A Systematic Literature Review of Geospatial Web Service Composition

Roberto dos Santos Rocha, Lívia Castro Degrossi, and João Porto de Albuquerque

University of São Paulo
Institute of Mathematical and Computing Sciences, São Carlos – SP, Brazil
rsrocha@usp.br
{degrossi,jporto}@icmc.usp.br

Abstract. Context: Geospatial Web services (GWSs) can be viewed as modular Web applications that provide services or geospatial data, information or knowledge. Combining GWSs and generating composite web services out of atomic services can provide users with a higher level of functionality. Aim: The aim of this study is to obtain an overview of research on geospatial web services composition through a systematic literature review (SLR). Method: A SLR was performed by formulating four research questions to evaluate the selected works. In total, 121 papers were accepted into the final set. Results: The most widely used GWS composition model is the syntactic followed by semantic service composition. Considering that geospatial domain is dynamic, it is necessary to understand the challenges and limitations of the semantic model in depth. Conclusions: There are a significant number of research projects being conducted in this specific Web service composition area targeting the geospatial domain. Our findings identified the need to improve the robustness of the empirical evaluation of existing research.

Keywords: Systematic Review, Secondary Study, Geospatial Services Chaining, Geospatial Services Orchestration, Web Services, Disaster Management, WPS, OGC.

1 Introduction

Over time, the amount and availability of geographic information has grown exponentially. This generates a new challenge, because it is necessary to integrate different information sources – authoritative data (e.g. sensor data) and volunteered geographic information (e.g. georeferenced social media messages) – in order to improve the situational awareness and support decision making [1].

Thus, and a new geospatial cyberinfrastructure (GCI) is needed to process and integrate geospatial information [17,19]. In this context, GCI can be improved by applying Service-oriented Architecture (SOA) and Web services technology that are also capable of processing and integrating geospatial information to tailor the results to the needs of the user.

The Web service composition confers the ability to compose services by manually or automatically generating a service composition plan [11]. Combining Geospatial Web Services (GWS) and generating composite web services out of atomic services can provide users with a higher level of functionality.
There are several different composition strategies based on existing composition platforms and frameworks. Several authors have provided comparative and critical reviews of Web service composition, including Mathkour et al [11]. In early works on GWS composition, there was an attempt to apply the same strategies in the geospatial domain. However, some of these studies failed to meet their requirements because geospatial Web service composition is different due to the following characteristics [7, 19]: (i) it consists of light-weight protocols, (ii) it has crowdsourcing capability, and mainly, (iii) it has the capability to process real-time geospatial data.

In an effort to understand the dynamics of the operation of a geospatial Web service composition, a systematic literature review (SLR) was conducted. The main contribution of this paper is that it presents the results of this investigation so that researchers interested in this area may have a broader picture of it and some important questions may be answered.

The remainder of this paper is structured as follows: Section 2 presents the basic concepts that are used in the rest of the paper. Section 3 reviews some studies similar to the SLR presented herein. Section 4 describes the methodology applied to conduct this SLR. Section 5 reviews the results and presents a discussion of the results. Finally, Section 6 concludes the paper and presents a discussion of the threats to the validity of this SLR.

2 Background

In this section, the basic concepts related to Web service composition (Section 2.1) and geospatial Web services composition (Section 2.2), which are used in the rest of the paper, are presented.

2.1 Web Service Composition

Web service composition provides the ability to develop services by manually or automatically generating a service composition plan to achieve a business goal, resolve a scientific problem or provide new service functionality [11].

There are several different composition strategies based on some existing composition platforms and frameworks. Many authors have provided comparative and critical reviews of Web service composition, including ter Beek [2], Dustdar et al [5], Gao et al [8], Mathkour et al [11], among others. Approaches developed in the literature for services composition can be classified into two categories [11, 18]: syntactic and semantic. In the next subsections, we will detail each category.

Syntactic Web service composition model

The syntactic model is an XML-based Web service composition that provides a stateless model of “atomic” synchronous or asynchronous message exchanges to allow business partners and their processes to interact with each other [11].

