A Panorama of the Semantic EAI Initiatives and the Adoption of Ontologies by these Initiatives

Julio Cesar Nardi[1, 2], Ricardo de Almeida Falbo[2], João Paulo A. Almeida[2]

[1] Research Group in Applied Informatics, Informatics Department,
Federal Institute of Espírito Santo, Campus Colatina, Colatina, ES, Brazil
[2] Ontology & Conceptual Modeling Research Group (NEMO), Computer Science Department,
Federal University of Espírito Santo, Campus Goiabeiras, Vitória, ES, Brazil

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

Abstract. Enterprise Application Integration (EAI) plays an important role by linking heterogeneous applications in order to support business processes within and across organizations. In this context, semantic conflicts often arise and have to be dealt with to ensure successful interoperation. In recent years, many EAI initiatives have aimed at addressing semantic interoperability challenges by employing ontologies in various ways. This paper aims to reveal, through a systematic review method, some aspects associated with semantic EAI initiatives and the adoption of ontologies by them, namely: (i) the business application domains in which these initiatives have been conducted; (ii) the focus of the initiatives regarding integration layers (data, message/service, and process); (iii) the adoption of ontologies by EAI research along the years; and (iv) the characteristics of these ontologies. We provide a panorama of these aspects and identify gaps and trends that may guide further research.

Keywords: enterprise application integration, semantics, ontology, systematic mapping.

1 Introduction

In order to be competitive and face changing economic conditions, enterprises need to be flexible and dynamic, which requires the use of information systems that can work together supporting business processes [1]. In this context, Enterprise Application Integration (EAI) plays an important role for linking separate applications into an integrated system driven by business models and the goals they implement [2].

Challenges in EAI arise, among others, from the fact that heterogeneous enterprise applications employ different data and behavioral models [3], leading to semantic conflicts. These conflicts occur whenever applications are built with different conceptualizations, which can impact the integration of data, messages/services, and processes. Despite many advances in EAI, semantic integration of enterprise applications remains a hard problem [4]. In this context, several approaches for semantic integration have been applied, using a variety of instruments, including domain vocabularies, taxonomies, ontologies, logical formalisms, and rules that specify policies, governance, etc. [3]. Among these approaches, ontologies have been acknowledged as an important means to address semantic EAI [4] [3], namely through promoting integration of different information system layers (data, message/ service, and process). In the context of semantic EAI, ontologies have been employed with the purpose of contributing to the establishment of common understanding.

This paper aims to reveal, through a systematic mapping [5], some aspects associated with semantic EAI initiatives and the adoption of ontologies by these initiatives, namely: (i) the business application domains in which the initiatives have been conducted; (ii) the focus of these initiatives regarding integration layers (data, message/service, and process); (iii) the adoption of ontologies by EAI research initiatives along the years; and (iv) the characteristics of the ontologies employed. These aspects are structured in six research questions that are investigated using 128 studies selected and analyzed according to a systematic review method.

This paper is organized as follows: Section 2 presents the main concepts used in this paper and clarify some important terminology regarding integration approaches; Section 3 presents the systematic review method adopted, and describes the main parts of the mapping protocol developed during the planning phase; Section 4 presents the results of the mapping, including the selection process, the classification schemas, and data synthesis; Section 5 discusses the findings and the mapping limitations; Section 6 presents concluding remarks and outlines further investigation.

2 Background

The various works in the literature refer to many aspects of enterprise application integration. In this section, we discuss some of the most salient concepts and terms in this broad area of research, in order to characterize the scope of our investigation and support the definition of the research questions that will be the subject of this work.

First of all, we should note that there are several definitions for the terms "integration" and "interoperability" referring to different or interrelated concepts, and these are often used indistinctively. Since we are interested in "application integration" as well as "application interoperability", we considered both terms in the searching string presented in Section 3, and throughout this paper, we use the term "integration" in a broad sense, involving both integration and interoperability.

