Glaserian Systematic Mapping Study: An Integrating Methodology

Gustavo Navas[1,2][a] and Agustín Yagüe[1][b]
[1]*Escuela Técnica Superior de Ingeniería de Sistemas Informáticos ETSISI, Universidad Politécnica de Madrid (UPM), Calle Alan Turing s/n, Ctra. de Valencia, Km. 7, Madrid, Spain*
[2]*Universidad Politécnica Salesiana, Morán Valverde S/N y Rumichaca, Quito, Ecuador*

Keywords: Glaserian Grounded Theory, Systematic Mapping Study, Qualitative Analysis, Software Development.

Abstract: This research arises as an answer to the limited classification capability that reaches the vast majority of selected articles within a Systematic Mapping Study (SMS) when studying the Grounded Theory (GT) in Software Development. The result of our research is Glaserian Systematic Mapping Study (GSMS). It is a methodology that combines SMS and Glaserian Grounded Theory (GGT), which is one of the two variants of the GT. Combining the robustness and sequential process of SMS with GGT and its iterative features, GSMS provides a more robust, flexible, iterative, and scalable methodology. SMS and GGT share two main activities, data collection and data analysis. However, they are conducted differently. The resulted integration takes advantage of this fact and maps both related activities and outcomes to produce a more robust and systematic methodology. In addition, our research formalizes equations to represent the typical data saturation of qualitative methods such as GGT. With GSMS, we were able to classify more articles than with SMS alone.

1 INTRODUCTION

When a Systematic Mapping Study (SMS) is applied, the results may not be sufficiently complete. Sometimes, the results are not significant enough because it is not always possible to fully classify the source data when document analysis is conducted. This low ranking ability could be the main reason why SMS fails, and some articles that could not be classified are discarded. Making an SMS with a higher percentage of coverage would imply a heavy preliminary phase of code definition and document classification. Classification is essential because the mapping results are not significant enough with a low volume of items.

This fact has been verified by us when conducting a SMS to study the application of Grounded Theory (GT) in Software Development. Our motivation arises because when conducting the SMS, the results obtained were not as promising as expected. In that initial work, we got similar results to other previous reviews of the literature on GT in software engineering were the classification rate of articles did not cover more than 55% of total number of papers, and according to (Stol et al., 2016), this is due to an inadequate GT application. Their arguments did not convince us because some of the discarded works had significant contributions, well-established theories on the subject they dealt with, and essential findings in software engineering and grounded theory. Our goal was to understand better the causes of this low level of classification rate. We changed the focus to look for a way to increase the classification capabilities of SMS without heavy coding processes.

We proposed an integration between SMS and GT to increase the coverage rate, providing a more robust classification mechanism that complements the rigour of systematic mapping studies with the flexibility of grounded theory through iterations. Our approach is based on Glaserian Grounded Theory(GGT) that starts data analysis without any preconceived notion. In our research, we started analyzing 70 articles and in the end, we were able to have a high classification rate and obtained theories about how to apply GT in software engineering.

The structure of this paper is as follows, Section 2 is the background and Section 3 describes the guidelines applied in the integration process. Section 4 develops the integration to produce what we call Glaserian Systematic Mapping Studies (GSMS) and Section 5 covers a case of application of the GSMS. Finally, some conclusions and future work are related.

[a] https://orcid.org/0000-0002-2811-0282
[b] https://orcid.org/0000-0002-4761-0901

Navas, G. and Yagüe, A.
Glaserian Systematic Mapping Study: An Integrating Methodology.
DOI: 10.5220/0011090500003176
In *Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2022)*, pages 519-527
ISBN: 978-989-758-568-5; ISSN: 2184-4895

2 BACKGROUND

This section presents a brief description of the SMS and GT. The word ”step” is being used when referring to SMS processes, ”stage” when referring to GT, and the word “phase” in the proposed methodology.

2.1 Systematic Mapping Study (SMS)

Systematic mapping is mainly used in medicine but is increasing its relevance in the research of Software Engineering (Petersen et al., 2015). The Systematic Mapping Study (SMS) is a rigorous review process of the scientific literature. SMS establishes a well-defined methodology that allows mapping scientific articles (Kitchenham et al., 2011) and reduces the bias of people’s opinions. As proposed by Petersen et al., there are six steps for the SMS processes: Step 1: **Research questions definition:** see Figure 1(a) P1. Step 2 **Conduct Search:** see Fig. 1(a) P2. Step 3: **Screening of papers**: see Figure 1(a) P3. Step 4: **Keywording:** see Figure 1(a) P4. Step 5: **Mapping:** see Fig. 1(a) P5. Step 6: **Synthesis:** see Figure 1(a) P6. Moreover, we have also included **Rigour and relevance assessment** Figure 1(a) P7 as an additional step of SMS proposed by Paternoster (Paternoster et al., 2014).