In the syntactic model of Web service composition, two approaches have been proposed [11]:
- **Orchestration Approach:** in this approach, there is a central coordinator (which may be another Web service) that is responsible for taking control of the Web services involved in composition and execution of their operations.
- **Choreography Approach:** in this approach, each Web service participates in a collaborative environment and knows at precisely what time it should execute its own operations and with whom to interact; i.e., a central coordinator that controls the involved Web services, and execution of their operations is not necessary.

**Semantic Web service composition model**

The semantic model is an ontology-based Web service composition that was proposed to support service composition in the semantic Web. Its goal is enabling applications to use annotations and appropriate inference engines to automate all stages of the Web service lifecycle [11].

In the semantic model of Web service composition, three approaches have been proposed [11]:

- **Uninformed Approach:** this approach comprises an uninformed search algorithm that does not take advantage of or utilize any information other than the goal predicates;
- **Informed Heuristic Approach:** in this approach, heuristic value is defined and used to choose which nodes in the search space will be expanded at the next step and which nodes will be ignored. In other words, this heuristic value requires selecting the best nodes from the search space in each step;
- **Genetic Approach:** this approach defines a genetic algorithm, the population-based meta-heuristic optimization algorithm, which uses four biologically inspired techniques, namely, mutation operation, crossover operation, natural selection operation, and fitness function.

**2.2 Geospatial Web Services Composition**

Generally, a geospatial Web service (GWS) can be viewed as a modular Web application that provides services regarding geospatial data, information, or knowledge [13]. As a Web service, a GWS involves three parts [13]: a user entity (consumer), a provider entity (provider), and a register entity (broker). Similar to a traditional Web service, the basic operations during the lifecycle of a GWS include publication, discovery, binding, invoking, and execution.

The heterogeneity of the data must be overcome for geospatial data to be accessed. Standardized open protocols and interfaces allow access to distributed and diverse data in a common way. The Open Geospatial Consortium (OGC) is a step toward interoperability of geospatial data; it develops protocols and open specifications that allow access to distributed and diverse data through interfaces [12, 19]. Some examples of OGC standards are a Web Map Service (WMS) to transfer maps; a Web Feature Service (WFS) for transferring geospatial data; and a Web Processing Service (WPS) for defining a set of standard interfaces to facilitate the publication and access of geospatial processes over the network [12, 16, 19]. WPS is an important component of GWS composition because it manages the processing and analysis of spatial data.
GWSs provide users with the capability of generating complex functionalities by composing atomic GWSs, and these new composite functionalities can also be published over the network as new web services, restarting the lifecycle of a GWS [6].

In previous works on GWS composition, there have been attempts to apply the same strategies in the geospatial domain. However, some failed to meet their requirements because of geospatial Web service composition demands, mainly the following characteristics [7, 19]:

- **Interoperability**: geoprocessing services with syntactic and semantic descriptions are machine-to-machine as well as machine-to-human discoverable and executable through standard protocols. Clients can thus perform both single and sequence geoprocessing functions compiled as workflows.
- **Lightweight protocols**: being lightweight, the protocols are easily adopted in existing and new applications.
- **Collaboration/Crowdsourcing**: data, information and knowledge from different users or communities can be explored and integrated to promote geoprocessing functionalities in an open environment.
- **Real-time**: the capability to process nearly real-time geospatial data sources that are provided by sensors, which produce large volumes of geospatial data.

3 Related Works

Despite the existence of several works on geospatial web service composition, no previous systematic reviews of the works covering this theme have been identified. On the other hand, in terms of the overall scope of Web service composition, there are some studies that have, to some degree, analyzed different specific aspects. These works are summarized in sequence:

- **Dustdar et al [5]**: this paper discusses the urgent need for service composition including the technologies required to perform service composition. Based on some currently existing composition platforms and frameworks, the authors define five categories of service composition: (i) static and dynamic composition (corresponding to design time and runtime composition); (ii) model-driven service composition; (iii) declarative service composition; (iv) automated and manual composition; (v) context-based service discovery and composition.

- **Mathkour et al [11]**: this paper introduces a taxonomy of Web service composition and provides a survey of Web service composition models and their approaches. Additionally, the authors present comparisons of different models and approaches for each model. This taxonomy is presented in Section 2.

