Process Integration in Semantic Enterprise Application Integration: a Systematic Mapping

Laylla D. Cerqueira¹, Julio C. Nardi², Monalessa P. Barcellos¹, Ricardo A. Falbo¹

¹ Ontology and Conceptual Modeling Research Group (NEMO), Department of Computer Science, Federal University of Espírito Santo– Vitória – ES – Brazil

² Informatics Department, Federal Institute of Espírito Santo, Colatina - ES, Brazil

{laylladuarte, monalessa, falbo}@inf.ufes.br, julionardi@ifes.edu.br

Abstract. Enterprise Application Integration (EAI) plays an important role by linking heterogeneous applications to support business processes within and across organizations. Semantic conflicts often arise in this context and have to be addressed to a successful interoperation. Besides, to properly support business processes, integration should deal with processes integration and cover the process layer. In this paper, we present a systematic mapping that investigated aspects related to EAI, particularly, the use of ontologies to address semantics in integration at process layer. The results provide a panorama of the research in this area.

1. Introduction

Organizations almost always use software applications to support business processes execution. In order to better support these processes and meet the organizations needs, applications need to be integrated. Enterprise Application Integration (EAI) is currently one of the main problems faced by organizations. More and more, applications need to work together to support complex business processes involving different business areas.

EAI at process layer, commonly referred to as business process integration, aims at creating a choreography engine that orchestrates data and message exchange between applications, resulting in a kind of workflow to better support business processes [Hanson et al. 2002]. It is very important because, in general, enterprise applications are built to address parts of business processes and they should be integrated to support the entire process or a set of related processes. Besides, process integration is fundamental to business process improvement [Berente et al. 2009].

However, the applications to be integrated are usually developed by different groups, which, many times, do not have any concern with integration. As a result, these applications, almost all, are heterogeneous, autonomous and distributed [Izza 2009]. Heterogeneity has been considered one of the most challenging issue, being the main source of semantic conflicts, which occur when applications use different meanings to the same information item, i.e., when information items seem to have the same meaning, but they do not [Watche et al. 2001]. To reduce these integration conflicts, EAI initiatives should address semantic aspects. Semantic integration, which is based on meaning, is more reliable than syntactical integration, which is based only on the processing of strings and union of schemes [Muthaiyah and Kerschberg 2008].
In semantic EAI, ontologies can be used to establish a common understanding about the domain of interest, serving as an interlingua to provide communication between applications [Calhau and Falbo 2010] and promoting integration at different applications layers (data, message/service, and process) [Nardi et al. 2013].

In this paper, we present a mapping study aiming at investigating semantic EAI initiatives that address integration at the process layer. In particular, we are interested in those initiatives that use ontologies as part of their approaches. With this systematic mapping we aim at providing an overview of this research topic considering the evidences about it in the literature [Kitchenham and Charters 2007]. This mapping study is a refinement of another one performed by Nardi et al. (2013), which investigated semantic EAI initiatives in general. We updated that study and, then, we selected the publications involving semantic EAI initiatives covering the process layer and analyzed them in more details.

This paper presents the mapping study and its main results. It is organized as follows: Section 2 provides the background for the paper, talking briefly about EAI and ontologies; Section 3 presents the research protocol used in the study; Section 4 presents the obtained results; Section 5 discusses the findings that emerge from the results; and, finally, Section 6 presents our final considerations.

2. Background

EAI is crucial for organizations, since applications increasingly need to work together to support business processes. For integrating enterprise applications, it is necessary to create a coherent information system architecture in which the various business processes, information storages and systems are integrated so that they appear seamless for the user. It is necessary, thus, to define an integrated system as a collection of subsystems that interact to form a whole, and whose properties emerge due to the interaction of its subsystems [Vernadat 2007; Pokraev 2009].

Many of these applications/(sub)systems, however, were not designed to work together. Contrariwise, they are often heterogeneous, autonomous, and distributed (HAD) applications. Heterogeneous means that each application implements its own data and process models. Autonomous means that applications may run independently of other applications. Distributed means that applications locally implement their models, which generally are not shared with other applications [Izza 2009]. This, therefore, contributes to make application integration a difficult and complex task.

