	CC13	Open science is becoming more and more accepted in scientific communities to be having many positive effects. These effects range from increased access and citation counts to facilitating technology transfer with the industry and fostering collaborations through open repositories. Academic publishing and knowledge sharing is meant to become more cost-effective	(MENDEZ *et al.*, 2020)	• Facilitate collaboration with industry
	CC14	Regular and continuous collaboration among researchers and practitioners regarding research evidence/contribution leads to better use of research results in practice	(BADAMPUDI *et al.*, 2019)	• Continuous collaborations among researchers and practitioners
	CC15	Researchers are not always reporting potentially useful findings in a manner that makes them readily accessible to the wider community	(BUDGEN *et al.*, 2018a)	• Accessibility to wider community
	CC16	Review by independent expert	(AMPATZOGLOU *et al.*, 2020)	[Expert support]
	CC17	The importance of interaction between researchers and practitioners to integrate evidence from research results (knowledge) and practitioners' opinion and experience has been identified in the evidence-based software engineering guidelines.	(BADAMPUDI *et al.*, 2019)	• Interaction among researchers and practitioners
	CC18	The practitioners' opinions need to be considered jointly for the adoption of the existing alternatives	(BADAMPUDI *et al.*, 2019)	• Consider practitioner opinions
	CC19	There needs to be an agreed validation process separate from the protocol piloting activity. Ideally, external reviewers should undertake this validation process.	(BRERETON *et al.*, 2007)	• Validate the results with external reviewers
	CC20	Use expert's opinion	(AMPATZOGLOU *et al.*, 2020)	[Expert support]

Continued on next page

Table 25 – continued from previous page

Cat.	ID	Text Passages	Reference	Codes
	CC21	We argue that, whenever there are the resources available, the entire selection of studies should be made independently by, at least, two people, and then compared. We also argue that the best choice would be for these two people to be experienced in carrying out SLRs in SE	(WOHLIN *et al.*, 2020)	• Involve more than one researcher
	CC22	Whenever feasible, data extraction should be performed independently by two or more researchers. Data from the researchers must be compared and disagreements resolved either by consensus among researchers or arbitration by an additional independent researcher. Uncertainties about any primary sources for which agreement cannot be reached should be investigated as part of any sensitivity analyses. A separate form must be used to mark and correct errors or disagreements.	(KITCHENHAM; CHARTERS, 2007)	• Involve more than one researcher
Documentation	DC1	The search should be documented as it occurs and changes noted and justified.	(KITCHENHAM; CHARTERS, 2007)	• Rigorous and complete documentation
	DC2	All forms of searching together with the rationale for their use, and the numbers found for each form should be clearly reported, including nil returns. A recommended option is to include the numbers in the process diagram	(BUDGEN *et al.*, 2018a)	• Rigorous and complete documentation
	DC3	Continuously update the classification schema to be able to classify all primary studies	(AMPATZOGLOU *et al.*, 2020)	• Rigorous and complete documentation • Continuously Update documentation schema
	DC4	Develop protocol	(AMPATZOGLOU *et al.*, 2020)	• Develop Protocol
	DC5	Make data available	(AMPATZOGLOU *et al.*, 2020)	• Data availability
	DC6	Systematic reviews document their search strategy so that readers can access its rigour and completeness.	(KITCHENHAM; CHARTERS, 2007)	• Document search strategy to enable the access to rigour and completeness
	DC7	Systematic reviews must be careful to report their search process fully or they will not be repeatable. Missing papers can have a significant impact on the stability of the results of a secondary study	(KITCHENHAM *et al.*, 2011)	• Careful report • Repeatability
	DC8	The inclusion/exclusion rules should be presented as a distinct element, such as a table, so that they can be readily recognised and cross-referenced.	(BUDGEN *et al.*, 2018a)	• Present the inclusion and exclusion rules
	DC9	The unfiltered search results should be saved and retained for possible reanalysis.	(KITCHENHAM *et al.*, 2004a)	• Results should be retained for reanalysis

Continued on next page

Table 25 – continued from previous page				
Cat.	**ID**	**Text Passages**	**Reference**	**Codes**
	DC10	There are several areas where better reporting is needed, including quality assessment, synthesis, and the procedures followed by the reviewers. Researchers, practitioners, teachers and journal referees would all benefit from better reporting of systematic reviews, both for clarity and also for establishing the provenance of any findings.	(BUDGEN *et al.*, 2018a)	• Better reporting is needed
	DC11	To provide as much information as possible about the SLR being updated	(WOHLIN *et al.*, 2020)	• Rigorous and complete documentation
	DC12	One important aspect of RRs is the way they are reported. Many authors argue that alternative mediums should be used, instead of traditional research papers format [...] although RRs present several variations on their methods and terminologies, 78% of them present results as a narrative summary reported in mediums that better fit practitioners' needs. For example: Contextual Summaries that limits the report to a one-page document; the Briefings that summarize the main findings of a systematic review in one section; [...]	(CARTAXO *et al.*, 2018)	• Polymorphic report
Tools and Support Techniques	TL1	To reuse the protocol from the SLR being updated	(WOHLIN *et al.*, 2020)	• Reuse the protocol
	TL2	Current strategies that support the automation of SLR activities, so that researchers can produce SLRs in SE more efficiently and cost effectively	(FELIZARDO; CARVER, 2020)	• Automation of SLR activities • Improve efficiency and effectivity
	TL3	Automation tools need to be able to work together and exchange data so SLR authors can choose the most appropriate set of tools for their review.	(FELIZARDO; CARVER, 2020)	• Automation Tools
	TL4	Expect to revise your questions during protocol development, as your understanding of the problem increases. A pre-review mapping study may help in scoping research questions.	(BRERETON *et al.*, 2007)	• Pre-revise the protocol
	TL5	Further studies and their findings will broaden our understanding in this area. They may provide support for how to effectively and efficiently conduct SLR updates, which will become an even more critical area since SLRs may need to be updated after some time to capture new development and research in a field.	(WOHLIN *et al.*, 2020)	• Improve Efficiency and effectivity
	TL6	Piloting the research protocol is essential. It will find mistakes in your data collection and aggregation procedures. It may also indicate that you need to change the methodology you intend to use to address the research questions.	(BRERETON *et al.*, 2007)	• Pilot the protocol
	TL7	Project timetable. This should define the review schedule	(KITCHENHAM; CHARTERS, 2007)	• Define review schedule
	TL8	Reuse existing schemas	(AMPATZOGLOU *et al.*, 2020)	• Reuse existing schema

Continued on next page

Table 25 – continued from previous page

Cat.	ID	Text Passages	Reference	Codes
	TL9	Some of these lessons-learned are: (i) to adopt software tools to support the updating process;	(WOHLIN *et al.*, 2020)	• Adopt tools to support update
	TL10	Study selection criteria are used to determine which studies are included in, or excluded from, a systematic review. It is usually helpful to pilot the selection criteria on a subset of primary studies.	(KITCHENHAM; CHARTERS, 2007)	• To pilot the selection criteria on a subset of primary studies
	TL11	The development of automation tools has been slow and fragmented in SE. To fully reach the potential of automating SLRs, researchers will need a sustained coordinated collaborative effort. Automation tools need to be able to work together and exchange data so SLR authors can choose the most appropriate set of tools for their review.	(FELIZARDO; CARVER, 2020)	• Automation tools development
	TL12	The objective of this chapter is to present current strategies that support the automation of SLR activities, so that researchers can produce SLRs in SE more efficiently and cost effectively	(FELIZARDO; CARVER, 2020)	• Automate SLR activities
	TL13	The results, in our view, suggest that the use of a decision support mechanism (as the 3PDF) to help the SE community decide upon the need for SLR updates is very important. We also put forward that the 3PDF should be adopted by the SE community to keep relevant evidence up to date and to avoid wasting effort with unnecessary updates.	(MENDES *et al.*, 2020b)	• Avoid unnecessary updates
	TL14	Tools supporting the review process	(AMPATZOGLOU *et al.*, 2020)	• Use support tools
	TL15	Use Brainstorming among team members	(AMPATZOGLOU *et al.*, 2020)	• Brainstorm among team members
	TL16	We would like to see greater co-operation among the different efforts along with agreed-upon standards to enable "plug and play" standardization among different tools. We propose that it would be beneficial if we could combine the strengths of existing tools (Marshall et al. 2014, 2018) and integrate them into the activities of the SLR process. For this proposal to work, these tools will have to work together and be able to exchange data/results.	(FELIZARDO; CARVER, 2020)	•Support tools integration
	TL17	Current software engineering search engines are not designed to support systematic literature reviews. Unlike medical researchers, software engineering researchers need to perform resource-dependent searches.	(BRERETON *et al.*, 2007)	• Support from search engine
	TL18	In the past decade, researchers have made major advances in automating the SLR process, aiming to reduce the workload and effort for conducting high-quality SLRs in software engineering (SE).	(FELIZARDO; CARVER, 2020)	• Automation of SLR activities • Reduce efforts and workload

Continued on next page

Table 25 – continued from previous page				
Cat.	**ID**	**Text Passages**	**Reference**	**Codes**
	TL19	An assessment made by 'pooling' the independent results from two or three reviewers is likely to be more reliable than if only one reviewer has performed this task, with perhaps some checking by another author . As noted earlier, the use of multiple assessments is now recommended	(BUDGEN *et al.*, 2018a)	• Use two or more reviewers
	TL20	The majority of papers only presented preliminary investigations, often describing an example of the tool in use, or a small experiment to assess its effectiveness. In addition, only two studies reported that an independent evaluation of a tool had been carried out. These results reflect the immaturity of the research area; however, they do lay the foundations of future work.	(MARSHALL; BRERETON, 2015)	• Immaturity of technology
	TL21	Based upon the evidence gathered, most of the tools identified are in early stages of development and usage. This has led to very little primary data regarding their effectiveness, and generally only speculation over their potential. An empirical investigation to assess the effectiveness of SLR tools could be a beneficial contribution to the topic.	(MARSHALL; BRERETON, 2015)	• Usage of technology

Table 26 – Translations to the characteristics for sustainable SLR

ID	Software Systems Passages	SLR Passages	Translation
	1st Order	1st Order	2nd Order
TR1	Green SE aims to create **reliable and durable** software that **meets users' needs** while reducing environmental impacts [DF4]	**Reliable and durable:** • SLR is a robust and reliable search method [CH2] • SLR creates auditable and repeatable results [CC12] **Meets to user needs:** • Researchers must consult target audience [CC4] • Researchers must consider practitioner opinions [CC18] • Validation of the results with external reviewers [CC19] • Researchers should choose practical problems and collaborating with industry are some of the possible solutions [CC3]	A sustainable SLR aims to **create reliable results** i.e., the threats to the validity of a sustainable SLR must be carefully managed, mitigated, and described so that the stakeholder can judge the quality of the reported results. Furthermore, SLR must **meets to stakeholders needs** i.e., **stakeholder needs must be translated into research questions to be answered by SLR**.