Figure 1(a) depicts SMS steps and their outcomes. On one side, P1, P2, and P3 share the same goal: collecting data and selecting the scientific articles to be analyzed. On the other side, P4, P5, and P6 deal with the analysis and mapping of results and finally, P7 is an approach to evaluate the scientific quality.

2.2 Glaserian Grounded Theory

Grounded Theory is a qualitative research methodology proposed by Glaser and Strauss in 1965 and consolidated in 1972 (Glaser and Strauss, 1973). GT is a methodology that generates a substantive theory about the topics under research, their concepts, and categories through constant and systematic comparison of data during the process. GT has evolved into two variants: Glaserian Grounded Theory (GGT)(Glaser, 1992) and Straussian Grounded Theory (SGT)(Van Niekerk and Roode, 2009).

The main difference between GGT and SGT is based on the role of researches and the starting point. GGT emphasizes an open attitude where theories do not come from researchers’ preconceptions; however, they come from the data. In the case of SGT, the researcher must apply a set of tools and procedures having an active role to use existing insights and experience during the research(Strauss, A. and Corbin, 1990). These differences can be summarized as follows GGT is independent of the researcher’s ideas, while the researcher’s views influence SGT. Within GT variants, behavior is the way of facing a problem or concern in an area of study. GGT’s behavior is given by the generation of concepts and relationships that explain and interpret its variation in an area of study. On the other hand, SGT describes the full range of behaviors (Sharma and Biswas, 2015).

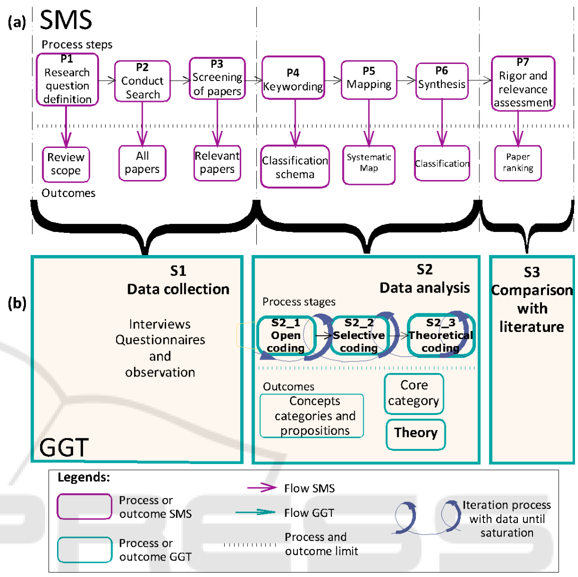

Figure 1: GGT and SMS integration.

GGT is the grounded theory approach selected due to, as it was stated by Stray et al.(Stray et al., 2016), the ease arising of research questions during data analysis, allowing concepts and categories to emerge from the data with more flexibility.

According to Glaser, GGT is a research methodology with two stages (Glaser, 1992), **Data collection** and **Data analysis**. Later, Adolph (Adolph et al., 2008) enriched this methodology, including **Comparison with the literature** as a new stage. These stages have been depicted as S1, S2, and S3 in Figure 1(b).

GGT comprises three substages: **Open Coding**, Figure 1 (b) S2_1. **Selective Coding** Figure 1 (b) S2_2. **Theoretical Coding** Figure 1 (b) S2_3 and produces two outcomes: **Core Category** & **Emerging Theory**. As previously mentioned, **Data analysis** is a data modeling process to discover information, extract conclusions, and support hypotheses that start with data gathered in the previous stage and end when a theory appears (Parizi et al., 2014). Concerning GGT, it could also be considered as a data analysis tool. It provides an iterative mechanism to build theories from data conducting.

3 INTEGRATION DESIGN

The integration of GGT stages and SMS steps increases the flexibility of SMS and systematizes the stages of GGT. As it was stated by (Finfgeld-Connett, 2014), this is a complex task because there are no previous recommendations on including GT in systematic reviews. The integration done by mapping between steps and stages has twofold objectives: i) to identify the relationship between step and stage, and ii) to determine the expected outcomes of the integrated research methodology.