Secondly, in the investigated literature, the distinction between intra- and inter-enterprise application integration is often present. Intra-EAI aims at integrating applications in the context of a single enterprise, while inter-EAI (also referred to as B2B integration) supports integration of applications of more than one enterprise, linked, in many cases, by a collaborative process [6]. Considering that most techniques and technologies that make up intra-EAI are also applicable to inter-EAI [6], we are interested in both intra- and inter-enterprise application integration and use "enterprise application integration" to refer to both.

Integration can concern one or several information system layers [3], such as: data layer, message/service layer, the process layer. Data layer integration concerns with moving or federating data between multiple databases, bypassing the application logic and manipulating data directly in the databases. Message/service layer integration addresses message exchange between information systems, which can occur in any tier, such as user interface, application logic or even in the data tier. Process layer integration, commonly referred to as Business Process Integration, views the enterprise as a set of interrelated processes, being responsible for handling message flows, implementing rules and defining the overall coordination of the execution.

Ontologies have been acknowledged as an important means for achieving semantic EAI [4] [3], since they aim at providing formal specifications of shared conceptualizations. Considering their level of generality, ontologies continuously range from top-level ontologies, through domain ontologies to application ontologies. Top-level ontologies (so-called foundational ontologies) describe very general concepts like space, time, object, event, etc., and are independent of particular domains or problems [7]. Domain ontologies describe concepts related to a generic domain, sometimes specializing concepts of a top-level ontology. Application ontologies, in turn, describe concepts related to a particular application [7]. Since these kinds of ontologies form a continuum, the borderline between them is not clearly defined. Thus, in this paper, we distinguish only between top-level ontologies - those developed considering theories of Formal Ontology and related areas, e.g. DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) and SUMO (Suggested Upper Merged Ontology) - and the rest (including various levels of generality usually referred as domain or application ontologies).

Finally, due to the potential of ontologies as a means to address semantic aspects, in last decades, many ontology implementation languages have been developed and many knowledge representation languages have been used for building ontologies, even they were not initially developed for this purpose [8]. So, it is important to know how ontologies have been designed and implemented in order to understand how appropriate these representations are for semantic EAI. In this context, we can cite knowledge representation languages such as first-order logic, frames and description logic. Based on them, there are some ontology languages, such as [8]: FLogic (Frame Logic), RDF (Resource Description Framework), and OWL (Web Ontology Language). Beyond these languages, ontologies are also developed using technologies associated to service description, such as OWL-S (OWL-based web service ontology) and WSMO (Web Service Modeling Ontology).

3 The Review Method and the Mapping Protocol (Planning)

This *systematic mapping* was conducted taking as basis the method for systematic literature reviews given in [5]. This method is known for its suitability for PhD studies, which is the context of this research, and the research group has expertise on it, although some limitations are known [5].

According to [5], a systematic mapping is a kind of secondary study, which offers a broad view of primary studies in a specific topic in order to identify available evidences. Thus, a *secondary study* is a study that reviews primary studies related to a set of specific research questions with the aim of integrating/synthesizing the evidences related to these research questions. The *primary study* is an empirical study investigating a specific research question.

A systematic mapping involves three phases [5]: Planning, Conducting and Reporting the mapping. *Planning* involves the pre-mapping activities, and encompasses the definition of the following items: research questions, inclusion and exclusion criteria, sources of studies, search string, and mapping procedures. These items compose the mapping protocol. *Conducting* the mapping is concerned with searching and selecting the studies, and extracting and synthesizing data from them. *Reporting* is the final phase and involves writing up the results and circulating them to potentially interested parties.

The mapping protocol is an important artifact in the review process. It is produced during the Planning phase and consumed during the other phases. The main parts of the mapping protocol used by this work are described as follows.

Research Questions. This mapping aims at answering the following research questions, considering the context of semantic EAI initiatives:

RQ1. What are the business application domains addressed?

RQ2. What is the distribution of studies according to the integration layers (data, message/service, and process layers)?

RQ3. Over the years, how wide has been the adoption of ontologies?

RQ4. What is the distribution of studies that use ontologies per integration layer?

