A Systematic Mapping on Discovery and Composition Mechanisms for Systems-of-Systems

Portirio Gomes¹, Everton Cavalcante¹,², Pedro Maia¹, Thais Batista¹, Kamilla Oliveira¹

¹DIMap, Federal University of Rio Grande do Norte, Natal, Brazil
²IRISA-UMR CNRS/Université de Bretagne-Sud, Vannes, France

{enghaw13, evertonranielly, pedropetrovitch}@gmail.com, thais@ufrnet.br, kamilla_anime@hotmail.com

Abstract—Systems-of-systems (SoS) represent a class of systems resulted from the interaction among independent systems that cooperate to form a larger and more complex system aiming at accomplishing global missions. An inherent characteristic of SoS is the high heterogeneity of their constituent systems, which are distributed, independent, and developed with different technologies. In addition, SoS are highly dynamic, so that their constituents are often partially known or even unknown at design time. As a consequence, these constituent systems need to be discovered, selected, and composed at runtime towards identifying the proper arrangements that contribute to the accomplishment of the global missions of the SoS. In this paper, we present the results of a systematic mapping aimed to investigate the existing approaches to discover and compose constituent systems within an SoS. Besides providing an overview of the state of the art on these topics, we shed light on important issues to be addressed by future research towards a more effective development of SoS.

Keywords—systems-of-systems; search; discovery; composition; systematic mapping

I. INTRODUCTION

In recent years, we have witnessed an increasing interest in the research and development of systems-of-systems (SoS), systems resulted from the interaction among other distributed, heterogeneous independent systems (the so-called constituent systems) that cooperate to form a larger and more complex system towards the accomplishment of global missions [1]. Each constituent system accomplishes its individual missions and can contribute to the accomplishment of the global missions of the SoS. However, the result of such an interaction is said to be more than the sum of the constituents as it enables the SoS to offer new functionalities that cannot be provided by any of these constituent systems working alone. Moreover, SoS are inherently dynamic, i.e., they can be composed and reconfigured at runtime.

SoS have been consensually distinguished from other complex and large-scale systems due to a set of characteristics inherent to this class of systems [1, 2]: (i) operational independence of constituent systems, which provide their own functionalities even when they do not cooperate with other systems within the scope of the SoS; (ii) managerial independence of constituent systems, which can be independently managed from other systems and from the SoS; (iii) evolutionary development of the SoS, which can evolve as a response to changes related to the environment, to the constituent systems, or to its functions and purposes; (iv) emergent behavior, which enables the SoS to provide new functionalities resulted from interactions among constituent systems and that are not localized into a single constituent; and (v) geographical distribution of constituent systems, which are physically decoupled and exchange only information among them. Altogether, these distinguishing features make SoS different from monolithic systems and call for a paradigm shift when architecting these systems [3].

A particular challenge to the construction of SoS is the high heterogeneity of their constituent systems, which are distributed, independent, and developed with different technologies and for diverse platforms. For this reason, the interoperability among constituent systems within an SoS has been one of the most important topics addressed in this context as it is determinant for promoting emergent behaviors [4, 5]. Therefore, abstracting away the inherent heterogeneity of these systems and making them interoperable are imperative issues to be tackled in SoS architectures.

The high dynamicity of SoS also raises challenges related to the construction of these systems. The concrete constituent systems that can compose an SoS at runtime are often partially known or even unknown at design time. As a result, these concrete constituents must be discovered, selected, and composed at runtime in order to identify proper arrangements of these systems that contribute to the accomplishment of the global missions of the SoS based on their capabilities.

Considering the aforementioned challenges, it is necessary to understand how to discover and compose constituent systems to form an SoS. To achieve this goal, we have conducted a systematic mapping (SM) aimed to investigate the existing approaches for discovering and composing constituent systems within an SoS. In this paper, we present the results of such an SM to (i) present an overview of the state of the art of research on these topics and (ii) shed light on important issues to be addressed by future research towards a more effective development of SoS.