Both methodologies share two main activities, "Data collection" and "Data analysis", but they have been conducted differently. The proposed integration of SMS and GGT takes advantage of this fact and maps both related activities and outcomes to produce a more robust and systematic methodology. GGT, as a more open research method, has driven the integration process. The resulting method has the same structure as GGT comprising three phases: Data collection, Data analysis, and Comparison with literature. The correspondence between SMS phases and GGT stages is shown in Figure 1. SMS steps are integrated into GGT stages to be systematically applied to produce and collect data in GGT that, lately, are analyzed in the "Data analysis" stage. Figure 1 shows the relationship between the SMS steps P1, P2, and P3 and the "GGT Data collection" stage. In the same way, steps P4, P5, and P6 are integrated into the "GGT data analysis" stage and, finally, step P7 into the "GGT Comparison with literature" stage.

The integration starts mapping the **"Data collection"** GGT stage and the SMS steps dealing with the acquisition of data sources. In traditional SMS, **P1** is the first step that could be mapped into GGT with questionnaires or interview elaboration over the topic under research. The integrated methodology corresponds to the definition of research questions because GT is very permissive (Zayour and Hamdar, 2016) and very diverse in data collection. On the other hand, some studies use documents to replace interviews and questionnaires (Adolph et al., 2008). **P2** step corresponds to Conduct Search. The integrated methodology has the same meaning as in the traditional SMS. It is equivalent to the writing and debugging process of the answers when conducting interviews or passing questionnaires for being analyzed in the conventional GGT. The outcome of this step is the list of articles found in the scientific libraries complying with the search string. The **P3** step is the paper screening by applying the inclusion and exclusion criteria to obtain the relevant papers that should be analyzed. In GSMS, it will be used like in the traditional SMS. The GGT perspective could be compared with the transcription process and classification of interviews and questionnaires to produce working artifacts. The outcome of this GSMS phase is the collected data that will be used as the input for the data analysis phase.

Data analysis is the procedure for conceptualizing and analyzing data, including identifying relationships. It is an integration process that increases the abstraction level that starts based on the results of the previous phase and ends when the theory appears (Parizi et al., 2014). The integration between GGT and SMS in data analysis is feasible because both share the goal of data modelling to form conclusions and formulate the hypothesis. The integration of "Data analysis" is more complex than "Data collection" because GGT allows multiple iterations, while SMS is sequential. Our proposal has unified them to provide an iterative process comprising GGT stages but adapting their scope and level of abstraction in terms of the SMS steps. The SMS processes **Keywording, Mapping** and **Synthesis**, shown in Figure 1 labels P4, P5, P6 comprise **Open**, **Selective** and **Theoretical coding** stages to produce the outcomes depicted as O4, O5, and O7 in the same figure.

Open, selective, and theoretical coding have different meanings depending on the SMS step even when they are applying the same concept. Therefore, the outcome of P4 (keywording) comprises several classification schemes (O4) that are used as inputs in P5 (mapping) to produce relationships between essential elements in terms of concepts, categories and propositions. These elements are classified and refined to obtain systematic maps (O5). Maps represent the inputs to P7 (Synthesis). They are combined to produce the Core Category, the Emerging Theory (O7), and could answer some research questions or identify new ones.

One of the contributions obtained by integrating GGT and SMS is the possibility to deal with those new research questions that may arise as an outcome during this phase. The number of iterations is directly related to the existence of unresolved questions. Finally, **Rigor and relevance assessment** SMS step, shown in Figure 1 label P6 in the figure, is only applied at the end of each iteration and integrated with the **"Comparison with the literature"** stage of the GGT to produce a paper ranking as the outcome O6 in the figure.

4 GLASERIAN SYSTEMATIC MAPPING STUDY (GSMS)

In brief, **Glaserian Systematic Mapping Study (GSMS)** could be summarized as a new approach, unifying the rigor and mapping of the SMS, with the flexibility of GGT for building theories. **GSMS** induces an iterative process in the Data Analysis phase that allows the construction of theories by promoting a deeper understanding of the results by generating conceptual propositions and including new research questions when they are discovered. **GSMS** applies GGT in two ways: as a qualitative research methodology and data analysis. As a qualitative research methodology(Glaser, 1992) that encompasses the SMS process, and second, as a data analysis tool used by SMS steps to build emerging theories (Van Niekerk and Roode, 2009), as is shown in Figure 2.