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

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

Inclusion and Exclusion Criteria. The primary studies selection was based on the following criteria, which were organized in one inclusion criterion (IC) and four exclusion criteria (EC). The inclusion criterion is: (IC1) The study addresses enterprise application integration under a semantic perspective. The exclusion criteria are: (EC1) The study is not written in English; (EC2) The study is an older version (less updated) of another study already considered; (EC3) The study is not a primary study (which excludes short papers, editorials, and summaries of keynotes, workshops, and tutorials); (EC4) The study is just published as an abstract.

Sources. We used automatic search to collect the studies. The search was applied in seven electronic databases that were defined based on systematic reviews in the Software Engineering area. The sources are: IEEE Xplore (http://ieeexplore.ieee.org), ACM Digital Library (http://dl.acm.org), SpringerLink (http://www.springerlink.com), Thomson Reuters Web of Knowledge (http://www.isiknowledge.com), Scopus (http://www.scopus.com), Science Direct (http://www.sciencedirect.com), Compendex (http://www.engineeringvillage2.org).

Search String. In order to define the search string, we used 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" or "interoperability" of enterprise software applications. The second group aims at capturing studies that deal with semantic aspects. Within each of the groups, the "OR" operator was used to allow for synonyms. The search string, as follows, was applied in three metadata fields (title, keywords and abstract) and suffered syntactical adaptations according to particularities of each source:

("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)

Mapping Procedures (Assessments). Before conducting the mapping, we performed a pilot test of the mapping protocol over a sample consisting of 35% of the studies, which was used to evolve the components of the protocol. Considering that the review process was conducted by one of the authors, an activity of validation was carried out by a second author using a different sample of 35% of the studies. Possible biases were discussed in periodic meetings.

4 Conducting the Mapping

This section describes the main steps that were performed in the mapping, including: search and selection, data extraction and data synthesis.

4.1 Search and Selection

The search process was conducted in the beginning of 2012, and, therefore, we looked for studies published until December 31th 2011. As a result, a total of 702 records were retrieved: 107 from IEEE Xplore, 16 from Science Direct, 17 from ACM Digital Library, 56 from Thomson Reuters Web of Knowledge, 232 from Scopus, 218 from Compendex, and 56 from SpringerLink.

After the search process, the selection process was conducted progressively in five stages. In the first stage, we have eliminated duplicated studies by examining titles and abstracts. In this stage, we had the highest reduction (almost 60%), since many studies are available in more than one source. In the second stage, we have applied the inclusion and exclusion criteria considering title and abstract only (resulting in a reduction of 15.5%). Although we have used language filter mechanisms on the source's search engines, some studies not written in English have been retrieved. Thus, we have also applied EC1 criteria in this stage. The resulting set of studies was refined in a third stage, which also considered the whole text (resulting in a reduction of 44.8%). After preliminary analysis, we noticed that only three studies published before 2001 remained in the end of the third stage (one published in 1993 and two published in 1995). Indeed, they did not characterized representative points of our sample, thus, in the fourth stage, we have eliminated these three studies and defined the lower boundary date as January 1st 2001. In the fifth stage we eliminated the fours studies for which we had no access to the full text.

Table 1 summarizes the stages and their results, showing the progressive reduction of the number of studies throughout the selection process (from 702 to 128 studies, with a reduction rate of about 81.7%).

Table 1. Results of the selection process stages.

Stage	Criteria	Analyzed Content	Initial N. of Studies	Final N. of Studies	Reduction per stage (%)
1st Stage	Eliminating duplications	Title and abstract	702	290	58.6%
2nd Stage	IC1, EC1, EC2, EC3 and EC4	Title and abstract	290	245	15.5%
3rd Stage	IC1, EC2, EC3 and EC4	Whole text	245	135	44.8%
4th Stage	Studies published before 2001	---	135	132	2.2%
5th Stage	Studies not accessed	---	132	128	3.0%

4.2 Classification Schema and Data Extraction

Before data extraction, we defined categories for classifying the studies according to the research questions, as follows.