Continues on the next page...

Table 26 – continued from previous page			
ID	**Software Systems Passages**	**SLR Passages**	Translation
	1st Order	**1st Order**	**2nd Order**
TR2	The term sustainable software can be interpreted in two ways: (i) **the software code being sustainable, agnostic of purpose**[1], or (ii) the software purpose being to support sustainability goals, i.e. improving the sustainability of human kind on our planet. [DF11]	**The software code being sustainable, agnostic of purpose:** • Creation of results that are generalizable for other contexts (organizations or domains) [CC7] • Production of results that are interesting for academia and industry by encouraging collaboration and creating ways for stakeholders participate actively in the process [CC1] [CC2] [CC3] [CC4] [CC14] [CC16] [CC17] [CC18] [CC20]	Sustainable SLR produces results that are agnostic of purpose, i.e., **SLR provides results that are useful to a wider community**. Sustainable SLR must consider the applicability and generalizability of results prioritizing that it can be reasonably applied to a definable group of researchers and/or practitioners and/or to the wider community.
TR3	Green or Sustainable Software is a software that **follow software development process** where practical application of sustainability aspects takes place. For instance, software that complies with customer software requirements **causing less changes during it scope**. [DF7]	**Follow a development process:** • SLR have a well-defined methodology [CH1] [CH5] • SLR is considered a robust and reliable research method [CH2] • Researchers must develop a protocol [DC4] and create a complete documentation [DC2] **Causes less changes during its scope:** • Pilot the protocol to avoid unnecessary changes in during studies selection, data extraction, or data synthesis [TL6] • Usage of iterative process to reduce the impact of changes in protocol [CH6]	A sustainable SLR **complies with conduction standards**, and the quality characteristics that are expected from any SLR developed by professionals. In addition, sustainable SLR must prioritize **to cause less changes in protocol during the conduction/reporting** i.e., it maintains the iterative process concentrating the main changes until the pilot test is conducted.
TR4	Sustainable software avoids non-essential deliverable during software development; Also is a project that avoid extensive documentations. Is a software that have a simple and flexible design and **fits to meet the long-term goals of project**. Is a software **that requires less effort for development and consumes less energy**. [DF6]	**Fits to meet to long-term goals of project:** • Academia and industry collaborations are important for better utilization of research results and for guiding future research directions [CC1] **Require less effort for development and consumes less energy:** • Researchers made major advances to reduce efforts and workload using automation of SLR activities[TL18]. • Avoid unnecessary updates and consequently effort waste [TL13]. • Pilot the protocol [TL4][TL6][TL10] and use an iterative process [CH6] avoid wasting efforts with mistakes in protocol/conduction.	Sustainable SLR should have **long-term goals** aiming at impacting the research area and the community around including researchers and SE practitioners over a long period of time i.e., and not only satisfy current researchers' needs. Also, sustainable SLR must also preserve the resources and create methods to **reduce the time and effort needed to conduct/update SLR**.
			Continues on the next page...

[1] **Agnostic of purpose** refers to code that are highly reusable independently of its context, for example, data structures like linked lists or binary trees.

Table 26 – continued from previous page			
ID	**Software Systems Passages**	**SLR Passages**	Translation
	1st Order	**1st Order**	**2nd Order**
TR5	Sustainable software is one that satisfy the that satisfy the requirements of the three dimensions of **society**, economy, environment with **responsible use of ecological, human, and financial resources**. In addition, a fourth dimension "human sustainability" should be also included. Most cited definition for sustainable software is one that **"meet the needs of the present without compromising the ability of future generations to satisfy their own needs"**. [DF3]	**Responsible use of financial resources:** • Researchers must avoid start SLR from scratch and reuse elements from other SLR. [TL1] [TL8] • Cost can be minimized through automation of SLR activities to improve efficiency and effectivity [TL2] **"Meet the needs of the present without compromising the ability of future generations to satisfy their own needs":** • Researchers must consult target audience [CC4] and consider practitioner opinions [CC18] to choose practical problems and collaborating with industry [CC3] to meet present needs. • Create results that are accessible to wider community [CC15] avoiding that SLR information fulfil only researchers particular needs.	A sustainable SLR should **make a responsible use of financial resources**, for this it must reuse elements from previous SLR and estimate the manual effort involved in SLR (e.g. the time of academics and specialists or practitioners in the SLR domain) always searching for alternatives to switch from manual labor to automated or semi-automated tasks. In addition, it must ensure that researchers manage human aspects of SLR, making decisions that reduce the negative impact on stakeholders. Furthermore, it must provide **information for a broad community**, including academy and industry, and **not just fulfilling their particular needs or from specific research groups**.
TR6	Green software is a software that **generates as tiny e-waste**[2] as possible during its operation and development. [DF2]	**Generates as tiny e-waste:** • Avoid waste information and effors reusing the existing schema generated in previous SLR [TL8]. • Reuse the protocol information to avoid start SLR from scratch [TL1].	A sustainable SLR should **avoid research waste** by publishing results and outputs that are useful for its target audience; and, allowing that its components (protocol elements) and other artifacts generated during its conduction be reused to avoid waste of information and effort.
			Continues on the next page...

2 **E-waste** is defined as the waste of energy or production of harmful emissions (e.g., printers, photocopier or other peripherals) that originates during the development and operation of the system (ERDELYI, 2013)

Table 26 – continued from previous page			
ID	**Software Systems Passages**	**SLR Passages**	Translation
	1st Order	**1st Order**	**2nd Order**
TR7	Sustainable Software is software whose direct and indirect **negative impacts on economy, society, human beings**, and the environment **resulting from development, deployment, and usage of the software is minimal and/or has a positive effect on sustainable development**[DF1].	**Negative impacts on economy, society, human beings, and environment resulting from development, deployment and usage is minimal:** • Researchers must adopt solutions to automate SLR activities and improve efficiency [TL2] reducing excessive efforts and workload [TL16] minimizing the amount of resources consumed to conduct/update, consequently, negative impacts in SLR process. • A better report for SLR is needed including quality assessment, synthesis, and the procedures followed by the reviewers [DC10] or they will not be repeatable[DC7], hence, a detailed report can minimize negative impacts in future updates/re-conduction. • Researchers should choose practical problems and collaborating with industry [CC3] reporting all potentially useful findings in a manner that makes them readily accessible to the wider community [CC15] minimizing negative impacts on SLR usage. **Has a positive effect on sustainable development:** • Since SLR follow the guidelines, it can provide reliable evidence [CH2] for academy and practitioners [CC5] and impact positively the academy (research) and industry (software development [CC1] providing support for decision-making process.	A sustainable SLR aims to identify, manage and mitigate threats to validity and provide means for researchers to make decisions that **reduce the negative impact on other researchers making them minimal**. For this, SLR must be compromised in ensuring the credibility of the results, generating a careful report to ensure that efforts needed to replicate or update SLR is minimal. **Sustainable SLR should have a positive effect in research development** minimizing the cost and maximizing results usability.

Continues on the next page...

Table 26 – continued from previous page			
ID	**Software Systems Passages**	**SLR Passages**	Translation
	1^st^ Order	**1^st^ Order**	**2^nd^ Order**
TR8	Green software is a software product that **over its whole life cycle are continuously assessable, documented and maintainable** [DF13], and **used for further optimization** of the software product [DF8].	**Over its whole life cycle are continuously assessable, documented and maintainable:** • Unfiltered search results should be saved and retained for possible reanalysis.[DC9] and researchers should make data available [DC9]. • SLR must contain a rigorous and complete documentation [DC1][DC2][DC3] that support auditability [CC12]. **Used for further optimization:** • SLR must enable to researchers repeat [CH1, CC12, DC7] and continuously update documentation schema [DC3] i.e., current SLR results and artifacts (like protocol [TL1]) must be used to optimize further SLR conduction	A sustainable SLR is a review that over its whole life-cycle must be **continuously assessed and continuously documented** i.e., it maintains a complete documentation and keep all relevant data available for possible audition/reanalysis, consequently, making it maintainable. In addition, all **elements from previous SLR (including implicit knowledge) should be reused in further update/replication of SLR**.
TR9	A sustainable software is a system that **manage resources used**[DF12]. In addition, sustainable software is aware of time behavior and **Resource utilization**. Other characteristics from software product quality that can be adapted to sustainability are: **learnability**, **accessibility**, **reusability**, **modifiability**, **adaptability**. Sustainability also must address usefulness, freedom from risk, support environmental risk mitigation, context coverage, context completeness, flexibility[...][DF14]	**Manage resources used:** • SLR researchers are aware of the necessity of reducing the workload i.e., efforts (time and effort) taken while conducting high quality SLR [TL18]. For better managing these resources the community proposed to automate SLR activities [TL2] [TL12] [TL18], avoid unnecessary updates [TL13] and creating a careful reports [CH4] **Learnable and accessible:** • Results of SLR must be accessible to wider community (academy or industry) [CC15] • Presents the results in different formats for better fitting in practitioners needs [DC12] **Reusable, modifiable and adaptable:** • SLR must contain a a rigorous and complete documentation [DC1][DC2][DC3] which allows reusability components like the protocol [TL1]. • SLR results should be retained to enable researchers to perform reanalysis (including modifications/adaptations) [DC9] or repeat SLR conduction [CC12].	A sustainable SLR **manages resources used** in its life-cycle. It also must create a documentation that is detailed, **easily understandable, auditable, and accessible to others (researchers, practitioners, and others)**. Furthermore, documentation should include all relevant data organized to be **easily reused, modified, and adapted** by other researchers.