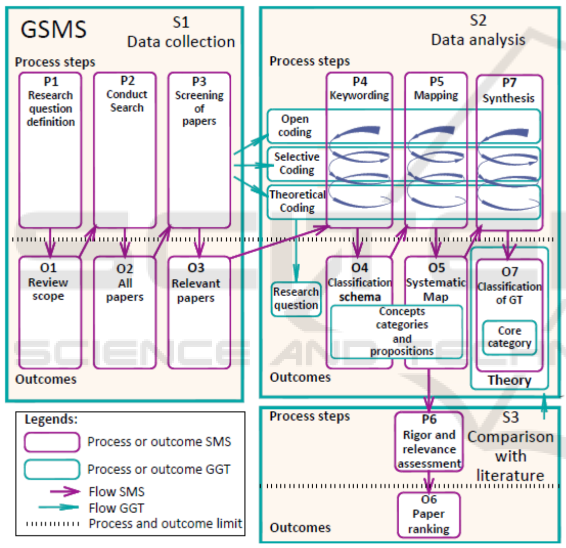

Figure 2: Glaserian Systematic Mapping Study (GSMS).

4.1 GSMS Data Collection Phase

This section describes the GSMS data collection comprising three subsections: Research questions definition, Conduct Search, and Screening of papers.

A. Research Question Definition Phase. The GSMS Research question definition phase is the initial research question or questions. In GSMS, the research question is recommended to be generic, using in its formulation the paradigm proposed by GGT, "What do we have here?" and it also affects its outcome "Review Scope" on the topic under study. Therefore, they drive the data analysis phase, as illustrated in Figure 2 label P1 with its corresponding outcome Review Scope Figure 2 O1.

B. Conduct Search Phase. In GSMS, there are several activities to perform "Conduct search phase". The first is to select the initial sources from scientific libraries like ISI WoS, IEEE, Scopus, or ACM. The second activity is to apply a rigorous and systematic procedure to define the search string (Petersen et al., 2015). Through carrying out a series of tests, this search string must be appropriate to ensure the result to produce the outcome O2, as is shown in Figure 2 P2 and O2.

The definition of the search string is critical and should look for a balance between precision and generality. If the topic under research is quite general, the search string should be general, but precise. However, if the search string is too detailed, some relevant publications could be discarded because of the inclusion/exclusion criteria. In our case, the research started with a complex search string like **("Grounded Theory" & "Software Development" & "requirement" & "design")**. Our results contained very few relevant publications due to being too specific. Later, we used a simplified expression with only two terms **("Software" AND "Grounded Theory")**.

C. Screening of Paper Phase. In GSMS, in "Screening of paper phase", Figure 2 label P3, a series of inclusion and exclusion criteria are applied to the scientific articles obtained in the previous phase. After this first filtering process, the obtained articles were passed to a snowballing process to identify additional ones (Kitchenham et al., 2011).

Within the GSMS, this is the last phase of "Data collection." As is shown in Figure 2 O3, two are the expected outcomes of this phase: i) the list of research questions and ii) the set of relevant papers representing the basis for the GSMS data analysis process.

4.2 GSMS Data Analysis Phase

The GSMS data analysis phase incorporates a series of iterations that starts with the data obtained in the previous phase and ends when a certain saturation level is achieved. The saturation level is independent of the researcher's criteria, and it should be only based on the systematic review of the elements produced in this phase. To consider that an iteration is finished, the following conditions must be verified: i), the iteration has led to one or more new research questions. ii), the iteration has answered at least one research question. iii), the iteration has fully answered a specific research question. It means that an iteration can require more than one loop through the corresponding processes before being considered finished. The data analysis phase is finished by achieving the saturation level at the end of an iteration when all research

Table 1: Outcomes of GSMS Data Analysis.

	Keywor..	Mapping	Synthesis
Open	Concepts	Concepts	Concepts
Selective	Catego-	Catego-	Catego-
Theoretical	Proposi-	Proposi-	Proposi-

questions have been wholly and thoroughly answered. This fact has been formulated in Section 4.2.

The three GSMS phases, Keywording, Mapping, and Synthesis, incorporate open, selective, and theoretical coding activities, and they come about in an integrated and coordinated way through iterations. The Rigor and relevance assessment phase is conducted at the end of all iterations. It is essential to highlight that open, selective, and theoretical coding give rise to different outcomes that are concepts, categories, and propositions, respectively. The concepts are basic ideas emerging from data, i.e., words, keywords, codes, notes or diagrams. They are used to relate them, creating categories (Adolph et al., 2008). These categories could be lists, relationships, or any other abstract elements based on the previous concepts. Finally, the propositions connect the concepts and categories, producing a discursive set of theoretical statements relating to them. They are validated through constant data comparison (Chun Tie et al., 2019). Table 1 presents the type of expected outcomes of each activity in the GSMS data analysis.