- **Zhao et al [19]**: this paper provides a comprehensive overview of the state-of-the-art architecture and technologies, as well as the most recent developments in the Geoprocessing Web (GW). GW is a broader topic that covers aspects of geospatial Web services composition. In addition, this paper tackles the achievements and challenges of the GW with respect to data, service, workflows and semantics.
All of the works presented herein are examples of traditional literature reviews related exclusively to traditional web service composition, except for the study by Zhao et al. [19] that deals with the Geoprocessing Web. Therefore, to cover what could be considered a gap in the literature, this paper presents the results of a systematic literature review related to GWS composition. Our work is different from others because it uses a replicable, scientific and transparent methodology for identifying different approaches related to geospatial service composition.

4 Method

We planned, conducted, and reported the review results by following guidelines proposed by Kitchenham and Charters [10]. According to these guidelines, a systematic literature review (SLR) includes several steps, which can be grouped into three main phases: (i) planning the SLR, (ii) conducting the SLR, and (iii) reporting the SLR. The details of planning and conducting the review are described in Sections 4.1 and 4.2, respectively. In Section 5, we present the results of the review.

4.1 Systematic Review Planning

The following steps are included in planning the SLR [10]: (i) identification of the need for a SLR, (ii) commissioning the SLR, (iii) specifying the research questions, (iv) developing a SLR protocol, and (v) evaluating the SLR protocol. In the following, steps (i), (iii) and (iv) are detailed because they are considered the main and mandatory steps of planning the SLR.

Need for a Systematic Literature Review

The motivation for conducting this systematic review was the perceived need to systematically extract, synthesize, and critically analyze the literature on GWS composition with respect to evaluating the published approaches. The SLR method was chosen for this research because it uses a replicable, scientific and transparent approach to avoid bias [3].

In addition, no SLR of this topic has been published. Therefore, we consider this to be an important and useful undertaking and consequently wish to share it with other potentially interested researchers. Thus, other research groups can start their work from this review, saving research time and resources to focus on proposals and evaluations of new approaches, using the existing works as a foundation.

Research Questions

Based on the research objectives in Section 1, our research was guided by the following four questions:

- **RQ1.** What is the composition model used?
- **RQ2.** In which application domain was the research carried out?
- **RQ3.** In which application domain was WPS used in the primary works?
- **RQ4.** What research methods are used in the primary studies?
Search and Selection Strategies

Table 1 presents the search string used to find papers related to this literature review. The keywords in the search string were derived from research questions RQ1, RQ2, and RQ3. This search string was submitted to the search mechanisms of the Scopus and IEEE Xplore databases. The search string was applied not only to the title field but also to the abstract and keywords fields to maximize the number of papers.

It is important to highlight that Scopus includes registers of the main conferences and journals published by digital libraries such as Springer, ACM, and IEEE Xplore, as well as several other important indexed vehicles. On the other hand, this characteristic may cause an overlapping of studies, since were used other database covered by Scopus, i.e., the IEEE Xplorer database. However, we decided to include it because the search engines of these databases are organized around completely different models [4]. Consequently, it is needs to search many different electronic sources, since no single source finds all of the primary studies [4,14].

Table 1. Search string

<table>
<thead>
<tr>
<th>Search string</th>
</tr>
</thead>
<tbody>
<tr>
<td>(“web service*” OR “workflow*” OR “service chain*” OR “web service orchestration” OR “orchestration” OR “composition” OR “web processing service” OR “web map service*” OR wps OR wms) AND (“geospatial*” OR “geoprocessing” OR “geoservice*” OR “spatial data infrastructure*” OR sdi)</td>
</tr>
</tbody>
</table>

To select only the list of papers appropriate to this SLR, the following inclusion and exclusion criteria were applied:

Inclusion Criteria (IC):
For inclusion, only peer-reviewed primary studies would be considered. In addition, the items were required meet both of the following criteria:

- **IC-1**: primary study whose content is truly related to Web service composition was included.
- **IC-2**: primary study whose content is truly related to the geospatial domain was included.