Continues on the next page...

Table 26 – continued from previous page			
ID	Software Systems Passages	SLR Passages	Translation
	1st Order	1st Order	2nd Order
TR10	A sustainable software is aims to **maintain its viability during a maintenance period or replacement by a new system** (with continuous monitoring of quality, and knowledge management). In addition, sustainability can also be considered in software usage, whose processes in the application domain are triggered by the software system as product. [DF9]	**Maintain its viability during a maintenance period or replacement by a new system:** • Researchers must avoid unnecessary efforts to re-conduct or update SLR [TL13] and create results that are auditable [DC9] and repeatable [CC12].	A sustainable SLR keep information organized and accessible **to maintain its viability in update/reconduction**.

Table 27 – Translations to the critical factors for sustainable SLR

Cat.	ID	Software systems Passages	SLR Passages	Translation
		1st Order	1st Order	2nd Order
Communication and Collaboration	TR11	Rich communication and collaboration [F1]	SLR process considers essential the *communication* among researchers, in addition, when possible it should involve more than one researcher (preferably by two experienced people) [CC10] to execute critical tasks e.g., selection of studies, data extraction and interpretation [CC21]. Researchers should *collaborate* to solve disagreements [CC13] and minimize the risk of removing studies that should be included [CC9]. Furthermore, all stakeholders should *communicate and collaborate* to ensure that team members have active participation in SLR process, especially in tasks like developing/reviewing the protocol [CC2] [CC19]	**Rich communication:** involves the communication among all **SLR stakeholders** (including researchers and SE practitioners) over the whole SLR process, for instance, to elaborate/revise the protocol, perform data analysis using consensus meetings to solve disagreements, or revise the final report to ensure understandability and completeness of documentation.
	TR12	Lack of customer's presence [F10]	Academy and industry can be considered as primary users of SLR results, hence, the effective participation of both is important to better utilize research results and for guiding future research directions [CC1]. This collaboration should be continuous [CC14] (i.e., occur over the entire SLR process) and the *presence of stakeholders* is essential to integrate evidence from research results (knowledge) and practitioners' opinion and experience. [CC17]	**SLR stakeholders (researchers and SE professionals) effective participation in SLR process:** means that stakeholders' opinions must be used to adjust project planning items (e.g., aims, scope, restrictions, etc.) and to validate the findings aiming to uptake SLR evidence into practice improving the usage of research findings into real scenarios.

Table 27 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR13	Insufficient knowledge of the customer [F11]	Integrate evidence from research results (knowledge) and practitioners' opinions and experience are guidelines proposed in evidence-based software engineering [CC17], hence, to develop an SLR it is necessary to consult target audience, [CC4], import expert evaluation, [CC8]. Therefore, much of the validity of results relies on previous knowledge of stakeholders, consequently, the *lack of knowledge of the stakeholders* about the research domain can directly impact SLR results.	**Knowledge of stakeholders about the research domain**: knowledge about previous solutions proposed, current research gaps, important open issues, trends from industry/academy, and other aspects that could guide researchers to address important research topics and provide useful evidence to SE community
	TR14	Lack of experience[F14]	*Experience from stakeholders* (researchers and SE professionals) in SLR conduction process is essential to avoid common pitfalls that may hinder the process and impact negatively results validity, for instance, electronic databases inefficiencies (KITCHENHAM; BRERETON, 2013), problems coverage due to the lack of standardization of keywords (AMPATZOGLOU *et al.*, 2019), etc. Hence, it is recommended to always use experts opinions [CC8] [CC10] [CC16] [CC20] and create systematic ways to double check SLR steps (e.g., using more than a single reviewer to compare decisions and interpretations [CC21] or consult experienced researchers to review the search strategy [CH3]).	**Experience of team members and stakeholders in SLR conduction:** experience includes knowledge about scientific research process mainly regarding the best practices to conduct/update SLR, consequently, the experience can support researchers to minimize bias, improve transparency, replicability, overall quality and avoid well-known pitfalls in process.
Knowledge management	TR15	Software reusability [F5]	SLR should be conducted in an auditable way to produce repeatable results [CC12]. Being repeatable means that documentation contains all details needed to replicate the results and data is openly available. The fact of SLR ensures a repeatable process also should allow researchers to reuse elements from the previous SLR [TL1] (e.g., search string, selection criteria, quality assessment, etc.) and reuse totally/partially the existing schemas [TL8] (e.g., data extraction schema).	**Improvement of SLR reusability:** refers to the careful documentation that allows researchers to reproduce/replicate SLR results and *reuse any components of the previous SLR* (e.g., protocol elements like search string, inclusion/exclusion criteria) or adapt research efforts like studies selection or data extraction to create new evidence.

Table 27 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR16	Usage of refactoring techniques [F6]	Refactoring techniques can be applied when: (i) researchers need to reuse SLR protocol elements [TL1]) and refactor inconsistencies or inefficiencies of previous SLR (e.g., refining search strings removing unnecessary keywords (NAPOLEÃO *et al.*, 2021)); (ii) when researchers need to reanalyze results from previous SLR [DC1] [DC9] to use them as a body of knowledge for their research; (iii) researchers need to continuously update and refactor SLR elements (e.g., classification schema) among iterations [DC3] to enable that all studies being correctly classified.	**Usage of refactoring techniques:** *refactoring* refers to using different ways to conduct SLR steps prioritizing the reuse of elements and avoiding starting from scratch. A *refactoring technique* refers to a systematic way to reuse elements from the previous SLR (e.g., search strategy, set of studies selected, quality assessment criteria, data extracted) that ensure the reliability of this process and allows that former data to be used as a baseline to answer new research questions.
	TR17	Efficient Knowledge sharing [F15]	Knowledge sharing/transfer mainly occurs in three moments: (i) when the research team maintains regular and continuous collaboration using brainstorming sessions [TL15] to solve conflicts, share their experience with team members and ensure that evidence/contribution are useful in practice [CC14]. (ii) When researchers provide open access to manuscripts, data collected and materials (e.g., scripts) to analyze data (MENDEZ *et al.*, 2020); and, (iii) when researchers report carefully the whole SLR process [DC7] externalizing details and any deviations from protocol that will be helpful for future researchers.	**Efficient sharing and knowledge transfer:** involves sharing knowledge in three different moments: (i) when stakeholders share their experience during consensus meetings; (ii) when researchers provide open access to data extracted (including raw data) and report carefully all decisions taken in SLR process; and (iii) when authors report lessons learned, their insights about the applicability of results in practice and provide recommendations for SE professionals.
Technical support	TR18	Lack of methodologies and tool support [F13]	Tool support is considered essential for SLR process to aid researchers to alleviate the amount of time- and effort to conduct SLR and current barriers especially in search phase(AL-ZUBIDY; CARVER, 2018). Despite researchers have proposed tools try to address some of the SLR barriers, the SLR process still suffers from a number of barriers and still lacks adequate tool support (AL-ZUBIDY; CARVER, 2018). Thus, the availability of these tools and their adoption is essential to support SLR process [TL9].	**Usage of tools to support SLR**: involves (i) tools to support the whole SLR process (StArt(FABBRI *et al.*, 2016), Parsifal[3], SLuRP(BOWES *et al.*, 2012)); (ii) tools to support specific tasks and save efforts by avoiding manual work (e.g., reference managers, text editors, spreadsheets); and (iii) automation or semi-automation tools (e.g., search string calibration, studies selection review).

[3] <https://parsif.al/>

Table 27 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
Technical support	TR19	Technology Maturity [F16]	Tools proposed for SLR over the years only presented preliminary investigations, often describing an example of the tool in use, or a small experiment to assess its effectiveness. Thus, there is a lack of studies reporting independent evaluation or intensive testing which reflects the *immaturity* of the research area [TL20][TL21]. Despite some advances in this area proving that more recent tools are covers more of the essential requirements than in the past (AL-ZUBIDY; CARVER, 2018) there is still a need for a joint effort among researchers and tool developers to build a comprehensive, free tool suite to support SLR (AL-ZUBIDY; CARVER, 2018)[TL16].	**Maturity of support technology:** maturity refers to producing support tools that are more than prototypes or proof of concept tools. Mature tools should be intensively tested and assessed in a wide range of scenarios (e.g., different domains, using qualitative and quantitative data, using heterogeneous data sources, etc.) aiming to deliver a mature product with (i) long-time support, (ii) solutions for complex SLR tasks (e.g., data summarization), and (iii) integration with other tools.
	TR20	Technology Accessibility [F17]	*Accessibility of SLR tools* refers to the creation of tools that minimizes the user learning curve. Marshall et al. (MARSHALL *et al.*, 2018) appraised SLR tools and revealed that some them have a long learning curve i.e. are confusing and difficult to learn. Accessibility can be extended to third-party tools e.g., digital libraries front-ends (i.e., interfaces, displaying results, guidelines, and filters) and back-ends (i.e., syntax, algorithms, rules, and optimization), because it impacts directly on researcher accessibility to studies (AL-ZUBIDY; CARVER, 2018). Hence, it is essential to have a better cooperation and create tools that are easy to use and create standards that allow them to be "plug and play"[TL16]	**Accessibility of support technology:** accessibility refers to developing tools that are easy to use (including an intuitive user interface/experience, comprehensive documentation/tutorials) and preferably being open-source and supported by the community.