A. GSMS Keywording Phase. The GSMS keywording phase looks for obtaining a Classification Schema, see Figure 2 label O4. A Classification Schema is a set of elements comprising concepts, categories, and propositions. Concepts are the result of applying abstraction during the **Keywording Open Coding**. Open coding is driven by the identified research questions. While performing the open coding activity, wording lists, memos, and codes are analyzed in-depth to identify concepts arising upon active research questions (Crabtree et al., 2009). The next activity is **Keywording Selective Coding** where categories emerge. Categories represent an upper level of abstraction built on top of concepts. The deepening of the analysis of these concepts gives rise to categories encompassing them. Categories could also be a list of relevant concepts. Finally, **Keywording Theoretical Coding** refers to the highest level of abstraction to define propositions to support emerging theories. Propositions determine theoretical knowledge based on consolidated statements built on concepts and categories. Summarizing, the outcomes of this phase are classification schemas for: concepts, categories, and propositions.

B. GSMS Mapping Phase. The GSMS mapping starts after the keywording phase with the goal of creating systematic maps. During the open coding activity, concepts becoming from keywording are mapped between them, generating new concepts to support the mapping. While selective coding is conducted, associations among existing categories and relationships between concepts and categories are identified to map new categories. Later, conducting the theoretical coding activities, new propositions could emerge from the maps. In this phase, some questions becoming from the previous iteration could be answered. The outcomes of this phase are mappings of concepts, categories, and propositions.

C. GSMS Synthesis Phase. In this GSMS phase, concepts, categories, and propositions that have arisen from the previous phases are the basis for the **Synthesis Open Coding** to produce more abstract and complete concepts. These concepts are deepened and analyzed for establishing the finals categories while conducting **Synthesis Selective Coding**. Finally, with constant data comparison, the final propositions are obtained in the **Synthesis Theoretical Coding**. It is shown in Fig. 2 P7. The outcomes of the GSMS synthesis can be one of these: i) Generation of a new research question emerging as part of the GSMS process. ii) An answer to a research question previously established, and iii) A fundamental proposition that will give rise to an emergent theory on a topic in a later iteration.

Keywording and/or mapping results are incorporated into the open coding activity of the synthesis. Their integration is achieved through selective coding, and then reaches the final proposition, that is, theoretical coding within the synthesis.

D. Formulation of Saturation in GSMS. Saturation has been modeled using two types of sets. **Q** is the set of containing all research questions generated. And, **AQ** represents the set of answers to a specific RQ. The GSMS Data analysis process ends when the following conditions are met at the same time: i) There are no new research questions in the iteration, ii) All research questions have been answered, and iii) No new elements have been added in the actual iteration to any of the sets.

These sets could be formalized as follows:

Equation 1. There are no new research questions in the iteration that could be reformulated in this way, the next iteration does not generate any new research questions. It is formulated as:

$$\left.\begin{array}{r} \text{Given iteration } k; \\ Q_k = \left\{ \sum_{i=1}^{n} RQ_i \right\} \\ \text{Given iteration } k+1; \\ Q_{k+1} = \left\{ \sum_{j=1}^{m} RQ_j \right\} \end{array}\right\} \Rightarrow \begin{array}{c} \text{No more questions when:} \\ Q_{k+1} - Q_k = \emptyset \wedge \\ Q_k - Q_{k+1} = \emptyset \end{array} \quad (1)$$

Let Q_k be the set of research questions at iteration k and Q_{k+1} the set at iteration $k+1$.

Equation 2. There is at least an answer to every research question in the set Q(k). It is formulated as:

$$\left.\begin{array}{r} \text{Let } Q = \left\{ \sum_{i=1}^{n} RQ_i \right\} \\ \text{where } RQ \text{ is a Research Question} \\ \forall RQ \in Q \exists \text{ an iteration } k. \text{ where is true that} \\ \exists A_k \neq \emptyset \bigwedge A_k = \left\{ \sum_{j=1}^{m} AQ_k(j) \right\} \\ \text{where } AQ_k(j) \text{ is an answer to } RQ_i \end{array}\right\} \quad (2)$$