The remainder of this paper is organized as follows. Section II details the conducted SM. Section III presents a brief discussion built upon the selected studies and points out some issues to be addressed by future research in this context. Section IV enumerates possible threats to the validity of the conducted SM. Finally, Section V provides some concluding remarks.
II. SYSTEMATIC MAPPING

According to Kitchenham et al. [6], an SM aims to investigate the literature with broader research questions in order to get a comprehensive overview of the state of the art (or state of the practice) on a given topic. Furthermore, this type of secondary study has been recently considered as an useful way for systematically identifying relevant research gaps in the literature and collecting evidences to commission further research [7].

An SM typically follows a systematic process divided in three basic steps [8]: (i) planning, which with the research questions to be answered, search strategy, selection criteria, and data extraction and synthesis methods; (ii) conduction, in which primary studies are identified, selected, and evaluated according to the previously established protocol; and (iii) reporting (or analysis), which aggregates extracted information from the relevant primary studies considering the research questions and outlines conclusions from them. The following subsections detail the application of each of these three steps, which were conducted in December 2014 and January 2015. In this SM, only studies published until December 2014 were considered.

A. Planning

In this phase, we have established a protocol describing the research questions, the search strategy to be adopted, the criteria to be applied for selecting the primary studies, and the data extraction and synthesis methods to be used.

Research questions. Aiming at finding primary studies to understand and summarize evidences about the existing discovery and composition mechanisms for SoS, the following research questions (RQs) were proposed:

**RQ1**: Which are the existing discovery mechanisms for SoS?

The goal of this RQ is to identify the existing mechanisms used to discover constituent systems, which can be further composed to form an SoS.

**RQ2**: How constituent systems can be composed to form an SoS?

The goal of this RQ is to identify systematic approaches for composing constituent systems (possibly after their discovery) to form an SoS.

Search strategy. In order to retrieve primary studies, we have used an automated search process performed over in five electronic databases (see Table I), which are among the most popular ones in Computer Science and Engineering and are able to ensure a high coverage of potentially relevant studies [9, 10]. Based on the defined RQs, four main keywords were initially identified, namely systems-of-systems, search, discovery, and composition. In addition, possible variations such as synonyms and singular/plural forms were considered, thus resulting in the following search string:

(system-of-systems OR systems-of-systems OR systems of systems OR systems of systems) AND (search OR discovery OR composition)

in which the main keywords were connected by using the AND logical operator and the possible variations and synonyms by using the OR logical operator.

Inclusion and exclusion criteria. Selection criteria were used to evaluate each retrieved primary study according to the defined RQs. The main goal was to include studies that are potentially relevant to answer the RQs and to exclude the ones that do not contribute to answer them.

The considered inclusion criteria (ICs) were:
IC1: The study presents and/or discusses scenarios, research challenges, or opportunities on discovery mechanisms to support the composition of systems within an SoS.
IC2: The study presents a mechanism for discovering constituent systems in an SoS.
IC3: The study presents an approach to compose constituent systems to form an SoS.

The established exclusion criteria (ECs) were:
EC1: The study is not directly related to SoS.
EC2: The study does not present discovery mechanisms in the SoS context.
EC3: The study does not present the composition process of systems in an SoS.
EC4: The study is a previous version of a more complete paper about the same research.
EC5: The study does not have an abstract or the full text is not available.
EC6: The study is a table of contents, foreword, tutorial, editorial, or summary of conference.
EC7: The study is not written in English, which is the most common language in scientific papers.

<table>
<thead>
<tr>
<th>Database</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEEXplore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td><a href="http://dl.acm.org">http://dl.acm.org</a></td>
</tr>
<tr>
<td>ScienceDirect.com</td>
<td><a href="http://www.sciencedirect.com">http://www.sciencedirect.com</a></td>
</tr>
<tr>
<td>Scopus</td>
<td><a href="http://www.scopus.com">http://www.scopus.com</a></td>
</tr>
<tr>
<td>Web of Science</td>
<td><a href="http://www.webofknowledge.com">http://www.webofknowledge.com</a></td>
</tr>
</tbody>
</table>

Data extraction and synthesis methods. In order to extract data from the selected primary studies, data extraction spreadsheets related to each RQ and other relevant information were built to synthesize the results and support drawing of conclusions. These data were independently extracted by the researchers regarding each RQ.