Table 27 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
Resources Usage	TR21	Develop and generate methods to consistently minimize their utilization of regular power and resources [F2]	During SLR conduction/update, some tasks are prone to consume more time- and effort (e.g., searching in databases, studies selection, data extraction, and assessing quality) (CARVER *et al.*, 2013). Hence, it is important to develop and use approaches that minimize the amount of time and effort needed to conduct SLR [TL2] (e.g., visual text mining(FELIZARDO *et al.*, 2017b) or machine learning to select studies (WATANABE *et al.*, 2020), or usage of snowballing for updating SLR (WOHLIN, 2016)). At the same time, these approaches move towards the reduction of workload while creating high-quality SLR [TL18] .	**Usage of techniques that minimize the resources consumption:** prioritize the use of techniques that requires less effort to produce high-quality output (e.g., usage of snowballing to update SLR). While conducting/updating researchers should design SLR to comprise techniques that automate time-consuming tasks (e.g., using techniques like machine learning).
	TR22	Efficient utilization of time and computing resources [F4] [F9]	SLR conduction consumes three main resources: time, personnel, and cost (NUSSBAUMER-STREIT *et al.*, 2021) whose "time" is the most reported resource and refers to the amount of time to execute SLR tasks. Efficiently managing these resources means (i) reducing barriers that impact negatively resource consumption (e.g., Lack of domain expertise, noisy surroundings, resource unavailability, poor internet, software incompatibilities, and limitations (NUSSBAUMER-STREIT *et al.*, 2021); (ii) using techniques that reduce workload by automating activities [TL2] [TL12] [TL18]); or (iii) finding alternative ways to conduct SLR tasks that consumes less time and effort (e.g., applying snowballing to update instead of traditional database search (WOHLIN, 2016))	**Efficient management/usage of resources:** "resources" refers to (i) physical resources (computers; Internet access; access to bibliographic bases, office supplies; and space or tools to manage virtual meetings) (ii) intellectual/human resources (how many and which stakeholders will compose the review team). " Efficient management" means providing researchers the necessary environment and resources needed (physical or intellectual) to execute tasks always designating tasks for team members according to their capacity.

Table 27 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR23	Run feasibility studies that assess the impact of the project on the environment [F7]	Pilot testing SLR protocol before advancing to conduction phase is essential to minimize negative impacts of problems in search strategy (KITCHENHAM; CHARTERS, 2007). Hence, it is essential to run pilot tests that help researchers to adjust questions as the understanding of this problems increases [TL4]. Furthermore, pilot tests support to find mistakes in your data collection and aggregation procedures or indicate changes in the methodology you intend to use [TL6] [TL10].	**Usage of feasibility studies :** "feasibility studies" refers to a more complex version of "pilot studies" that should include other aspects (beyond the protocol) to appraise the viability of conducting SLR. Feasibility studies should comprise a deeper investigation of (i) the existence of the previous SLR conducted in the same domain and the possibilities of reusing this data; (ii) the expected effects of results on the practitioner's decision-making process; (iii) the capability of SLR of identifying gaps and trends in the current evidence to underpin future research in the domain; (iv) the availability of resources to support SLR from cradle-to-grave.
	TR24	Sustainable maintenance of the software [F8]	Maintenance of SLR refers to the update process (NEPOMUCENO; SOARES, 2018) that should be conducted whenever there is sufficient evidence and changes in state of the art in a given domain (MENDES *et al.*, 2020b). To perform SLR maintenance, researchers should make use of the auditability of SLR[CC12] and reanalyze data and protocol elements [DC9] aiming to reuse elements and minimize efforts. Furthermore, it is recommended to adopt alternative strategies to update these studies (e.g., snowballing (WOHLIN, 2016) or visual text mining (GARCÉS *et al.*, 2017))	**Maintenance of SLR:** before updating/reconducting SLR, researchers should check the relevancy of the topic and evaluate the impact of providing up-to-date information for readers. During the SLR update, researchers should reuse as much information as possible (such as protocol, data, etc) and minimize efforts to update SLR.
	TR25	Iterative development [F12]	SLR process appears to be sequential, but it involves iterations [CH6]. Iterations during SLR conduction sometimes are not explicitly documented in the final report, however, cyclic characteristics of some phases or subsets of activities in SLR are very important to fit SLR results as the knowledge of participants grows. Iterations are usually described in protocol revision (after pilot testing) and sensitive activities like data extraction, especially when novices are involved in the SR. [CH7]	**Usage of iterative process to conduct SLR:** use *iterations* to overcome difficulties faced during SLR conduction (e.g., the inexperience of the research team, or lack of knowledge about the domain). Iterations lead the research team to revise multiple times protocols items before undertaking the full review, in addition, the iterative processes aid researchers to identify new important studies in each iteration and collect relevant information as knowledge of the research team increase.

Table 28 – Translations to the guidelines for sustainable SLR

Cat.	ID	Software Systems Passages	SLR Passages	Translation
		1st Order	1st Order	2nd Order
Communication and Collaboration	TR26	Schedule daily stand-up meeting to encourage the communication of team members. [GD1] Arrange face to face meetings and use web technologies such as online chats, webinars, email, and discussion forums for strong interactions between customer and developers. [GD2] Arrange social activities for providing opportunities to the team members for informal communication. [GD3] Plan interaction sessions among developers and customers to understand the complexity of software project. [GD4] Encourage informal communication across the team members. [GD5] Perform regular team sessions for effective software development plans (schedule, iteration, quality, and sustainability). [GD7]	SLR conduction should foster authors discussion [CC6] by adopting brainstorm sessions among team members [TL15] and other techniques or technologies that support information exchange. In addition it should involve more than one researcher [CC10][CC21] to execute tasks that are often subjective (e.g., data extraction). It is essential the continuous collaborations among researchers and practitioners [CC14] to facilitate that all stakeholder are aligned with research domain and report results that are more useful;	Schedule regular meetings to encourage communication/collaboration to brainstorm with researchers and stakeholders through web technologies (e.g., Slack, web conferences, etc.) or face-to-face meetings, to better comprehend the research domain and share solutions to improve the SLR process.
	TR27	Always code the software in pairs (pair programming). [GD27]	Involve more than two researchers for data extraction [CC10][CC21]	Always use two or more reviewers to double check the results of the conduction process
	TR28	Feedback from customers during retrospective meeting should be recorded to improve the overall quality of software. [GD6]	SLR process should consult target audience [CC4] i.e., academia and industry whose both must collaborate to produce useful results [CC1] [CC2] [CC3] [CC4] [CC14] [CC16] [CC17] [CC18] [CC19] [CC20]	Feedback from target audience during SLR conduction should improve the overall quality of SLR
	TR29	Design simple and user friendly software according to customer's needs. [GD28]; Update the design as per the changes in customer's requirements. [GD29]	During SLR conduction research group should consult target audience [CC4] for designing research to comprise with target audience needs and foster their utilization for guiding future research directions [CC1]. Hence, practitioners' opinions need to be considered [CC18] always piloting the protocol [CC19] and performing changes whenever necessary.	SLR design must comprise all stakeholder's requirements and provide evidence that is user-friendly that meets users' needs.

Table 28 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR30	Manage to redesign software system according to the volatile customer's requirements. [GD30] Follow short development cycles/iterations [GD31] Involve the customer throughout software development to validate the work in iteration. [GD32]	While conducting an SLR researchers should always consult target audience [CC4] to design research that serve as a guide for future research [CC1] always choosing practical problems to comprise the target audience needs [CC3]. Furthermore, an iterative process can be adopted [CH6], especially in protocol validation, always using stakeholders opinions to establish an consensus regarding the protocol [CC2];	Follow short conduction/update cycles/iterations managing the SLR redesign along the process according to the stakeholders requirements and using their opinion to validate the results in each iteration
	TR31	Identify sustainability stakeholders who issues objectives, constraints and consideration about the system under development. [GD25]; Use acceptance testing to validate the requirements of customer. [GD33]	Always use expert opinion [CC16][CC20] and consult target audience [CC4][CC18] to gather multiple perspectives of SLR process. Furthermore, it is important to conduct an agreed validation process separate from the protocol piloting activity. Ideally, external reviewers should undertake this validation process [CC19].	Consider multiple perspectives while conducting SLR identifying with stakeholders issues, objectives, constraints, considerations, and apply acceptance tests to validate requirements with stakeholders and target audience.
Knowledge management	TR32	Effective knowledge management [GD8]; Avoid knowledge loss by effective knowledge sharing [GD16];Provide knowledge sharing strategies [GD13]; Provide effective knowledge sharing platforms, for example, social media [GD18]	Knowledge management is essential to transfer the know-how of SLR process and make it easier for further updates (FELIZARDO *et al.*, 2020a). A better managing SLR information includes creating a detailed report (documentation)[DC7][DC10][DC1][DC2][DC3] and making data available [DC5], for example, through open science practices. Nevertheless, knowledge management is not only restricted to reporting and, it also extends to discussions among authors (including practitioners) discussion to share opinions and experiences [CC6][CC18]	Enhance knowledge management by using sharing platforms and effective strategies to avoid knowledge loss and improve sharing among team members.