EAI can be performed at different layers [Izza 2009]: data, message/service, and process. Data integration deals with moving or federating data between multiple data stores. Integration at this layer assumes bypassing the application logic and manipulating data directly in the database, through its native interface. Message (or service) integration addresses messages exchange between the integrated applications. Process integration views enterprises as a set of interrelated processes and it is responsible for handling message flows, implementing rules and defining the overall process execution [Izza 2009]. It constitutes the most complex integration approach and, according to Berent et al. (2009), different from data and message/service integration, process integration is often not explicitly defined.
Semantic conflicts can occur in any of these layers, arising whenever applications are built with different conceptualizations. To avoid them, the meaning of the interchanged information has to be understood across the systems to be integrated. In this context, the use of ontologies as an inter-lingua for explaining implicit and hidden knowledge is an useful approach to overcome these conflicts [Watche et al. 2001].

Considering their generality level, ontologies can be classified as: top-level ontologies (so-called foundational ontologies), which describe very general concepts like space, time, object, event, etc., and are independent of particular domains or tasks; domain ontologies, which describe concepts related to a generic domain (e.g., Electrocardiogram); task ontologies, which describe the conceptualization related to a generic task or process (e.g., Diagnosis); and application ontologies, which deal with concepts related to a particular application (e.g., a medical ontology for heart diseases built on Diagnosis and Electrocardiogram ontologies). Ideally, domain and task ontologies should be defined from top-level ontologies and application ontologies should be based on domain and task ontologies [Guarino 1998].

It is important to clarify some terminological aspects used along this paper. Despite the definitional differences/interrelations between integration (as the act of incorporating components into a complete set in a way to form a new system constituting a whole and creating synergy [Izza 2009]), and interoperability (as the ability of applications/components to exchange data and services preserving the constituents parts as they are [Vernadat 2007]), these terms are often used indistinctively [Nardi et al. 2013]. In this paper, therefore, the term “integration” is adopted in a broader sense, covering both integration and interoperability meanings.

Finally, we need to define what a mapping study is. It is a secondary study, i.e. a study that is based on analyzing research papers (referred to as primary studies), designed to give an overview of a research area through classification and counting contributions in relation to the categories of that classification. It makes a broad study in a topic of a specific theme and aims to identify available evidence about that topic. In order to reduce bias and ensure the study repeatability, mapping studies adopt a systematic approach [Kitchenham and Charters 2007] [Petersen et al. 2015].

The mapping study presented in this paper was performed following the process defined in [Kitchenham and Charters 2007], which includes: planning, when the research protocol is defined; conducting, when the protocol is executed and data are extracted, analyzed and recorded; and reporting, when the results are recorded and made available to potential interested parties. Next section presents the main parts of the research protocol used.

3. The Research Protocol

The goal of this mapping study is to investigate EAI initiatives considering semantic aspects and addressing the process layer. By "addressing the process layer", we mean that the we are interested in EAI initiatives in which data/service exchange is made in a way that integrates parts of a process or different process in a workflow. For achieving the goal of this study, we defined eight research questions presented in Table 1.
<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>When and in which type of vehicle have the studies been published?</td>
<td>Offer an understanding on when and in which types of vehicles the studies have been published.</td>
</tr>
<tr>
<td>RQ2</td>
<td>Which types of research have been done?</td>
<td>Identify the research type according to the classification defined by Wieringa et al. (2005).</td>
</tr>
<tr>
<td>RQ3</td>
<td>What have been the business application domains addressed in the semantic EAI initiatives?</td>
<td>Identify the business applications domains that have been supported by semantic EAI initiatives addressing the process layer.</td>
</tr>
<tr>
<td>RQ4</td>
<td>Have ontologies been adopted in the semantic EAI initiatives?</td>
<td>Investigate if ontologies have been used in semantic EAI initiatives and the purpose of using them.</td>
</tr>
<tr>
<td>RQ5</td>
<td>What kinds of ontologies (considering their generality level) have been used?</td>
<td>Identify the kinds of ontologies used in semantic EAI initiatives addressing the process layer and verify if there is predominance of some kind.</td>
</tr>
<tr>
<td>RQ6</td>
<td>Which languages/ formalisms have been used to represent the ontologies?</td>
<td>Identify how ontologies have been represented in semantic EAI initiatives.</td>
</tr>
<tr>
<td>RQ7</td>
<td>How process integration has been addressed in the semantic EAI initiatives?</td>
<td>Investigate the technological strategies and integration approaches used to perform semantic integration at the process layer.</td>
</tr>
<tr>
<td>RQ8</td>
<td>Have systematic approaches been used to conduct these semantic EAI initiatives?</td>
<td>Verify whether or not the initiatives have been followed systematic approaches to perform semantic integration at process layer.</td>
</tr>
</tbody>
</table>