B. Conduction

In this phase, the primary studies were searched, selected, and evaluated according to the established protocol, resulting
in a set of possibly relevant studies. During the search process, the generic search string has undergone minor changes in order to make it compatible with the specificities of each electronic database engine. Afterwards, the automated search procedure was performed over the selected electronic databases by searching for all studies that have matched the adapted search string. The automated search was limited only to title, abstract, and keywords fields.

The selection of the studies was divided in two main phases. The first phase encompassed reading title, abstract, and keywords of the studies retrieved from the electronic databases, whereas the second phase encompassed the full reading of the filtered studies. In order to minimize the effect of any bias or misinterpretation, the selection involved four researchers that have individually performed the selection activities. Moreover, each study was evaluated twice, each time by a different researcher.

After retrieving the studies from the electronic databases, studies indexed by more than one database were removed and then the researchers have performed the first selection filter studies based on the selection criteria (ICs and ECs) against the information available in title, abstract, and keywords. Next, an agreement meeting was conducted in order to compare the results and solve existing conflicts, thus resulting in a consensual preliminary selection. Afterwards, the full text of the filtered studies was read and the selection criteria were applied again. Nine studies were considered as relevant to this SM and selected for data extraction (see Table II).

Table III summarizes the number of primary studies retrieved from the databases, the number of studies selected as relevant for the SM, as well as the precision\(^1\) and index\(^2\) rates [20] regarding each electronic database. Scopus has indexed the majority of the relevant studies (8/9) although IEEEXplore and Web of Science have also presented good rates since they have indexed almost half of these studies (4/9). Nevertheless, all sources have presented low precision rates as we have selected a small amount of primary studies compared with the total of retrieved studies (i.e., 9/259 = 3.47%). This fact can indicate that the string used in the automated search could be refined in future revisions of this work aiming at increasing its accuracy.

C. Reporting

In this phase, we have summarized the results of the SM considering the RQs and the extracted/synthesized data. We herein present a general panorama regarding the included studies and answers to each RQ based on their analysis.

\(^1\)Precision rate is the ratio between the number of relevant studies retrieved from a database and the total of studies retrieved from it.

\(^2\)Index rate is the ratio between the number of relevant studies retrieved from a database and the total of relevant studies.
<table>
<thead>
<tr>
<th>Source</th>
<th>Number of studies Retrieved</th>
<th>Number of studies Selected</th>
<th>Precision rate</th>
<th>Index rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEEExplore</td>
<td>118</td>
<td>4</td>
<td>3.39%</td>
<td>44.4%</td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td>14</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>ScienceDirect.com</td>
<td>13</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Scopus</td>
<td>196</td>
<td>8</td>
<td>4.08%</td>
<td>88.9%</td>
</tr>
<tr>
<td>Web of Science</td>
<td>79</td>
<td>4</td>
<td>5.06%</td>
<td>44.4%</td>
</tr>
</tbody>
</table>

Demographic data. As SoS represent a relatively new research field, it is interesting to observe when studies addressing discovery and composition within SoS were published. Fig. 2 shows the distribution of the selected primary studies per publication year. Most studies (7/9) have been published in the last six years, thus indicating a recent interest of the scientific community related to these topics.

We have also observed the venues where the selected studies were published. We have noticed that 77.78% (i.e., 7/9) of these studies were published in conference proceedings, thus possibly indicating that they require a stronger validation/evaluation. It is also important to highlight that almost half of the selected studies has appeared in three venues where a significant number of works related to SoS have been published, namely the IEEE International Conference on Systems of Systems Engineering – SoSE (S2 and S9), the Annual IEEE International Systems Conference – SysCon (S6), and the recent International Journal of Systems of Systems Engineering – IJSSE (S5).