Table 28 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR33	Create knowledge base [GD14]	Scientific knowledge storage currently relies on data availability [DC5] and rigorous and complete documentation [DC7][DC10][DC1][DC2][DC3]. Nevertheless, it is necessary to strengthen the creation of approaches to improve knowledge sharing (e.g., Open Science Framework (FOSTER; DEARDORFF, 2017) or PROSPERO (SCHIAVO, 2019)) making it possible to learn from previous experiences and share knowledge with other researchers possibly interested in the topic.	Create a knowledge base to improve knowledge management and allow SLR stakeholders to learn from previous experiences of the research group.
	TR34	Acquire knowledge by hiring skillful and knowledgeable software engineers [GD15]; Support from senior managers. [GD11]	Development of SLR should be in consultation with librarians or others with relevant experience [CH3], Hence, it is essential use expert opinion [CC8][CC16][CC20] to incorporate its knowledge into results and to validate sensitive parts of the process.	Include in the research team or get support from skillful and experienced researchers/stakeholders to acquire knowledge about the research domain and SLR process.
Technical support	TR35	Provide sustainable software engineering tools and methods to software engineers [GD39]	The need for tools to support SLR process is claimed by SE community by several opinion papers (KITCHENHAM; BRERETON, 2013) and ratified by other researchers that states the importance of using support tools [TL14]. However, most of the tools for SLR are prototype and are in the early stages of development [TL20][TL21], hence, a better tool support is needed (AL-ZUBIDY; CARVER, 2018).	Provide sustainable SLR tools and methods to researchers and practitioners
	TR36	Avoid tools and methods that affect software engineers in their work negatively [GD38]	Since tools proposed to support SLR process are mostly limited and there are only small experiments to assess their effectivity [TL20] researchers should avoid those that impacts negatively their work. SLR authors still report a number of limitations with the existing tool options which leads authors to perform a lot of manual work. For instance, removal of false positives, manual cleaning of references filtering the retrieved results. Most tools are general-purpose tools and not well-adapted to SLR search. Even so, many of these tools were still helpful and works reasonably well (AL-ZUBIDY; CARVER, 2018).	Avoid tools and methods that affect researchers and practitioners in their work negatively

Table 28 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR37	Demonstrate the current tools and methodologies applicability to software engineers. [GD12]	SE researchers surely would benefit from tools to manage the SLR process, but existing tools need independent validation i.e., the proposed tools need to be evaluated by groups other than those who developed them before they can be unreservedly recommended. (KITCHENHAM; BRERETON, 2013). Hence, it would be beneficial to demonstrate the effectiveness of these tools empirically [TL21].	Demonstrate the current tools and methodologies applicability to researchers and practitioners
Resources Usage	TR38	Develop reusable prototype instead of throw-away prototype. [GD20]	It is essential to develop artifacts that are reusable, for example, the protocol [TL1] that could be reused in update or as a baseline for other SLR. Furthermore, during the whole SLR process is important to reuse existing schemas [TL8], including the data generated and collected in pilot testing.	Develop pilot tests for SLR that are reusable and allow researchers to build them incrementally, avoiding throwing away the information and wasting resources reconducting unnecessarily any SLR step.
	TR39	Refactor the software system whenever required [GD21]; Adopt component-based development strategy for reuse of the existing components in the development of a new system to minimize time, cost, and efforts [GD22]; Highly reusable assets are prone to be optimized, as is their greenability [GD23]; If an asset is easy to modify, it is likely to keep (and not worsen) its greenability [GD24].	Artifacts generated from SLR process, including the protocol [TL1], results [DC9] and other existing schemas [TL8] should be stored for future reanalysis or updates. Thus, rigorous and complete documentation [DC1][DC2][DC3] and storing assets appropriately is essential to foster further reuse.	Adopt a component-based development strategy to maximize the reuse of existing schemas/assets and minimize time, cost, and efforts of the research team whenever a refactor is needed.
	TR40	Sustainability is more than only maintainability of software. A sustainable product should have a lower impact when introducing changes (reusability) or should improve usability that can make the product life-cycle longer [GD34]	SLR should be reusable during its whole life cycle, it includes produce a reusable protocol [CC1] and most of schemas [TL8], always avoiding start from scratch.	Conduct SLR that has a lower impact when introducing changes and improve the usability of results making its life-cycle longer

Table 28 – continued from previous page				
Cat.	**ID**	**Software Systems Passages**	**SLR Passages**	**Translation**
	TR41	Reliability has a potentially positive interaction with sustainability [GD35]; Quality attributes such as maintainability, portability and usability promote software perdurability [GD37]	Reliability of results is considered a strong characteristic of SLR method [CH2] that is mainly reinforced by the auditability of the results [DC9] and the rigorous and detailed documentation [DC1][DC2][DC3]. These characteristics allow researchers to trust in results to guide their research and create updateable results which make SLR more perdurable.	Enhance the quality of SLR concerning maintainability, usability, and reliability of results positively interacts with sustainability and promotes SLR perdurability.

APPENDIX

D

A Maturity Model for Sustainable Systematic Literature Reviews

SANTOS, V.; CAPILLA, R.; NAKAGAWA, E.Y.; A Maturity Model for Systematic Literature Reviews, ARXIV (pre-print), Pages. 1-8, 2022. DOI: 10.48550/arXiv.2206.11936

Abstract

Systematic literature reviews (SLR) have been widely adopted to synthesize evidence in a reliable and unbiased manner. The process of conducting an SLR is rigorous and well-known but some SLR do not provide high-quality results because some of the steps are not carried out while others may lead to inaccurate results. In other cases, the reproducibility of the review is hard to achieve. These inherent problems in some SLR demand to identify the weakest parts of the process and those that may introduce bias. Today, several techniques (e.g., snowballing, selection of the right studies, good search strings, etc) can be used to produce and systematize better the production of SLR. The main goal of this work is to identify *"how to improve the maturity of the SLR process"* and hence, identify the most common problems SLR suffer in terms of quality. To do so, we suggest a mapping between the SLR steps and the well-known CMMI (Capability Maturity Model Integration) model to identify the most common SLR practices and which of them are more critical if they are not enacted properly. Our results identify those common steps, and we came up with a mapping around nine goals and 46 practice areas that may require attention aiming to produce high-quality SLR.

D.1 Introduction

The software engineering (SE) community has adopted Systematic Literature Review (SLR) over the last years (MENDES *et al.*, 2020a) as a technique to summarize evidence from primary studies and support researchers in outlining the state of the art of a given research topic (KITCHENHAM *et al.*, 2015). SLR provides important benefits, such as the possibility of dealing with information from different studies in an unbiased manner (KITCHENHAM; CHARTERS, 2007; NIAZI, 2015), production of auditable and repeatable results (KITCHENHAM *et al.*, 2011; BUDGEN *et al.*, 2018a), and the opportunity to identify research gaps and also perspectives for future investigations (KITCHENHAM *et al.*, 2015).

At the same time, although a SLR defines a systematic process to be followed, some SLRs lack of enough rigor and this may lead to several problems, such as: poor documentation (ZHOU *et al.*, 2016; BUDGEN *et al.*, 2018a), lack of repeatability (KITCHENHAM *et al.*, 2011), bias in studies selection, data extraction, summarization (AMPATZOGLOU *et al.*, 2019), poor coverage and lack of generalization of the results (AMPATZOGLOU *et al.*, 2019). Other problems refer to unintentional bias induced by SLR author, so for instance, non-expert authors apply the SLR steps in a wrong way, or very expert authors believe some studies should be excluded or included without doing this selection in a systematic way. These problems may demand to re-conduct some activities (e.g., studies selection, data extraction, data synthesis) when they reproduce or update an SLR or even when they plan a new one. Therefore, a systematic process like an SLR may fail and some steps lack of a certain degree of maturity leading to poor quality results.

The SE community has discussed over years the barriers in conducting SLR studies (CARVER *et al.*, 2013), its main threats to validity (AMPATZOGLOU *et al.*, 2019), as well as other inherent problems like time- and effort-consuming issues (FELIZARDO; CARVER, 2020). Some problems of SLR studies resulted in the proposal of several practices (including methods, techniques, and tools) to mitigate them. For instance, there are tools supporting the SLR process (MARSHALL; BRERETON, 2013) to partially automate some SLR tasks (FELIZARDO; CARVER, 2020), such as: selection of the primary studies (WATANABE *et al.*, 2020), strategies to support updates (MENDES *et al.*, 2020b), strategies to create useful SLRs for the industry in a timely manner (e.g., rapid reviews (CARTAXO *et al.*, 2018)), and different ways to transfer the SLR findings to practitioners (e.g., evidence briefings) (CARTAXO *et al.*, 2018).

The main goal of this work is to contribute improving the SLR process and identify those practice areas that exhibit less quality. Our vision paper introduces the idea of a maturity model for SLR studies (MM4SLR) by identifying the priority areas aimed to conduct high-quality SLRs. To achieve our goal, we first analyzed a bunch of SLR studies in which we identified 84 practices, which we synthesized and distilled into 39 distributed around nine goals. After, we mapped these key practices with the CMMI[1] (Capability Maturity Model Integration) to understand how mature (CHRISSIS *et al.*, 2011) a given practice is and which should be the next level of maturity to be achieved by any SLR.

The remainder of this paper is organized as follows: Section D.2 presents the background and related works; Section D.3 presents our approach and research method; Section D.4 details part of our approach regarding the process areas and key practices that will be used in MM4SLR

[1] <https://cmmiinstitute.com/>

discussed in Section D.5. In Section D.6, we provide an initial validation of the proposed model and discuss our main findings in Section D.7. Finally, we outline our conclusions and future work in Section D.8.

D.2 Background and Related Works

Process models and maturity are two concepts that have been widely explored in SE and other areas to improve a given process and achieve the quality of products (TEAM, 2010). For instance, the 3C reuse model (WEIDE *et al.*, 1991) defines three elements (i.e. concept, content, and context) in which reusable components can be studied according to specific guidelines. Other well-known models for software development define several steps and phases (i.e. waterfall, iterative, spiral, agile, V-model, DevOps, among others) aimed to deliver better software. In other domains, process models like the Software Process Improvement Capability Determination (SPICE - ISO/IEC 15504) (DORLING, 1993) defines a framework around a reference model and two dimensions (i.e., process and capability) to assess software development processes and help organizations and software development teams achieve process improvement.

Another popular maturity model widely used in software development processes is the Software Engineering Institute (SEI)'s Capability Maturity Model (CMM) (PAULK, 2009) and updated in the 2000s as the Capability Maturity Model Integration (CMMI) model (CHRISSIS *et al.*, 2011) [2]. CMMI is a reference model for appraising software process maturity and a normative model for helping software organizations to evolve their processes from ad-hoc and chaotic processes to mature and disciplined ones (CHRISSIS *et al.*, 2011). The CMMI process defines five maturity levels: Initial, Managed, Defined, Quantitatively managed, and Optimizing.