Equation 3. Saturation is reached when there are no more answers to each question in the set Q(k). This condition is formulated as:

$$\left.\begin{array}{r} \text{Let } Q = \sum_{i=1}^{n} RQ_i \text{ considering } \forall RQ \in Q \\ \text{Given the iterations } k. \text{ It is true that} \\ \exists A_k(i) = \sum_{j=1}^{m} AQ_j \ \& \ A_{k+1}(i) = \sum_{j=1}^{n} AQ_j \\ A_k(i) \text{ is an answer to } RQ \text{ in iteration } k \bigwedge \\ A_{k+1}(i) \text{ is an answer to } RQ \text{ in iteration } k+1 \\ \text{It is true that } A_k(i) = A_{k+1}(i) \end{array}\right\} \quad (3)$$

Data Analysis Phase Iteration and Loops. GSMS data analysis is the most complex process in the methodology. It comprises iterations with well-stated goals. Iterations are the mechanism to reach the appropriate saturation level in the research. Each iteration receives inputs and produces outputs, and the output of one iteration is the input of the next, except in the first one, where the iteration's input is the output of "Data collection". To achieve the expected goals, iterations could require one or more loops. In GSMS, we use the term loop to refer the complete execution of the phases Keywording, Mapping, and Synthesis. Once the loop is finished, it is evaluated whether the goal of the iteration has been achieved or not. In the case of not, a new loop starts; but in the case of achievement, the process is moved to the next phase. Figure 3 (b) shows an example of an iteration with one loop. We used activity diagrams to model iterations, due to the existence of the fork/join framework that supports the branches that could happen during the execution of an iteration.

The constant comparison of data and its theoretical sampling is present in each one of the loops within the iterations; without them, it would not have been possible to delve into the process of finding theories through propositions. The proposed GSMS establishes the possibility that questions coming from a general scope evolve dynamically to be specific, abstract, and challenging ones.

4.3 Comparison with Literature Phase

The set of answers to the questions arising in the process are the input elements for this phase, having the scientific rigour and industrial relevance assessment phase as part of it. It is conducted at the end of each iteration when one or more answers to research questions were provided. There are some relevant considerations about this phase: **First**, It allows reviewing tertiary articles, thus establishing a difference with the primary articles obtained in data collection. It looks for to confirm or deny the findings of the iteration through other works around the subject of study. **Second**, the answered questions must provide a set of articles to compare with the arising emerging theories. **Third**, GGT establishes that should not be pre-established validation criteria at the beginning of the research; however, at the end of each iteration it is the time to give rise to answer a question and the findings, that will be compared and validated, to look for similarities and differences.

A. Rigor and Relevance Assessment Phase. For the elaboration of the rubrics for scientific rigor and relevance industrial, we took the recommendations given by (Ivarsson and Gorschek, 2010). The rigor and relevance were applied to the answers of each research question through the coding activities.

5 CASE STUDY

This section provides an example of the application of GSMS to study the use of grounded theory in software development. It comprises two sections; the first explains how the data collection was conducted, and the second presents how the data analysis was performed.

5.1 Data Collection

Data collection was applied as described in section 4.1. The search string used was **("Software" AND "Grounded Theory")**. After the filtering process, 70 research papers were selected. Figure 3 (a), depicts an activity diagram with the process applied. This research is based on a previous systematic mapping conducted by the authors to study where and how Grounded Theory was applied in Software Development through the following research questions: **RQ1:**, Where is the Grounded Theory (GT) appropriate within the Software Development study?, **RQ2:** Is the GT applied correctly in the process and tasks of Software Development?, **RQ3:** Is the GT useful for Software Engineers in the industry?

5.2 Data Analysis Iteration Example

This section describes the application of the first iteration of GSMS data analysis in our research. Each iteration is presented with the following structure: i), inputs received from the previous iteration or phase as appropriate. ii), the application of SMS steps and GGT stages. iii), the corresponding outputs for the next iteration/step. The input of the first iteration is the output of the ”Data collection” and comprises: three research questions (**RQ1, RQ2, RQ3**) and 70 papers to be analyzed. It is important to remember that open, selective, and theoretical coding give rise to concepts, categories, and propositions, respectively. Figure 3 (b) shows the Activity Diagram of iteration.