Fig. 3 shows the application domains addressed by the SoS presented in the selected primary studies. We have noticed that a third of these studies use SoS inserted into the military domain (3/9). Indeed, most of the SoS found in the general literature are in such a domain, where the SoS concept started to gain popularity as a strategy to deliver new capabilities resulting from the collaborative work of existing complex systems [1]. Other domains such as smart grid, health care, and transportation have been found in the selected studies since they are scenarios where SoS have also been developed to address important challenges such as criticality and complexity of software systems. Finally, another third of the studies has not specified the target application domain. For this reason, there are no evidences to state that the approaches proposed in these studies are generic enough and hence can be applied to any type of SoS and/or application domain.

RQ1 – Discovery mechanisms for SoS. We have noticed that two studies (S1 and S2) suggest using a registry (or lookup) service responsible for registering information about the constituent systems available at the environment and facilitating their discovery. The main concern is to query this registry for the capabilities offered by constituent systems and provide information about how to interact with each constituent system, for example. It is noteworthy to mention that this approach is rooted upon service-oriented architectures (SOA), which have been identified as a possible standard architectural style for SoS [5]. In SOA, (i) service providers publish descriptions of the services that they provide in a registry, (ii) clients search for services by querying this registry, (iii) the registry provides information about the required service to clients, and (iv) clients and service providers are bound. According to studies S1 and S2, a registry within an SoS would follow the same principles in terms of managing information published by constituent systems with respect to the capabilities that they provide. In this perspective, other constituent systems or another entity (e.g., a discoverer) could query the registry to further enable the constituent systems to interact with each other. Nevertheless, these studies do not provide details on how to represent such an information about the capabilities of constituent systems to be made available at the registry.

Another important approach proposed by studies S4 and S5 is related to the use of semantic techniques aiming at contributing to the automation of the discovery and composition processes while eliminating ambiguities and improving accuracy. These studies assume that each available constituent system uses an specific ontology to semantically enrich information about it, e.g., by defining a vocabulary of the concepts related to its capabilities (functionalities). However, due to the inherent heterogeneity of constituent systems and the existing differences between the concepts defined in each specific ontology, the approaches introduced in these two studies adopt a shared, common ontology to which each ontology regarding a constituent system is mapped. Additionally, semantic Web services based on the
defined ontologies are used to describe capabilities, non-functional properties, and interfaces of constituent systems.

On the other hand, study \textit{S7} proposes a decentralized approach in which constituent systems (called federates) are peers and have knowledge about each other in terms of location, e.g., network address to connect with. Although this approach seems to be advantageous in that it does not require a single registry responsible for managing all information about the available constituent systems, it is limited with respect to dynamism. As a consequence, a given constituent system needs to know information such as Internet address, interfaces, required parameters, etc. to communicate with another constituent system. This condition is not well suited for the highly dynamic environment of an SoS, in which constituents may unpredictably enter into and leave.

**RQ2 – Composition of constituent systems within SoS.** We have identified two main ways approaches to address the composition of constituent systems. First, studies \textit{S2} and \textit{S4} are concerned with the use of mediators, which would be responsible for transmitting (and translating) data between constituent systems, as well as performing negotiations to enable these systems to collaborate with each other. Considering the spontaneous interactions among constituent systems, which are dynamically discovered and integrated into an SoS, study \textit{S4} argues that mediators cannot be specified or implemented at design time, i.e., they have to be synthesized and deployed at runtime in order to cope with the highly dynamic environment of an SoS. Furthermore, mediators can drive the interactions among constituent systems by using obligations. As proposed in study \textit{S2}, obligations specify the constraints to be fulfilled by constituent systems in order to allow their proper interaction and composition within the SoS. In study \textit{S2}, an obligation is expressed by a reference to a constituent system (i.e., such an obligation is assigned to the system) and a set of actions that it performs, thus describing its capabilities. These descriptions of obligations are managed by the registry entity, which uses them to search for constituent systems able to fulfill the obligations.