Exclusion Criteria (EC):
The following EC were used to guarantee that the papers were truly appropriate for the context of this SLR:

- **EC-1**: the primary study is not electronically available on the web, i.e., paper is not available for free download through institutional subscriptions that the authors currently have.
- **EC-2**: the primary study is not presented entirely in English.
- **EC-3**: the data register identified after applying the search string does not actually refer to a scientific paper, but to some non-peer reviewed publication, such as technical reports, books and book chapters, proceedings prefaces, and journal editorials.
4.2 Systematic Review Conduction

This section presents the main steps for conducting the SLR. The steps are grouped into two parts: (i) identification and selection of the primary studies and (ii) data extraction and synthesis of the primary studies.

Identification, selection of primary studies

Identification and selection of the primary studies were based on the adapted strategy proposed by Rocha and Fantinato [14], which consists of the three steps shown in Figure 1 and is described as follows.

![Search Strategy Diagram]

Fig. 1. Search Strategy.

The search process was performed in electronic databases with the goal of identifying papers that are important for this study (Step 1). A search string (Table 1) was adjusted and applied to the search engines of each one of the used data sources. A set of papers in a special issue focusing on “Towards a Geoprocessing Web” of the Computers & Geosciences Journal was used as a mechanism for evaluating the search string. The searches returned 1,499 primary studies, when considering the sum of the two data sources and eliminating duplicate results. This is considered a large number, resulting from the strategy to maximize the chances of finding candidate primary studies as explained in the SLR planning.

The search was applied by the first time in April 2014. A very few specific recent works published in 2014 were not considered since only works regarding to complete years were considered in this study scope, i.e., until 2013.

Table 2. Electronic databases used

<table>
<thead>
<tr>
<th>Source</th>
<th>URL</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Xplore</td>
<td><a href="http://www.ieeexplore.ieee.org">www.ieeexplore.ieee.org</a></td>
<td>269</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Scopus</td>
<td><a href="http://www.scopus.com">www.scopus.com</a></td>
<td>1,230</td>
<td>195</td>
<td>119</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>1,499</td>
<td>198</td>
<td>121</td>
</tr>
</tbody>
</table>
We first assessed the title and abstract and applied the inclusion and exclusion criteria to each study and thus obtained a set of relevant papers (Step 2). Finally, the 198 studies were read in full, and the inclusion and exclusion criteria were applied, which resulted in the inclusion of 121 studies (Step 3). Table 2 shows these data in more detail. The Scopus source had the highest return, 119 primary studies, corresponding to 98% of the selected studies.

Data extraction and synthesis

A set of common attributes was used to extract data from each primary study; these attributes included the title, author(s), publication year, and venue, among others. In addition to these common attributes, a series of other specific attributes was used during the paper analysis and filling in the form; these specific attributes were used to extract data by taking into account each one of the different research questions, for which specific extraction strategies were also established, as shown in Table 3.

Moreover, the following OGC domains were used to categorize the primary studies in terms of their application domain (RQ2 and RQ3): Aviation, Built Environment & 3D, Business Intelligence, Defence & Intelligence (D&I), Emergency Response & Disaster Management, Energy & Utilities, Geosciences & Environment, Government & Spatial Data Infrastructure, Mobile Internet & Location Services, Sensor Webs, and University and Research.

Table 3. Attributes collected for data extraction

<table>
<thead>
<tr>
<th>RQ</th>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>Basic info</td>
<td>Title, authors, publication year, and venue (journal or conference proceedings).</td>
</tr>
<tr>
<td>RQ1</td>
<td>Web service composition</td>
<td>The Web service composition model of the primary study based on the taxonomy proposed by Mathkour et al [11].</td>
</tr>
<tr>
<td></td>
<td>model</td>
<td></td>
</tr>
<tr>
<td>RQ2</td>
<td>Application domain</td>
<td>Which OGC-based application domain is the primary study based on?</td>
</tr>
<tr>
<td>RQ3</td>
<td>OGC WPS standard</td>
<td>Does the primary study use an OGC interface standard?</td>
</tr>
<tr>
<td>RQ4</td>
<td>Research method</td>
<td>The empirical research methods/data analysis techniques used in the selected study based on the taxonomy proposed by Hevner et al [9].</td>
</tr>
</tbody>
</table>