Considering that the results of the study performed by Nardi et al. (2013) included publications until 2012 and informed the integration layers addressed in each publication, we decided to use the same search string used by Nardi et al. (2013), and, then, select only publications addressing the process layer. The search string has two groups of terms that were joined in a conjunction with the AND operator. The first group includes terms that aim to capture studies related to integration/interoperability of enterprise software applications. The second one aims at capturing studies that deal with semantic aspects. Within each group, the OR operator was used to allow for synonyms. The search string is:

("application integration" OR "application interoperability" OR "enterprise system integration" OR "enterprise system interoperability" OR "integration of information system" OR "interoperability of information system" OR "integration of application" OR "interoperability of application" OR "interoperability of enterprise application" OR "interoperability of enterprise system" OR "integration of enterprise application" OR "integration of enterprise system" OR "interoperability of business application" OR "interoperability of business system" OR "integration of business application" OR "integration of business system" OR "integration of heterogeneous system" OR "integration of heterogeneous application" OR "interoperability of heterogeneous system" OR "interoperability of heterogeneous application" OR "interoperability of information system" OR "integrated application" OR "interoperable application" OR
"integrated enterprise system" OR "interoperable enterprise system" OR "information system integration" OR "information system interoperability" OR "enterprise system integration" OR "enterprise system interoperability" OR "business system integration" OR "business system interoperability") AND (semantic OR semantics OR semantically).

As sources of publications, seven digital libraries were searched, namely: Scopus (www.scopus.com), Engineering Village (www.engineeringvillage.com), ACM (dl.acm.org), IEEE Xplore (ieeexplore.ieee.org), Web of Science (ISI of Knowledge) (apps.webofknowledge.com), Springer Link (link.springer.com), and ScienceDirect (www.sciencedirect.com).

Publications selection was performed in four steps. In Preliminary Selection and Cataloging (S1), the search string was applied in the search mechanism of each digital library (we limited the search scope to title, abstract and keywords metadata fields). After that, publications indexed by more than one digital library were identified and duplications were removed. In Selection of Relevant Publications – 1st filter (S2), the abstracts of the selected publications were analyzed considering the following inclusion (IC) and exclusion (EC) criteria: (IC1) the study addresses an EAI initiative that considers semantic aspects; (EC1) the publication is not written in English; (EC2) the publication does not have an abstract; (EC3) the publication is a copy or an older version of an already selected publication; (EC4) the publication is not a primary study; (EC5) the study was published only as an abstract. In Selection of Relevant Publications – 2nd filter (S3), the full text of the publications selected in S2 was read and analyzed considering the cited inclusion and exclusion criteria. Publications whose full text was not available were also excluded (EC6). As a result of S3, we updated the study reported in [Nardi et al. 2013]. Then, to focus on publications addressing EAI initiatives covering the process layer, in Selection of Relevant Publications – 3rd filter (S4), we applied an additional inclusion criterion: (IC2) The publication presents a semantic EAI initiative addressing the process layer. Thus, we were able to narrow the scope and consider only semantic EAI initiative addressing the process layer. As a result, we obtained the publications object of the study described here.

4. Data Synthesis

This study considered publications until December 31st 2015. 170 publications were obtained in S1. After duplication removal, 85 publications remained. 34 publications were selected in S2 and 19 in S3. During S4, by applying IC2 to publications selected in S3, 8 were selected. By applying IC2 to the publications selected by Nardi et al. (2013), 32 of them were selected. As a result, 40 publications addressing semantic EAI initiatives and covering the process layer are object of this study.

Figure 1 illustrates the process followed and the number of publications selected in each step. Next, a data synthesis to each research question is presented. Due to space limitation, we do not provide in this paper the list of selected publications.
RQ1. When and in which type of vehicle have the studies been published?