Second, studies \textit{S6}, \textit{S8} and \textit{S9} present different approaches to compose constituent systems. The authors of study \textit{S6} employ genetic algorithms to determine the best possible way of interconnecting constituent systems in order to form the SoS, i.e., experts identify the constituent systems and their possible interconnections, and the algorithm determines an architecture with the best interconnections between such systems. In turn, the authors of study \textit{S8} propose an approach in which constituent systems and their behavior are modeled as bigraphs, a formalism to describe interactions among concurrent systems. Alternatively, the authors of study \textit{S9} use a model-based approach composed of two phases, namely (i) system design, in which constituent systems and their behavior are modeled, and (ii) system assembly, in which such systems are selected according to the needed behavior to form an SoS. Although these three studies are different by the nature of the approaches that they propose, we have observed some similarities with respect to: (i) the formal definition of constituent systems; (ii) explicit specification of the possible interactions among systems; and (iii) interactions handled at a higher, conceptual level, not at the implementation level. Nevertheless, these approaches require statically defining these interactions, a condition that may not hold in the dynamic environment of an SoS since it may be impracticable or even impossible to define all possible interactions at design time.

**III. Discussion**

After performing the SM, we have noticed that (i) the approaches for discovery and composition in the SoS context are incipient, and (ii) the literature does not present specific studies dedicated to these topics yet. In the following, we shed light on some important issues identified from the studies analyzed in this SM to be addressed in future research on SoS.

**Use of capabilities for discovering and composing constituent systems within SoS.** In the systems engineering context, a capability has been commonly understood as the ability of a system to provide a given functionality by performing a set of tasks [3]. In this perspective, capabilities can play an important role in discovery and composition processes within SoS. For instance, such capabilities can be used to determine which and how these systems should interact within the scope of the SoS in order to accomplish its missions. As an example, the approach proposed by Iacobucci [21] considers a set of required capabilities and defines systems that are candidate to meet such capabilities, so that the composition of these different constituent systems can form alternative architectures for the SoS. Nonetheless, despite some of the studies analyzed in this SM highlight the use of capabilities as means of enabling the discovery process, they lack of details on how describing such capabilities. Therefore, it is also important to provide means of modeling capabilities of systems aiming at properly reflecting the functionalities that they offer.

**Use of service-orientation principles in SoS.** A recent SM about SoS software architectures [5] has identified that SOA seems to go in the direction of becoming a standard architectural style for SoS due to a considerable number of studies exploring the use of these architectures. As an example, Simanta et al. [22] state that the existing approaches and techniques developed to support service identification, publication, and discovery in SOA can support SoS due to to the existing similarities between these types of systems. In addition, this architectural style can facilitate the integration of constituent systems within an SoS as well as deal with inherent heterogeneity, distribution, and interoperability issues. Nevertheless, we argue that SOA cannot be used as-is for architecting SoS. As reported in Section II-C, two studies identified in our SM (\textit{S1} and \textit{S2}) have proposed using a
registry service for storing information about the constituent systems available in the environment and querying it for the capabilities offered by constituent systems as well as for information about how to interact with each constituent system, for example. However, we believe that the inherent complexity of SoS hampers the use of such a registry entity as it was originally proposed in SOA. This registry should be improved in order to accommodate not only information regarding capabilities of constituent systems, but also about other important aspects to be considered when integrating constituent systems to form an SoS, e.g., missions, priorities, constraints, etc.

**Understanding the influence of the different categories of SoS on discovery and composition.** In the literature, SoS have been broadly categorized into four different classes [3]: (i) directed, in which the SoS is controlled by a central entity and the behavior of its constituent systems is subordinated to such a central control and its purposes; (ii) acknowledged, in which goals, resources, and central control of the SoS are all recognized, but constituent systems retain their independent management and their behavior is not subordinated to the central managed purpose; (iii) collaborative, in which constituent systems voluntarily collaborate in a greater or lesser degree in order to address shared, common interests; and (iv) virtual, in which there is no central control and universal purposes and the constituent systems are completely unaware of their participation within the SoS. Almost all analyzed studies consider that constituent systems are somehow aware about their participation in the SoS. However, the proposed approaches may not apply to the other categories or even a single approach may be not enough to address all of them. For instance, static discovery mechanisms and a centralized registry are solutions that can meet the characteristics of directed SoS, in which constituent systems and interactions among them are well defined and the SoS has control over such systems despite their operational and managerial independences. On the other end of the spectrum, the existence of a centralized or even distributed registry would not be adequate for virtual SoS, in which constituent systems are completely unaware of their participation within the SoS and their capabilities are discovery at runtime.