5 Results

This section presents the results and analyses produced by conducting this SLR according to the protocol presented in Section 4. As a result of implementing the protocol, we selected 121 primary studies, 40 from journals and 81 from
conference proceedings. Due to space limitations, a complete list of literature used in this SLR can be found at: http://selected-papers.sourceforge.net/

Figure 2 shows the temporal distribution of the primary studies (from 2005 to 2013), separated by publication type, i.e., conference or journal. It is noteworthy that only 33% were published in journals and 67% were published in conference proceedings. These are indications that the Geospatial Web Service Composition area is still at an early stage of research and development and under a maturity search process. Moreover, Figure 2 shows that the largest number of publications, including journal and conference proceedings, was recorded in 2012. In addition, 53% of the publications were from in the last four years 2011 to 2013.

Furthermore, an additional analysis was performed for the specific event and journal in which each primary study was published in order to verify which vehicles, due to their nature, could be causing a greater number of works related to geospatial Web service composition. According to the data analysed, Geoinformatics (22% of conference proceedings) and IGARSS\(^1\) (14% of conference proceedings) were considered the most representative vehicles for publishing works related to GWS composition.

On the other hand, the most representative journals, in descending order by number of publications, are Computer & Geosciences (30% of journals), Transactions on GIS (12% of journals), and Computer and Environment Urban Systems (10% of journals). These three represent 52% of all journal publications.

The following sections show the structured results specifically related to the responses of the research questions.

5.1 RQ1 – What is the most widely used composition model?

As shown in Figure 3, the most widely used geospatial Web service composition model is the syntactic model, which accounts for 51% of all selected papers, followed by the semantic model, which accounts for 40%.

The use of the syntactic model for GWS composition is directly related to the characteristics of the composition model. According to Mathkour et al [11], the syntactic Web service composition model is a static model, which is generally

\(^1\) International Geoscience and Remote Sensing Symposium
intended for use in industry, while the semantic Web service composition model is a dynamic model generally used by the academic community. Moreover, the semantic model requires full automation. This automation must address one problem [11]: the gap between concepts used by humans and the data interpreted by the computer.

Concerning the syntactic model of GWS composition, the orchestration approach is the most widely used. It is carried out by an orchestration engine (OE), which coordinates the interactive services involved and is controlled by a document containing chaining instructions in a certain description language. The business process execution language (BPEL) is provided by a ‘de facto’ web service orchestration (WSO) standard. However, Stollberg and Zipf [15] noted several problems with the use of BPEL in combination with current WSO versions, for example, communication protocols and the transfer of raw binary data.

The studies that used semantic services composition model were classified as “Semantic Web Service Composition Model”. No further classification was established because several of these works did not provide a detailed description of their approaches (i.e., uniformed approach, informed heuristic approach or genetic approach). Considering that geospatial domain is dynamic, it is necessary to understand the challenges and limitations of the semantic model. The semantic model aims to automate all phases of the life cycle of the composition of geospatial services. Thus this model can provide the flexibility to this domain.

Moreover, we found four works that combine syntactic and semantic service descriptions for chaining geographic services. In addition, we found three works that could be classified as artificial intelligence (A.I.) models, i.e., A.I. planning-based and fuzzy models. These papers present a solution for the automatic composition of geospatial web services that relies on semantic web services technology and A.I. planning techniques to generate composite GWSs.
5.2 RQ2 – In which application domain was the research carried out?

As observed in Figure 4, the Emergency Response and Disaster Management (52%), Sensor Web (12%), and Agriculture (6%) domains occurred the most frequently.

According to [12], Emergency Response and Disaster Management are different domains of activity with different information sharing requirements. In both domains, however, there is a need to rapidly discover, share, integrate and apply geospatial information [12]. Thus, these domains are the most common found in papers published in journals and conference proceedings, accounting for 45% and 31% of publications, respectively.