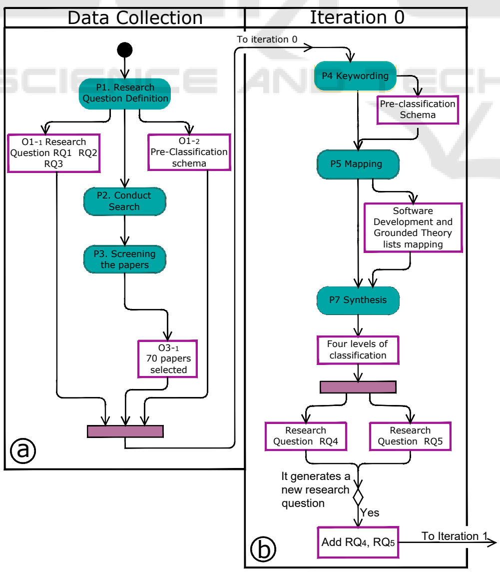

Figure 3: (a) Data collection (b) Data analysis Iteration 0.

Keywording has become within the GSMS a constant comparison of data and abstractions through the open, selective, and theoretical coding. The keywording tries to find a classification schema in a GSMS, but this must arise from previously established knowledge existing in the literature. For the initial list of topics in open coding, we started from the two concepts used to conduct the search phase: “software development” and “grounded theory”. These concepts were used because both are consolidated in the scientific literature. The open coding of keywording resulted in two lists, one for Software Development terms obtained from Swebok v3.0 (that is the body of knowledge for Software Development)(Bourque and Fairley, 2014) and the other corresponding to the variants of Grounded Theory. (Urquhart, 2001). The application of selective coding produced categories. These categories were the lists that emerge from Software Development and Grounded Theory. In the case of Software Development, the list contains the main 10 SD processes obtained from Swebok v3.0. The Grounded Theory list has only two elements representing the GT variants: Glaserian Grounded or Classical Grounded, and Straussian Grounded or Evolved Grounded.

Finally, theoretical coding seeks to establish propositions. These propositions, that we call pre-classification schema are more abstract and deeper levels of the categories obtained in selective coding, prove that they encompass several concepts and are a more specific, refined proposition and an evolution in **Software Development** and **Grounded Theory**. Unfortunately, in iteration 0, these lists by themselves cannot generate relevant categories and propositions.

Mapping starts with its open coding and is based on the elements of the Pre-classification schema, assigning a code to each item in both lists. Later, the selected articles were cataloged within the codes of the pre-classification schema. Selective coding in mapping is interested in summarizing the number of articles that can be cataloged within the two lists of the Pre-classification schema.

Finally, we can code 41 articles corresponding to (58,5%) in the scope of software development and 34 articles corresponding to 48.57%, in grounded theory. Theoretical Coding requires establishing propositions based on the results previously obtained. It analyzed the mapping results about the classification of the articles. It was determined that 20 articles (28.6%) had some codes from both lists, 18 (25.7%) papers were not on either list, 19 (27.1%) were coded only with terms of the Software Development list, and 13 articles (18.6%) were coded only with terms of the grounded theory list.

The obtained **Outcomes** of this iteration 0 are:

$$
\begin{cases}
Q_i = \{\sum_{i=1}^{n} RQ_i\}; \\
\text{Where } Q_i \text{is set RQ at iteration i} \\
Q_{DC} = [RQ_1, RQ_2, RQ_3]; \\
\text{Where } Q_{DC} \text{ is set RQ from Data collection} \\
Q_0 = Q_{DC} + [RQ_4, RQ_5]; \\
Q_0 = [RQ_1, RQ_2, RQ_3, RQ_4, RQ_5]; \\
A_0 \text{ is the set of AQ of iteration 0,} \\
A_0(1) = \varnothing; \quad A_0(2) = \varnothing; \quad A_0(3) = \varnothing; \\
A_0(4) = \varnothing; \quad A_0(5) = \varnothing; \\
\text{Where } A_0(k) \text{ is the set of Answers to } RQ_k
\end{cases}
\quad (4)
$$

Synthesis and Outcomes we did not obtain enough representative results when classifying the documents. Even when the two topics were consolidated in the specialized literature, this first classification rate was too low. The synthesis highlighted the lack of proper classification of the selected works because most of our articles could not be classified. It lead us to propose two new research questions: **RQ4:** Is there a way to categorize the documents within GT in SD to increase the rate of cataloged papers? **RQ5:** Is there any other way to categorize software than the ones provided by the pre-classification schema?
None of the saturation conditions is met, therefore it is needed to move to another iteration.
Comparison with Literature: Analyzing the existing literature reviews on software engineering about the variants of GT (Kroeger et al., 2014; Stol et al., 2016), they also had a low rate of classification, probably because it has not been adequately deepened.