**Use of semantic-based techniques for integrating constituent systems into an SoS.** Studies S4 and S5 advocate the use of ontologies to semantically enrich information about constituent systems mainly in terms of their capabilities. By using an ontology, a given constituent system can be described in an unambiguous way, thus avoiding misinterpretation about the provided functionalities while improving accuracy and automation of discovery and composition processes. However, the operational and managerial independences of constituent systems pose a challenge on the use of ontology-based techniques as each of these systems may have their own ontology and respective meaningful information. Therefore, addressing this heterogeneity is important as it can contribute to the achievement of a full interoperability among constituent systems [23]. Moreover, additional studies are necessary to verify the applicability of languages and methods currently used to specify ontologies and compose semantic-based services in the SoS context.

**Use of mediators as interaction elements within an SoS.** In SoS, mediators can be used to specify and guide the interactions among constituents systems as well as to promote interoperability among such systems within the SoS. Mediators not only transmit data (as connectors in traditional systems), but they are also able to process such exchanged data and to coordinate the interactions among constituent systems. Unlike constituent systems, mediators are under control of the SoS. Therefore, it can create, remove or change mediators at runtime, thus being able to rearrange how constituents can interact among each other. However, as discussed in study S4, these mediators need to be synthesized and deployed at runtime in order to cope with the high dynamic environment of SoS. In addition, they can be enriched with the specification of obligations (as proposed in S2) as constraints to be fulfilled by constituent systems in order to allow their interaction within the SoS.

**Addressing emergent behavior.** As previously mentioned, one of the main features that makes SoS distinct from other classes of systems is the so-called emergent behavior. In fact, an SoS relies on emergent behaviors to achieve its purposes as they enable it to provide new functionalities resulting from the collaborative interactions among its constituent systems. However, this brings up two important issues. First, emergent behaviors are associated to the interactions among constituent systems, i.e., they only appear when such systems interact within the scope of the SoS. Some emergent behaviors can be either foreseen, i.e., they can be determined by specifying interactions among constituent systems or representing interaction patterns (the ways in which they interact), or unforeseen, i.e., they dynamically appear in the context of the SoS. On the other hand, both foreseen and unforeseen emergent behaviors may be desirable or even undesirable, so that the result of the interactions among constituent systems within an SoS can be respectively positive or negative over its operation [24]. Consequently, any approach comprising the composition of constituent systems towards forming an SoS must (i) maximize desirable behaviors, thus fostering the accomplishment of the missions of the SoS through the interactions among its constituent systems, and (ii) minimize undesirable behaviors, which may affect the accomplishment of the missions of the SoS and have a negative impact on important quality attributes, such as performance, security, and reliability. Nonetheless, in spite of their importance, emergent behaviors currently represent one of the most difficult open research issues in SoS and have been recurrently neglected by most of the existing approaches in this context.
IV. Threats to Validity

The conducted SM and its results may have been affected by some threats to validity. In the following, we discuss some of these limitations.

**Incompleteness of study search.** The completeness of this SM may have been affected by missing relevant studies. In order to reduce this threat, we have used electronic databases (see Table I) that are among the most relevant available sources in Computer Science and Engineering [8, 10]. However, there are still limitations. First, some studies may have been missed due to technical limitations of the automated search engines, an issue that is out of our control. In addition, we have not performed a snowballing [25], an useful technique that consists of checking the reference lists of the read studies aiming to find additional studies that were not retrieved in the automated search procedure. Therefore, other possibly relevant studies could have been identified and considered in this SM.