In addition, a more depth analysis showed that events of hydrometeorological origin might have contributed to this evidence, given that the literature reveals that this kind of event represents about 30% of disasters occurring around the world.

5.3 RQ3 – In which application domain was WPS used in the primary studies?

WPS provides standardization rules for invoking geospatial processing services inputs and outputs (requests and responses). According to the data presented in Figure 5, 51% of the primary studies use the WPS interface standard. “Emergency Response & Disaster management” was the most common item found, found in 16% of the primary studies; and “Sensor Web” the second one, found in 10% – together, they represent to 26% of all studies that use the WPS standard. In addition, of the works that use the WPS interface, 70% are syntactic Web services composition models.

5.4 RQ4 – What research methods are used in the primary studies?

The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods [9]. According to the data presented in Figure 6, data extraction and synthesis found application evidence
in the primary studies selected for the thirteen evaluation methods this SLR focused on.

“Prototype” was the most common item found (35%), followed by “Scenario,” which was found in 29% of the primary studies. “Case Study” was found in 11% of the primary studies.

An individual analysis by publication vehicle shows that “Prototype” and “Scenario” are the most relevant methods for both publication vehicles, i.e., journals and conference proceedings. A prototype is a step towards an implemented working system in a similar way as a model is [9]; and scenario is a descriptive method of evaluation that should only be used for especially innovative artifacts for which other forms of evaluation may not be feasible. Considering these definitions, the results demonstrate that the approaches are not yet mature enough for application in industry because the studies are being conducted in the laboratory, by means of prototypes and scenarios or short examples.

“Case study” is indicative of the highest level of maturity because it is applied in real scenarios. Of the total case studies, 77% were applied in the last three years: 2010 to 2013. A more depth analysis showed that case study stands out as a trend for the coming years.
6 Conclusion

This paper reports on the results of a systematic literature review of Web service composition in geospatial domain. For this purpose, we followed the guidelines proposed by Kitchenham and Charters [10] and selected papers from two important data sources – Scopus and IEEE Xplore.

The findings of this study established that the scientific literature about geospatial Web service composition is growing, with a significant growth in the number of publications over the last three years (2011–2013).

Considering that geospatial domain is dynamic, it is necessary to understand the challenges and limitations of the semantic model of service composition. The semantic model aims to automate all phases of the life cycle of the composition of geospatial services.

Moreover, the union of the syntactic and semantic models using A.I. planning techniques can be a solution for the automation of the stages of the lifecycle of geospatial Web services composition, and thus should be explored.

This study showed that all selected studies present any kind of preliminary evaluation. However, this apparent benefit is diminished when we consider the low level of evidence of the proposed approaches. The most commonly employed evaluation methods are academic studies, i.e., studies that are being conducted in the laboratory, by means of prototypes, scenarios or short examples. As a means to improve accuracy and enable generalizations of the primary studies, further research into this topic should include some form of empirical assessment, such as case study.

This systematic review has some limitations. To the extent that we performed a SLR, the potential for incomplete identification of relevant studies and publication bias must be considered. We are aware that the selection process may not have captured all of the relevant studies, mainly the studies that have not been indexed. To minimize this threat, two reputable databases were used, namely Scopus and IEEE Xplore. The wide list of different publication forums returned indicates that the search coverage was sufficient. Moreover, we used a gold standard, which consists of collection papers from a special issue of the Computers & Geosciences, to evaluate our search string. Thus, we have tried as far as possible to ensure that all relevant papers were selected.

In addition, some subjective decisions may have occurred during paper selection and data extraction because some primary studies did not provide a clear description or proper objectives and results, making it difficult to objectively apply the inclusion/exclusion criteria or the impartial data extraction.

Our future work will continue to report this SLR based on an updated set of GWS composition studies and develop methodological and technical supports to effectively reuse existing empirical evidence in the GWS composition domain.

Acknowledgement. João Porto de Albuquerque is grateful for FAPESP (grant no. FAPESP 2012/18675-1), CAPES (grant no. 12065-13-7) and Heidelberg University (Excellence Initiative II / Action 7) for providing funding for his contribution to this paper.
References