Figure 2 shows the distribution of studies per year. It is possible to note an increasing in 2007 and a peak in 2008. After 2008, the number of studies decreased until 2010 and remained relatively stable until 2013. None study addressing semantic EAI initiatives covering the process layer published in 2014 or 2015 was found. Regarding publication vehicle, 22 studies (55%) were published in scientific events and 18 (45%) in journals.

RQ2. Which types of research have been done?

Following the classification suggested by Wieringa et al. (2005), all the analyzed studies are Solution Proposals. 4 (10%) studies are also Evaluation Research, since they have been applied into a production environment, and 36 (90%) studies are also
Validation Research due to the use of a proof of concept, experiment, prototype or similar to evaluate the proposal.

**RQ3. What have been the business application domains addressed in EAI initiatives?**

Considering the business application domains in which semantic EAI initiatives were applied, 13 (32.5%) studies just make reference to generic scenarios (e.g., business-to-business). The other 27 (67.5%) studies presented their solution proposals in the context of specific business application domains. From these studies, 15 different categories of business application domains were identified: 4 studies are related to e-Business, 3 to Manufacturing, 2 to Telecom, 3 to Virtual Production and 5 to Product Lifecycle Management. The other 10 categories were found in one study each, namely: Aerospace, Hospital, Weather, Oil, Power Marketing, Airline, Logistics, Education, Software Engineering, and Pharmaceutical.

**RQ4. Have ontologies been adopted in the semantic EAI initiatives? If so, what is the purpose of using them?**

28 studies (70%) use ontologies as reference models to assign semantic in EAI initiatives: 10 (25%) use ontologies to assign semantic to data, 13 (32.5%) to data and service, and 5 (12.5%) to data, service and process. In one study the authors state that ontologies are used, but its use is not explained. 11 studies (27.5%) do not use ontologies. From these, one uses formal description language to address semantic aspects and the 10 remaining (25%) use other approaches (e.g., business application features).

**RQ5. What kinds of ontologies (considering their generality level) have been used?**

Table 2 presents the percentage of studies per kinds of ontologies. “Unspecified” refers to studies that use ontologies but do not specify their kinds and, thus, it is not possible to identify them.

<table>
<thead>
<tr>
<th>Kinds of ontologies</th>
<th>Studies that used(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Ontology</td>
<td>45%</td>
</tr>
<tr>
<td>Top Level and Domain Ontology</td>
<td>24%</td>
</tr>
<tr>
<td>Domain Ontology and Application Ontology</td>
<td>21%</td>
</tr>
<tr>
<td>Top Level, Domain and Application Ontology</td>
<td>3%</td>
</tr>
<tr>
<td>Unspecified</td>
<td>7%</td>
</tr>
</tbody>
</table>

**RQ6. Which languages/formalisms have been used to represent the ontologies?**

The studies adopt several languages/formalisms to represent ontologies, ranging from Semantic Web languages to more simplistic data representation techniques. The following languages/formalisms were identified: OWL (4), XML (3), WSMO (1), RDF (5), OWL-S (4), OWL and OWL-S (2), OWL-S and XML (1), XML and Topic Maps (4), Lisp, WSMO and OCML (1). Finally, one study uses a language proposed in the own study, and 3 other studies propose the use of ontologies, but do not make commitment to any specific language/formalism.

**RQ7. How process integration is addressed in the semantic EAI initiatives?**

The integration approaches presented in the analyzed studies can be categorized into design-time and run-time approaches. The first one regards integration at conceptual
level, i.e., conceptual models are used to represent/communicate the integration design. The second one refers to integration during process execution, by using a process engine, for example. The two approaches can be combined in a design&run-time approach, when the integration is addressed at conceptual level (integration models are built) and also during process execution. 6 (15%) studies use design-time approaches, 10 (25%) use run-time approaches, and 24 (60%) combine both in design&run-time approaches.

**RQ8. Have systematic approaches been used to conduct the semantic EAI initiatives?**

29 studies (72.5%) were conducted without following a systematic approach. Only 11 studies (27.5%) used approaches guiding the steps to be followed in the integration. There are initiatives that use approaches proposed in previous works, such as [Jankovic et al. 2008], which uses an approach proposed in the Athena Interoperability Framework [Berre et al. 2007]. Others propose the systematic approach used, such as [Li et al. 2009].