**Number of selected primary studies.** As reported in Section II-B, we have selected only nine studies as relevant to this SM. In spite of being a threat to the validity of our study, this small number of selected studies might be due to the fact that SoS represent a relatively new research field and hence more research efforts on the investigated topics are necessary. At the same time, such a number may indicate that the performed search process had limitations in terms of retrieving all relevant primary studies. Nonetheless, we believe that the findings coming from the analysis of these nine studies are still able to provide a reasonable overview about the state of the art and drive further research on discovery and composition of constituent systems in SoS.

**Bias on study selection.** In order to make the results of this SM reproducible, the protocol presented in Section II-A clearly established the search terms used in the automated search, the search sources that were considered, and the criteria used to select the primary studies. However, different researchers tend to have different understandings on these criteria, so that the results of the study selection performed by different researchers are likely to be varied. Although the drawn conclusions may have been influenced by the researchers’ opinions, we have attempted to mitigate the effect of any personal bias or misinterpretation by adopting a multiple-revision strategy. Therefore, all studies were evaluated twice, each time by a different researcher in both selection phases (see Section II-B).

**Inaccuracy of data extraction.** Bias on data extraction may result in inaccuracy of the extracted data items, thus affecting the analysis of the selected studies. We have attempted to reduce this bias by clearly defining the data items outlined in the data extraction spreadsheets. In addition, the data items to be extracted in this SM were discussed among the researchers and agreed upon their meaning.

**Bias on data synthesis.** Not all studies sufficiently and clearly describe the details of information to be extracted as data items aiming at supporting the answers to the defined RQs. Therefore, we have had to infer certain pieces of information regarding data items during data synthesis. In order to minimize the inaccuracy of such inferences, we have conducted discussions aiming at solving any disagreement and clarifying potential ambiguities.

**Quality of the selected studies.** One of the reliable mechanisms of increasing the level of confidence in the findings of secondary studies is to establish some criteria for assessing the quality of the selected primary studies as means of ensuring that the collected evidences are relevant and present scientific value [26, 27]. Nevertheless, quality assessment is not an essential task in an SM as the goal of this type of secondary study is to provide a broad overview of the studied research topic and hence all primary studies related to it should be selected [8]. For this reason, such a procedure was not applied over the selected studies. However, a future revision of this SM can encompass quality assessment to reflect the validity of primary studies and increase the reliability of the results of the SM itself.

V. Conclusion

The construction of SoS faces new, important challenges coming from the distinguishing characteristics of this emergent class of systems. The high dynamicity of SoS and the inherent heterogeneity of their constituent systems call for systematic approaches to make such constituents interoperable while abstracting away their heterogeneity, as well as to dynamically discover and compose them at runtime towards the accomplishment of the global missions of the SoS. In this context, we have conducted an SM aiming at investigating the existing mechanisms to discover and compose constituent systems within an SoS. As main findings coming from the nine analyzed studies, we have concluded that (i) the discovery and composition approaches in the SoS context are still incipient and (ii) the literature does not present specific studies dedicated to these topics. Moreover, we have noticed that most studies attempt to take advantage of existing approaches proposed out of the SoS context, in particular the adoption of service-orientation principles and semantic-based techniques. Nevertheless, important concerns pertaining to SoS have still been neglected, such as: (i) how to effectively model capabilities of constituent systems in order to reflect the functionalities that they provide and allow their discovery and further composition; (ii) how the different categories of SoS (directed, acknowledged, collaborative, virtual) influence the discovery and composition processes; and (iii) how to consider emergent behaviors (desirable and undesirable ones) resulted from interactions among constituent systems. Therefore, we believe that the panorama presented in this paper can open perspectives to new, important issues to be addressed in future research, thus contributing to a more effective development of SoS.
ACKNOWLEDGMENT

This work was partially supported by INES – Brazilian National Institute of Science and Technology for Software Engineering (http://www.ines.org.br), funded by CNPq under grant 573964/2008-4.

REFERENCES