One key component of CMMI is the *process area*, also known as Key Process Area (KPA). A process area is a cluster of related practices that, when implemented collectively, satisfies a set of goals considered necessary for improvement in that area. Another important component is the *specific goal* that refers to unique characteristics that must be addressed to satisfy that process area. These goals are satisfied through the execution of specific practices. Together with CMMI, international standards, such as ISO/IEC 12207[3] and ISO/IEC 15504-7[4], have also contributed to support organizations to go towards more mature processes.

Related works: Inspired by CMMI and these standards, maturity models have gone beyond software development as an attractive means in multiple areas that need tools to deliver more mature processes (WENDLER, 2012). For instance, Lee *et al.* (2007) present a business process maturity model (BPMM) where different maturity levels are associated with the influence of process areas and process improvement. The authors map the CMMI levels and KPAs (Key

[2] CMMI's latest version v2.0 was released in 2018
[3] ISO/IEC 122207: <https://www.iso.org/standard/63712.html>
[4] ISO/IEC 15504-7: <https://www.iso.org/standard/50519.html>

Process Areas) to concepts and characteristics of business processes. Those characteristics were identified by interviewing experts from four different business sectors.

Moreover, in Shen *et al.* (2021), the authors propose a capability maturity model (CMM) method to assess the performance of carbon emissions in cities to reduce carbon emissions. The proposed method helps to identify low-carbon city (LCC) strategies to assess in the performance of different LCC dimensions and identify the right KPAs and indicators in the context of LCC.

Overall, CMM can provide significant gains in productivity and quality (PAULK, 2009) as maturity models provide a means to evaluate which process areas demand more attention to make effective decisions (SHEN *et al.*, 2021), and a path for progressively maturing the processes, prioritizing actions to be done, and introducing a cultural change in organizations (CRAWFORD, 2021).

D.3 Approach and Research Method

In this work, we adopt the Grounded Theory (STRAUSS; CORBIN, 1998), which is a qualitative method where data is collected and analyzed to generate a new theory based on the comparison of concepts and produce higher-order themes. The data extracted is coded and grouped around similar concepts to derive meaningful categories that are used to derive the theory.

In a similar vein, thematic analysis (CRUZES; DYBA, 2011c) is used to cluster key practices into higher-order themes (Specific Goals and Process Areas) so we can identify which SLR steps can be mapped to process areas. Figure 13 shows the four steps to define our maturity model MM4SLR.

Step 1. Sampling data collection:, we selected two well-cited and well-known tertiary studies (AMPATZOGLOU *et al.*, 2019; ZHOU *et al.*, 2016) to extract the main practices. We selected these two studies based on our experience and knowledge in the field and because we did not find other tertiary studies that include the use of guidelines and definitions from well-known studies discussing the SLR process and because they analyze a set of SLRs conducted in the SE area.

Regarding the aforementioned tertiary studies, Ampatzoglou *et al.* (2019) and Zhou *et al.* (2016) analyze 14 relevant papers that suggest for conducting SLRs (BUDGEN *et al.*, 2018a; REVIEWS; DISSEMINATION, 2002; CRUZES; DYBA, 2011b; IMTIAZ *et al.*, 2013; KITCHENHAM *et al.*, 2004b; KITCHENHAM; CHARTERS, 2007; KITCHENHAM; BRERETON, 2013; PETERSEN *et al.*, 2008; PETERSEN *et al.*, 2015; SILVA *et al.*, 2011; WOHLIN, 2014b; DYBA *et al.*, 2005; KITCHENHAM, 2004; BIOLCHINI *et al.*, 2005). Both studies identify the most common threats to validity of the SLR process and strategies to mitigate them. Therefore, we can identify gaps in current SLR practice. In short, Ampatzoglou *et al.* (2019) analyzed 165 secondary studies published from 2007 to 2016, and Zhou *et al.* (2016) assessed

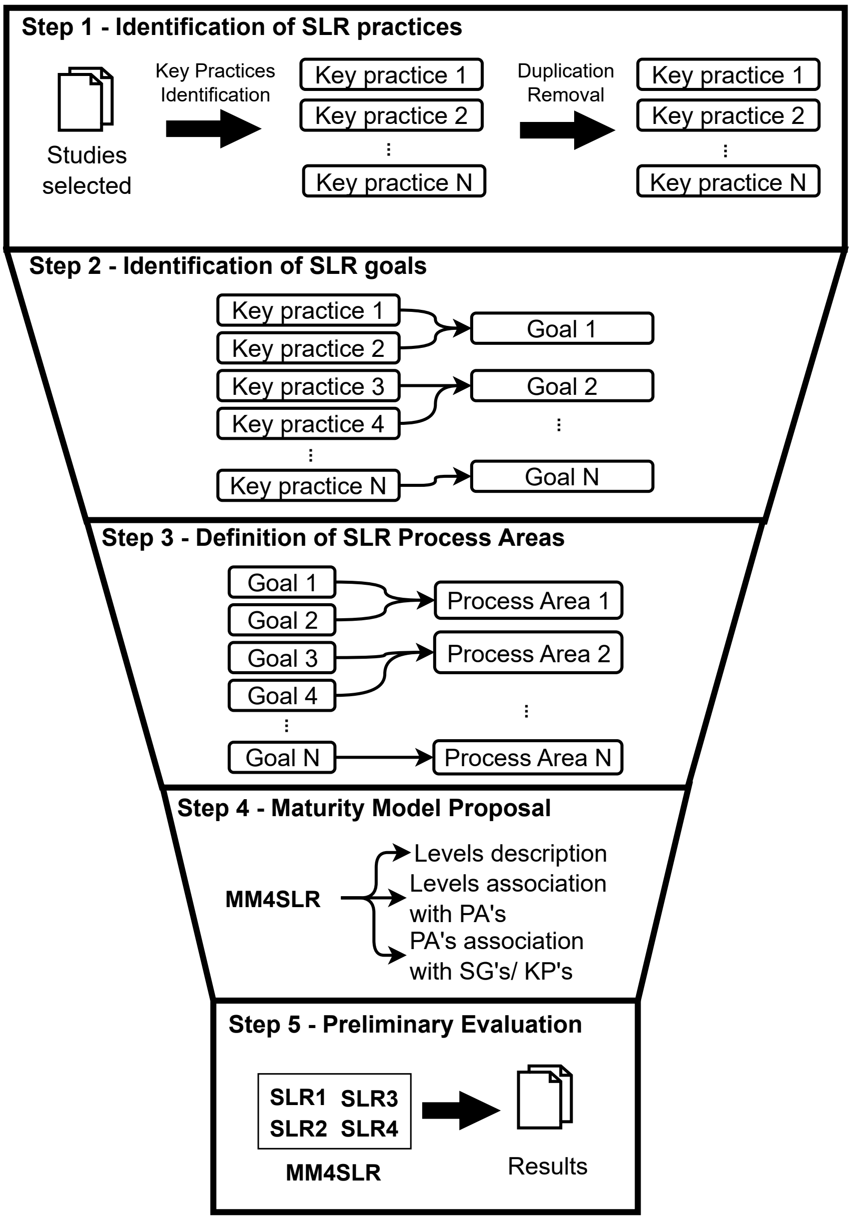

Figure 13 – Research method used to define MM4SLR

about 316 studies published from 2004 to 2015. We carefully examined the results of both studies and we identified 84 practices[5] that could be adopted to mitigate issues SLRs. We also synthesized the practices analyzed into 39 key practices that are discussed in Section D.4.

Step 2. Data coding: In this step, we grouped the 39 key practices into nine specific goals (detailed in Section D.4). We grouped similar practices that address the same threat to validity. For instance, several practices can be adopted to ensure good coverage of a given SLR (i.e., all relevant primary studies were considered or the use of snowballing techniques). As a results, these nine goals represent critical points in the SLR process that may deserve more attention from researchers (shown in Table 34).

[5] The list of 84 practices is available on <https://doi.org/10.5281/zenodo.7749269>

Step 3. Cluster codes into categories and derive higher-order themes: In addition, we generated clusters of specific goals that address common issues, which resulted in five process areas (detailed in Section D.4). For example, those specific goals associated with the definition of strategies for searching studies were assigned to the process area of Project Planning (PP). Hence, these areas represent broader topics that, once satisfied, could contribute to improving the maturity of the SLR process.

Step 4. Theory: We analyzed the process areas and, inspired in CMMI, we derived MM4SLR encompassing five levels of maturity for SLRs.

D.4 Process Areas, Goals, and Practices

Tables I to V present the 39 key practices that we mapped to nine specific goals (shown in Table 34). We also associated the specific goals with five process areas (i.e., Project Planning, Technical Support, Results Documentation, Process and Product Quality Assurance, and Communication Management), which are presented below. In the tables, the rows in red represent mandatory activities in conducting an SLR, those in green represent optional activities, and the yellow ones are recommended practices.

Project Planning (PP) is the first area and covers the entire SLR planning, which defines how the review will be conducted. SLR planning concerns, for instance, the need for a review, the amount of resources needed (including budget, time, and human resources), and the review protocol. Planning an SLR is critical because decisions at this stage significantly impact on the entire process. In this area, we identified eight practices (see Table I) associated with these goals. We marked three out of eight as mandatory activities or Key Process (KP1, KP2, and KP8) that any SLR shall fulfill, one as optional (KP3), and the other four ones as recommended practices. Please note that tools could help to analyze the selected venues in practice area KP8 but in most cases, manual intervention is needed (e.g., solving the discrepancies).