**5. Discussions**

This section discusses the raw data presented in the previous section.

Looking at the types of vehicles where the studies have been published (RQ1) and the types of research that have been done (RQ2), we can infer that the topic investigated in this mapping has been explored and discussed with relative degree of maturity. Usually, journals require more mature works, and the homogenous distribution of the studies between scientific events (55%) and journals (45%) can be understood as a sign of that. On the other hand, the fact that only 4 (10%) studies discuss an evaluation in a real scenario (Evaluation Research) is an indicative that the semantic EAI initiatives addressing the process layer have not yet transposed the migration barrier to practice.

With respect to the business application domains where EAI initiatives have happened (RQ3), we can notice that they are very diverse. This points out that semantic EAI at process layer is a problem that runs through several business domains.

As for systematic approaches to perform semantic integration (RQ8), we can notice that there are few works following systematic approaches for performing initiatives of integration at the process layer. Taking into account the studies that applied systematic approaches, all of them consider process models in some extent, but only two approaches ([Calhau and Falbo 2010] and [Shangguan et al. 2008]) use ontologies to address integration at the process layer. 7 of the 11 identified approaches start by doing reverse engineering of applications to be integrated and, after that, integration requirements are elicited. In these cases, applications to be integrated are previously established and the requirements are identified considering them. The other 4 approaches start with integration requirements elicitation and, then, recover models and functionalities from the applications to be integrated. In these cases, requirements are used as a basis to select the applications to be integrated and their portions to be considered. All the approaches consider the use of ontologies to assign semantics to applications elements.

Combining the findings for RQ3 and RQ8, we can conclude that we need to embark efforts to develop general systematic approaches for guiding EAI at the process layer. A systematic approach can help structuring the integration process into different
abstraction levels and define guidelines on how to perform the various integration activities. This is essential for establishing an engineering approach for EAI.

Regarding the use of ontologies (RQ4 and RQ5), semantic aspects are addressed by using them in most studies. This can be understood as an evidence of the importance of ontologies as an instrument to achieve semantic integration. There is a predominance of domain ontologies, and all the studies that use this kind of ontology apply them assigning semantics to system elements (data, services and process elements). Although ontologies are predominant for treating semantics, other kinds of model are also used, such as business application features [Kulkarni and Sreedhar 2006], service visual representation [Yeung 2011] and business process representation [Rouached et al. 2009]. Therefore, reference models are essential to address semantic EAI covering the process layer, helping to assure the appropriate communication between applications.

Despite most of studies adopt ontologies, only 5 (17.2%) use them to assign semantic to processes aspects. In [Calhau and Falbo 2010], domain ontologies are used to assign semantic to information handled by processes as inputs and outputs, but not to the processes directly. In [Alazeib et al. 2007], ontologies addressing general process concepts and specific application domain concepts are used to create a process template that serves as a reference to represent the business processes involved in the integration. In [Madhusudan 2004], a domain ontology is used to describe services and data involved in business processes. In [Minguez et al. 2011], in turn, a domain ontology provides the conceptualization used as a basis to process modelling. Finally, in [Shangguan et al. 2008] domain ontologies are used to describe services and functionalities related to the process flows. From these 5 studies, only two present domain ontologies addressing the processes involved in the integration. These results show that even in semantic EAI addressing the process layer, the use of ontologies has been focused on data and service layers, which, in a certain way, corroborates the statement of Berente et al. (2009) who say that process integration is often not explicitly defined and occurs as a consequence of data and service integration. We argue that task ontologies could be helpful to process integration, since they can be used to describe generic processes and, thus, could be applied to assign semantic to process activities, inputs and outputs. However, none of the investigated studies use task ontologies.

Regarding languages/formalisms used to represent the ontologies (RQ6), the focus has been on using machine readable languages, in particular those from the Semantic Web. 16 studies (40%) use RDF, OWL or/and OWL-S. However, there are also studies addressing integration regardless technologies. Another noteworthy aspect regards languages for web service ontologies, such as OWL-S and WSMO, which are used in 9 studies (22.5%). This reinforces the strong relation existing between integration at the process and the message/service layers.