6 CONCLUSION

This paper represents a step forward to applying systematic mapping analysis by enriching it with grounded theory practices. The result is what we call Glaserian Systematic Mapping Study. It combines the rigour of systematic mapping and the flexibility of grounded theory. GSMS is more powerful than SMS because coding activities are conducted in each iteration, allowing new knowledge to emerge.

This publication contributes to in the following aspects: i) We have not found previous attempts to formalize the GT processes. ii) This formalization creates a new way of defining saturation through three equations in terms of research questions, their answers, and the concepts comprising the answers. iii) Research questions can be deepened as the iterations progress, thus achieving answers to deeper and more specific questions. iv) The answers that emerge as part of the iterations can be confronted by other findings compared with the literature, allowing these findings to be validated. v) Findings can be validated in each iteration according to their application through the rigor and relevance assessment stage.

GSMS incorporates the SMS scalability and the GGT systematization. In the GSMS, the SMS steps have been enriched with the data analysis tools provided by the GGT, giving more depth to the results, especially the steps of Keywording, Mapping, and Synthesis. It is also able to evaluate the scientific rigour and industrial relevance of the results across iterations.

The GSMS improved the classification rates compared to SMS. It also has the advantage of adding new research questions that arise without having to restart the research process. In our case, applying SMS, only 55.7% of the articles were classified, but applying GSMS our classification rate exceeded 80%.

REFERENCES

Adolph, S., Hall, W., and Kruchten, P. (2008). A Methodological Leg to Stand on: Lessons Learned Using Grounded Theory to Study Software Development. CASCON '08, pages 13:166–178, NY, USA. ACM.

Bourque, P. and Fairley, R. E. (2014). *Guide to the Software Engineering Body of Knowledge (SWEBOK(R)): Version 3.0.* IEEE CS Press, CA, USA, 3rd edition.

Chun Tie, Y., Birks, M., and Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE open medicine*, 7.

Crabtree, C. A., Seaman, C. B., and Norcio, A. F. (2009). Exploring language in software process elicitation: A grounded theory approach. In *ESEM 2009*.

Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative research*, 14(3):341–352.

Glaser, B. G. (1992). *Basics of grounded theory analysis*. Mill Valley, Calif. : Sociology Press.

Glaser, B. G. and Strauss, A. L. (1973). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

Ivarsson, M. and Gorschek, T. (2010). A method for evaluating rigor and industrial relevance of technology evaluations. *Empirical Software Engineering*, 16(3):365–395.

Kitchenham, B. A., Budgen, D., and Pearl Brereton, O. (2011). Using mapping studies as the basis for further research - A participant-observer case study. *Information and Software Technology*, 53(6):638–651.

Kroeger, T. A., Davidson, N. J., and Cook, S. C. (2014). Understanding the characteristics of quality for software engineering processes: A Grounded Theory investigation. *IST*, 56(2):252–271.

Parizi, R. M., Gandomani, T. J., and Nafchi, M. Z. (2014). Hidden facilitators of agile transition: Agile coaches and agile champions. In *2014 8th Malaysian Software Engineering Conf., MySEC 2014*, pages 246–250.

Paternoster, N., Giardino, C., Unterkalmsteiner, M., Gorschek, T., and Abrahamsson, P. (2014). Software development in startup companies: A systematic mapping study. *IST*, 56(10):1200–1218.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology*, 64:1–18.

Sharma, R. and Biswas, K. K. (2015). Functional Requirements Categorization Grounded Theory Approach. In *ENASE 2015*, pages 301–307.

Stol, K.-J., Ralph, P., and Fitzgerald, B. (2016). Grounded theory in software engineering research: A Critical Review and Guidelines. In *ICSE '16*, pages 120–131.

Strauss, A. and Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques. *Newbury Park, CA: Sage*.

Stray, V., Sjøberg, D. I., and Dybå, T. (2016). The daily stand-up meeting: A grounded theory study. *Journal of Systems and Software*, 114:101–124.

Urquhart, C. (2001). An encounter with grounded theory: Tackling the practical and philosophical issues. In *Qualitative research in IS: Issues and trends*, pages 104–140. IGI Global.

Van Niekerk, J. C. and Roode, J. (2009). Glaserian and Straussian Grounded Theory: Similar or Completely Different ? In *SAICSIT'09*, number 10, pages 96–103.

Zayour, I. and Hamdar, A. (2016). A qualitative study on debugging under an enterprise IDE. *Information and Software Technology*, 70:130–139.