Some of the recommended practices marked in yellow are done in many SLRs, which can improve the quality of the studies. However, in other cases where low-quality SLRs are produced, some of these recommended practices are done wrongly (e.g., bad selection of studies) or simply neglected (e.g., conduction of a pilot or trial search to adjust the search string). Finally, for KP3 marked as an optional activity, it could be also categorized as recommended but we found that some SLRs do not provide such classification. In the case of the snowballing technique (KP7), it ensures the completeness of the studies selected but in many case, this task is not performed or the number of snowballing iterations is not enough. Overall, this process area can guide researchers in planning their reviews, avoiding common pitfalls, and offering directions to achieve results that impact their target audience positively.

Technical Support (TS) is another process area that refers to the adoption of supporting tools and search engines (of databases) during the conduction and update of SLRs. Hence,

Table 29 – Project Planning (PP) - Process Area

ID	Key Practice	SLR Activity	SLR Phase
KP1	Evaluate whether the research goals/questions are well-motivated	Define the need	Planning
KP2	Specify the inclusion/exclusion criteria	Protocol definition	Planning
KP3	Define a classification schema for the papers selected	Protocol definition	Planning
KP4	Use keywords from abstracts	Search strategy definition	Planning
KP5	Define appropriate years to limit the search space	Search strategy definition	Planning
KP6	Conduct trial searches to adjust the search string	Search strategy definition	Planning
KP7	Use snowballing technique	Search strategy definition	Planning
KP8	Manually scan the selected venues to check based on title, abstract, and full paper or combination of them	Paper selection	Planning

Table 30 – Technical Support (TS) - Process Area

ID	Key Practice	SLR Activity	SLR Phase
KP9	Use of tools to facilitate the search process	Study selection	Conduction
KP10	Select the most relevant digital libraries (DLs)	Search strategy definition	Planning
KP11	Consider the particularities of the interfaces of the DL to adjust the search string	Search strategy definition	Planning
KP12	Use tools for bibliography management	Study selection	Conduction

TS is essential to mitigate several problems such as the high-time consumption. Carver et al. (CARVER *et al.*, 2013) highlighted that the most time-consuming task is the search for studies in the databases, selection of studies, data extraction, and assessment of the quality of studies. Several specific tools have been proposed to support some of these tasks (e.g., SLuRp (BOWES *et al.*, 2012), SLR-tool (FERNÁNDEZ-SÁEZ *et al.*, 2010)) but most of the researchers use spreadsheets and reference managers to organize the results and remove duplicate entries (AL-ZUBIDY; CARVER, 2018). In this light, specific tools supporting the SLR activities are still immature (MARSHALL; BRERETON, 2013) and more efforts are needed to enhance and integrate them (FELIZARDO; CARVER, 2020).

In this process area, we identified four KPs as shown in Table II. We categorized two of them as mandatory (KP10 and KP11), i.e., activities that must be performed, while the other two (KP9 and KP12) are optional. We did not include in the discussion tools providing advanced visualization capabilities as most of them are still immature and research tools.

Results Documentation (RD) is the process area that provides ways to create a reliable and complete documentation of the conducted SLR, aiming that SLR results effectively impact the target audience. To create such documentation, we identified six practices, such as: documenting in detail the whole process, providing access to the entire raw data (for anyone that intends to replicate or update the SLR), and documenting the details of the SLR process. However, the SE community has noticed the lack of impact of the SLR results in SE practitioners (BADAMPUDI *et al.*, 2019) so they proposed new strategies for SLR documentation such as evidence briefings (CARTAXO *et al.*, 2018) aiming to impact the industry. From the six KPs shown in Table III, four of them were considered mandatory as SLR researchers need to publish

Table 31 – Results Documentation (RD) - Process Area

ID	Key Practice	SLR activity	SLR phase
KP13	Document the inclusion/exclusion criteria explicitly in the protocol	Final report	Reporting
KP14	Make data publicly available for replication	Data availability	Reporting
KP15	Document in detail the review process in a protocol	Final report	Reporting
KP16	Document and explain the details in conducting the review and how the disagreements where solved	Final report	Reporting
KP17	Use software engineering guidelines for reporting the studies	Final report	Reporting
KP18	Document the results of the search and selection process	Data availability	Reporting

all the details of the study for replication purposes and the techniques used. Only one key practice (KP14) was considered as recommended because not all SLR make available the data used and another KP (KP17) was marked as optional as we did not find much SLRs using SE guidelines.

Process and Product Quality Assurance (PPQA) groups encompasses 17 practices (see Table IV) focused on improving the quality of the SLR. In particular, these practices aim to mitigate bias in the SLR process, such as handling inconsistencies, ensuring the validity of results, ensuring the reliability of the venues and studies analyzed, and avoiding bias in data extraction among others. We acknowledge that many of these key practices require additional effort so this is why most of them are marked as recommended. For instance, we can perform pilot studies to adjust the search process using different keywords. Moreover, involving several researchers to solve the disagreements or cross-check the results is sometimes not easy to find. In addition, we found three mandatory key practices (KP20, KP23, and KP27) as tasks that must be always performed to achieve a minimum level of quality, while the other four key practices (KP21, KP29, KP30, and KP31) were considered optional. For instance, we could use sensitivity analysis formulas to know if the set of papers selected is the right ones. As more recommended practices are adopted more quality and less bias the SLR will have.

Communication Management (CM) is the last process area and aims at handling the communication among the researchers involved in the SLR. The need for communication is required to avoid misunderstanding of the previous KPs. From our analysis, we distilled four key practices that may improve communication issues and solve conflicts among researchers. For instance, using systematic voting, adopting kappa statistics, brainstorming possible interpretations of findings, and cross-checking data extracted from studies. CM provides means to gather and report insights from different perspectives, making the contribution of the SLR stronger. As shown in Table V, we identified one mandatory practice (KP37), two recommended ones (KP36 and KP38) that can improve the quality of the communication process, and one optional (KP39) that rarely happens.

Goals and process areas: In addition, from the classification and clustering of the aforementioned key practices, we derived higher-order themes where a set of specific goals map to each process area and its associated key practices, such as shown in Table VI. In this table, we can see that, for instance, the Project Planning (PP) process areas address goals SG1 and SG2

Table 32 – Process and Product Quality Assurance (PPQA) - Process Area

ID	Key Practice	SLR Activity	SLR Phase
KP19	Search exhaustively the related work to familiarize with the field; identify comparable studies; and, identify relevant publication venues and influential papers	Define the need	Planning
KP20	Remove duplicate studies and older versions of newer ones	Study selection	Conduction
KP21	Compare the selected studies to a gold standard	Protocol validation	Planning
KP22	Ask for an independent expert to review the search process	Protocol validation	Planning
KP23	Cross-check for the completeness of the studies selected	Study selection	Conduction
KP24	Perform pilot data extraction to test the agreement between researchers	Protocol validation	Planning
KP25	Perform pilot data analysis and interpretation	Protocol validation	Planning
KP26	Assess about the validity of primary studies and their impact using statistics	Data synthesis	Conduction
KP27	Discuss if the selected research method (SLR or SMS) answers the goals/research questions of the study	Protocol validation	Planning
KP28	Double-check the results to find inconsistencies in the search method	Protocol validation	Planning
KP29	Continuously update the schema until it becomes stable to classifies all the primary studies	Data extraction	Conduction
KP30	Conduct reliability checks (e.g., post-SLR surveys with experts)	Data extraction	Conduction
KP31	Use sensitivity analysis to select the right set of papers	Paper selection	Conduction
KP32	Check if results are inline with previous studies	Data synthesis	Conduction
KP33	Involve more than one researcher in the review process	Study selection	Conduction
KP34	Allocate more than two reviewers in data extraction process	Data extraction	Conduction
KP35	Allocate more than one expert to perform the classification to reduce bias	Study selection	Conduction

Table 33 – Communication Management - Process Area

ID	Key Practice	SLR Activity	SLR Phase
KP36	Use systematic voting	Study selection	Conduction
KP37	Discuss how to solve discrepancies of potential selected papers	Study selection	Conduction
KP38	Identify experts' disagreement level with the kappa statistic	Study selection	Conduction
KP39	Brainstorm with the authors about possible interpretations of the findings in an absence of related studies	Data synthesis	Conduction

aimed to adopt systematic methods and ensure the study complies with the needs and research goals. With regard to the Technical Support (TS) process area, this concerns goal SG3 and the use of tools to facilitate the conduction of the SLR. For SG3, we know two key practices are optional but any tool used to reduce the human burden to analyze the data is very welcome.

Specific goals SG4 and SG5 share some of the key practices. With these two goals, we wanted to highlight that all the data and the steps of the protocol used must be publicly available. However, for replication or update purposes, the practice area KP16 was not included in SG5 as it can be different if the SLR researchers are different authors.

The goals that refer to the process and product quality assurances (PPQA) are three. In SG6, we address the coverage and completeness of the studies selected and represented by the KPs shown in Table VI. The snowballing technique can help in this goal to find missing studies.

Table 34 – Specic Goals identified from key practices

ID	Specific Goals	Process Area	Associated KPs
SG1	Adopt systematic methods to search and manipulate studies	PP	KP2, KP4, KP5, KP6, KP7, KP8
SG2	Ensure that the study design complies with the research goals	PP	KP1
SG3	Adopt tools to support the search process	TS	KP9, KP10, KP11, KP12
SG4	Provide detailed documentation of the SLR process	RD	KP13, KP15, KP16, KP18
SG5	Make sure that all data is available replication and updates	RD	KP14, KP15, KP18
SG6	Ensure the coverage and completeness of the studies	PPQA	KP19, KP20, KP23, KP29, KP31
SG7	Adopt strategies to minimize bias	PPQA	KP22, KP28, KP30, KP3, KP35
SG8	Use techniques to ensure the validity of the results	PPQA	KP26, KP27, KP30, KP33, KP34
SG9	Use different stakeholders' opinions to double-check decisions taken during the planning/conduction of the SLR	CM	KP37, KP38, KP39

Additionally, we felt important to adopt strategies to minimize the bias (SG7) of the studies reviewed and ensure the validity of the results (SG8). Both goals are crucial because for SG7 sometimes we discuss studies that are not relevant for the research questions addressed or we select studies that do not fit well under the inclusion/exclusion criteria. Furthermore, goal SG8 must be supported by valid studies selected (KP26) to derive meaningful results using the right studies (KP27), but reliability checks (KP30) can be very helpful to cross-check the initial results. Such cross-checks are often performed by more than one researcher or expert on the topic (KP33 and KP34). Our last goal (SG9) addresses communication management issues that are basically supported by the three KPs shown in Table VI, which mainly bring experts to discuss possible discrepancies and interpretation of the results. We excluded KP36 (i.e., systematic voting) as there are several forms to achieve a consensus so, even a recommended practice, there might be different ways to come up with an agreement.