Concerning process integration (RQ7), there is a predominance of design&run-time combined approach, indicating a concern with integration not only at implementation level but also at conceptual level. In fact, semantic aspects should be addressed since initial phases of the integration initiative. It should be assigned during the initial phases (analysis) and kept in the following ones (design and implementation) [Calhau and Falbo 2010].

Several technical strategies have been used to perform integration. Strategies based on services use technologies such as Enterprise Service Bus (ESB) and
middleware to provide tools communication through service exchange. Strategies based on process manager use a specific component (e.g., a process engine) to orchestrate service exchange in a workflow to support process execution. Strategies based on modelling, in turn, involve the use of models to represent the integration at conceptual level. All the 34 studies that use run-time or design&run-time approaches adopt strategies based on services. From these 34 studies, 14 (41%) also use process managers as components in charge of controlling service exchange in a workflow. These results show the strong relation between service integration and process integration. In fact, process integration is usually obtained from connections among services. Thus, service-oriented strategies are favorable to process integration, since services can be connected in a way that supports processes execution.

Strategies based on modelling are used in 17 (50%) studies. All the 6 studies that apply design-time approaches use models to perform conceptual integration. 11 (45.8%) of the 24 studies using design&run-time approaches adopt Model Driven Development (MDD), being 4 associated with process manager and 7 with service-based strategies. Although there is a strong relation between service and process integration, the last one occurs in a higher abstraction level. Thus, conceptual models are a suitable approach to deal with that. However, it is also necessary to address process integration at lower levels. MDD is a promising strategy to go from conceptual level to implementation level, decreasing the level of abstraction through models transformation.

6. Final Considerations

Organizations use many applications simultaneously to accomplish their business processes. A challenge for them is to integrate those applications to better support their businesses. EAI solutions can help in this task, providing a middleware for supporting and integrating business processes. An EAI solution works as a connecting interface between different applications. It has a set of functionalities gathered together in a single and complete package, which can provide better performance and business processes refinement [Al-Ghamdi and Saleem 2014].

For coping with semantic issues that arise in integration initiatives, ontologies can be used to assign meaning to the shared elements. Besides addressing semantic issues, integration initiatives should achieve integration at process layer. By doing that, processes supported by the integration solution can run in a continuous workflow that connects different parts of a process or different processes.

In this paper we presented a mapping study that investigated semantic EAI initiatives addressing the process layer. The results of the mapping provide a panorama of research related to processes integration in semantic EAI initiatives. Summarizing, semantic EAI initiatives have been used ontologies (mainly domain ontologies) to assign semantics mainly to data and services. Service-oriented solutions (such as ESB and middleware) have been applied to provide communication between applications, being associated to process managers (such as workflow engine) that orchestrate services to support the processes execution. Models have been used to support integration at conceptual level and also to create integration solutions based on model transformation (MDD).

Some gaps in the research topic investigated can be pointed out: (i) the lack of systematic approaches for guiding integration at the process layer; (ii) task ontologies
have not been used to support process integration; and (iii) lack of a general conceptualization about business processes.

Apart from the study reported in [Nardi et al. 2013], which served as starting point to our study, we did not found any other related work. There are similarities and differences between the study performed by Nard et al. (2013) and ours. As for similarities, both studies investigate semantic EAI and the ontologies use in this context. Concerning differences, we can highlight the study focus and investigated aspects. [Nardi et al. 2013] aimed to provide a panorama concerning semantic EAI in general, while our study focuses on semantic EAI addressing the process layer. Thus, in our study we analyzed in depth how integration at process layer has been addressed and how ontologies have been used to aid semantic integration in this context. We also investigated other aspects not addressed in [Nardi et al. 2013], such as research types and use of systematic approaches to conduct semantic EAI. Finally, [Nardi et al. 2013] considered studies published until 2012, and we extended the coverage to studies published until 2015.

Acknowledgement

This research is funded by the Brazilian Research Funding Agency CNPq (Processes 485368/2013-7 and 461777/2014-2) and FAPES (Process 69382549/2014).

References


Kitchenham, B., Charters, S. Guidelines for performing systematic literature reviews in software engineering. TR EBSE-2007-01, School of Computer Science and Mathematics, Keele University, 2007.