D.5 A Maturity Model for SLR Process

Deriving a theory is sometimes not easy but based on the tables discussed earlier and inspired in the CMMI maturity levels, this section presents an initial version of MM4SLR. We relied on CMMI 2.0 where the model defines six maturity levels (numbered from 0 to 5): incomplete, performed, managed, defined, quantitatively managed, and optimizing. For each maturity level, we mapped the process areas distilled in our approach to produce MM4SLR.

Regarding CMMI 2.0 **Level 0 (a.k.a. Incomplete)**, we did not include any process area for this level as it lacks any kind of systematization and will produce SLR with a lack of rigor in the process and the results as well.

Level 1 - Performed: The main concern of this level is the management of the SLR process. This level addresses two process areas: PP (which deals with SLR planning), and RD (which concerns the SLR documentation) making the review process managed and controlled with standardized procedures. This means that the SLR process accomplishes requirements towards its maturity and, hence, it can be considered managed. At this level, the quality of the SLR is unknown as it depends on the specific KPs enacted and the reliability of the results is questionable.

Level 2 - Managed: At this level, the SLR process is well-defined and documented and the main focus of this level is to assure the SLR quality. The main difference is the Technical Support (TS) that is used to manage the planning, conduction, and reporting of SLR. At this level, there is a deeper comprehension of quality control mechanisms that are regularly mentioned, but the SLR quality in this level is still a challenge.

Level 3 - Defined: At this level, the SLR is conducted in a proactive manner making every effort to continuously improve their quality. To fulfill the requirements of this level, the SLR process must address two process areas: PPQA (which refers to quality assurance) and CM (which deals with communication). The SLR process assures the SLR quality, and bias are minimized in handling disagreements. Many practices refer to double-checking SLR activities or running assessments, such as the quality appraisal of primary studies and sensitivity analysis. CM is also assigned to this level because it is a well-defined activity that enhances the communication issues between SLR researchers and other stakeholders.

Level 4 - Quantitatively Managed: At this level, the SLR process is quantitatively managed by metrics that ensure a systematic control of quality and effort applied in each activity. This level prioritizes the use of metrics, for instance, the sensitivity formula proposed by Zhang and Babar (2010) that is defined as the proportion of studies retrieved by the search strategy adopted and the number of relevant studies retrieved. Specifically, statistical analysis of the results becomes essential and the impact of each primary study can be measured.

Finally, **Level 5 (a.k.a. Optimizing)** is hard to achieve and was mapped to MM4SLR as it refers to the highest level of maturity in which researchers are concerned with continuously evaluating and improving their SLR process. As SLRs are produced every certain number of years when the state of the art advances and updates of a previous SLR are uncommon, it seems hard to achieve a continuous evaluation of SLR studies and also because the immaturity of some SLR tools does not support the improvement of the overall SLR process. Consequently, we focused on CMMI levels 1 to 4 to produce MM4SLR where relevant KPs are identified for different process areas.

D.6 Validation

This section provides an initial validation of our model using four SLRs, as summarized in Table 35. To determine the maturity level of each SLR, we adopt the criteria that each process area should have all the mandatory KPs and at least the 50% of the overall number of KPs of each process area to consider whether a study fulfills a given process area. We analyzed these SLRs but we anonymized the sources to avoid authors could be disappointed from our rating. We selected SLR1 and SLR2 from two top venues ranked in the first quartil (Q1) while SLR3 and SLR4 were published in other journals non-ranked in the Journal Citation Reports.

For instance, in SLR1, the study has all the mandatory items in red and has 5 out of 8 KPs. Therefore, SLR1 fulfills the Project Planning (PP) process area. After the evaluation, we can observe in Table VII that this SLR fulfills the following process areas: PP, TS, and RD. In the case of PPQA, there is one mandatory KP (KP23) for which we could not find clear evidence so we excluded it likewise the communication management (CM) process area. Consequently and according to MM4SLR, we can state that SLR1 meets Level 2 (Managed) but it requires more KPs to achieve Level 3. Regarding SLR2, we mapped to the same maturity level with some small variations in the recommended and optional KPs compared to SLR1.

The case of SLR3 is a bit different. We rated it with Level 0 (Incomplete) because from our evaluation, there is one mandatory KP (KP16) not fulfilled so this SLR cannot be in Level 1 (Performed). As the authors did not discuss how the disagreements in reviewing and classifying studies was done, we believe the SLR cannot reach that particular maturity level. Although one could think it is only one mandatory KP not addressed, this KP is important to avoid bias in the selection and classification of the final set of studies.

Finally, we rated SLR4 as Level 2 (Managed) as it fulfills with the mandatory KPs for that maturity level. This SLR has other interesting KPs belonging to the PPQA process area that other studies do not have (e.g., KP36, voting system).

D.7 Discussion

Well-defined and well-known guidelines and processes to conduct SLRs exist (KITCHENHAM *et al.*, 2015) but many researchers do not follow them systematically. At the same time, they are increasingly conducting and publishing SLRs without much concern for the quality of the process and product (i.e., SLR results).

From a broader perspective, it is not difficult to find SLRs published that are in maturity levels 1 or 2, which may produce in some sense SLRs with low or average quality. Probably, most of SLR that are evaluated by specialists belong to Levels 2 or 3, meaning that their authors are concerned with the quality of the SLR and results as well. On the other hand, SLR processes and key practices in Level 4 are more uncommon but reachable. Our MM4SLR helps researchers

Table 35 – Maturity level of sample SLR studies

Key practice	SLR1	SLR2	SLR3	SLR4	Key practice	SLR1	SLR2	SLR3	SLR4
KP1	X	X	X	X	KP21	-	X	-	-
KP2	X	X	X	X	KP22	-	X	-	-
KP3	-	X	-	X	KP23	-	-	-	-
KP4	X	X	X	X	KP24	-	X	-	-
KP5	X	-	X	X	KP25	-	X	-	-
KP6	-	X	-	X	KP26	X	-	-	-
KP7	-	X	-	-	KP27	X	-	-	X
KP8	X	X	X	X	KP28	-	-	-	-
KP9	X	-	-	-	KP29	-	X	-	-
KP10	X	X	X	X	KP30	-	-	-	-
KP11	X	X	X	X	KP31	-	-	-	-
KP12	-	-	-	-	KP32	-	-	-	-
KP13	X	X	X	X	KP33	X	X	-	X
KP14	-	-	-	-	KP34	X	X	-	-
KP15	X	X	X	X	KP35	X	X	-	X
KP16	X	X	-	X	KP36	-	-	-	X
KP17	-	X	-	-	KP37	X	X	-	-
KP18	X	X	X	X	KP38	-	X	-	-
KP19	-	X	-	-	KP39	-	-	-	-
KP20	X	X	X	X					
Maturity Level						(2) Managed	(2) Managed	(0) Incomplete	(2) Managed

to evaluate and understand which KPs are still missing to achieve a higher maturity level and which effort will be required. Regarding this effort, it could be easier to roughly estimate how much time could take and how many researchers are needed. This may also depend on the tool support used, the number of studies to analyze, or if snowballing is used, for instance.

The tables presented in this work and the three types of KPs according to their importance are a first attempt to systematize the quality of SLR studies. Moreover, they can help to update existing SLRs when the technologies and the state of the art advance, even if reviews/authors are different from the original SLRs. We do not want to introduce extra complexity or additional tasks in typical SLR processes but rather, make the process more understandable and highlight the quality aspects that make the difference between poor and good SLRs. In addition, we expect to verify the adequacy and distribution of the proposed specific goals and KPs in each process area as some KPs may belong to more than one process area.

Finally, one crucial point for discussion is to agree on how many KPs must be considered for a given SLR to achieve a particular maturity level. In our evaluation, we relaxed the number of KPs to 50% but each organization could decide how much this percentage should be and tune the proposed rating to a higher level, specifically in those process areas that encompass more recommended items such as the case of PPQA, because the number of mandatory and recommended KPs is very unbalanced.

D.8 Conclusion and Future Works

This paper introduced the notion of a maturity model (called MM4SLR) to improve the current practice of SLR studies. MM4SLR clarifies the mandatory and recommended activities that may help researchers to conduct SLRs. As each KP requires a certain effort (some of them assisted by tools), we advise on the most important KPs to produce better quality SLRs. We also emphasize the role of quality and communication aspects as crucial to produce better SLRs. In addition, we argue the need for more tools to help SLR researchers in their tasks to achieve a certain level of automation that could help in future replication processes. Overall, we encourage addressing as much as KPs to ensure the completeness, accuracy, and validity of the results in future SLRs. All in all, MM4SLR serves as a first attempt to guide SLR researchers in achieving better quality reviews based on the right fulfillment of the KPs.

As research agenda for the future, we suggest the following topics: (i) additional validation of the model with more SLRs or even systematic mapping studies (SMS); (ii) metrics to estimate the effort and number of researchers needed for different SLR; (iii) means to achieve easier Levels 4 and 5 with new tools that can automate some of the existing manual tasks; (iv) rules to know how many KPs and which ones are necessary to achieve a given maturity level; and (v) evolution/refinement of MM4SLR with near KPs or update/remove existing ones based on new goals.

Acknowledgments: This work was supported by FAPESP (2015/24144-7, 2019/23663-1), Capes (PROEX-11308091/D), and CNPq (313245/2021-5).

ICMC USP
SÃO CARLOS