Classification schema concerning integration focus. This schema is based on [3] and encompasses three categories: Integration at data layer, Integration at message/service layer, and Integration at process layer. So, depending on the focus of the integration approach, the study is classified as one of these layers or any combination of them.

Classification schema for kinds of ontology. This schema encompasses two categories: Top-level ontology and Low-level ontology. According to the generality level of the ontologies, discussed in Section 2, a study is classified as using a Top-level ontology if a foundational ontology is used. On other hand, a study is classified as using a Low-level ontology, if a domain or application ontology is used. A study can be classified in both categories if it employs both top- and low-level ontologies.

Other classification schemes. Concerning the categories for business application domains and ontology languages, we collected unstructured data without a pre-defined classification (the categories were only defined during data analysis), in order to deal with the large variety of possibilities. In order to collect data about business application domains, we looked for use cases, examples used for describing the proposed solutions, domains that motivated research initiatives, and so on. Regarding ontology languages, we looked for the formalisms used to represent ontologies, such as OWL, OWL-S, first-order logic, among others. After that, during data synthesis, we analyzed the content and defined the categories. This process was iterative, and the resulting categories were evaluated in periodic meetings. This process involved five steps: (1) analyzing content; (2) defining categories; (3) evaluating categories; (4) classifying studies; and (5) evaluating the classification schema.

The data extraction process consisted in analyzing and collecting data of each selected study, and organizing them in a data collection form, shown in Table 2.

Table 2. Data collection form.

Field	Description	Classification schema
ID	Unique identifier	Not applicable
Bibliographic reference	Authors, title, conference or journal, and publication year	Not applicable
Business application domain(s)	Business application domains where study was applied	Not defined a priori
Integration focus	The integration layer(s) which is(are) the focus of the study	[Integration at data layer, Integration at message/service layer, or Integration at process layer]
Kind(s) of ontologies	Kind(s) of ontologies used in the study	[Top-level ontologies, or Low-level ontologies]
Ontology language(s)	Languages/formalisms used to implement/create ontologies	Not defined a priori

4.3 Data Synthesis and Results

Semantic EAI Efforts over the Years. In order to offer a general view about the efforts in semantic EAI area, we present in Fig. 1, a distribution of the selected studies (128) per published year. We can note a growth in the number of published studies from 2001 to 2008, which is characterized by two moments of relative stabilization: from 2001 to 2003, and from 2004 to 2006. After 2008, when we have observed the largest number of published studies, the number of studies decreased until 2010 and remained stable in 2011.

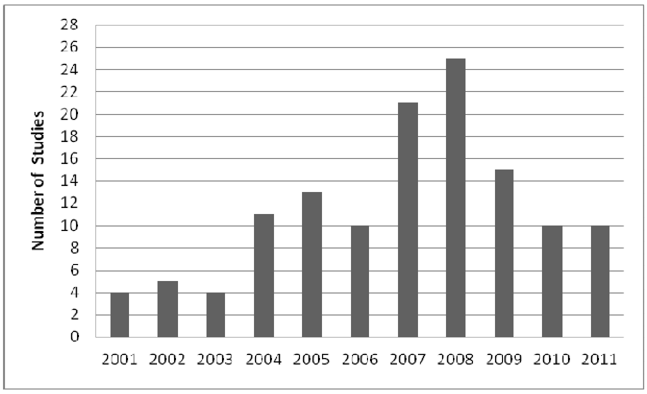

Fig. 1. Distribution of the selected studies over the years

Business Application Domains in Semantic EAI (RQ1). Considering the business application domains in which semantic EAI initiatives were applied, we identified that about 76.6% of the studies presented their solution approaches in the context of specific business application domains. The other 23.4% of the studies were classified as "General", since they just make reference to generic scenarios like "business-to-business", "e-commerce", "business", etc. Considering the approaches that were developed in the context of specific application domains, we have identified 19 categories of business application domains, which are presented in Fig. 2 together with the percentage of studies per category. The "Other" category was introduced to group business application domains that had no representative occurrence (only one paper), such as: Aerospace, Importing and Exporting, Content Publishing, Video Mail System and Software Engineering.

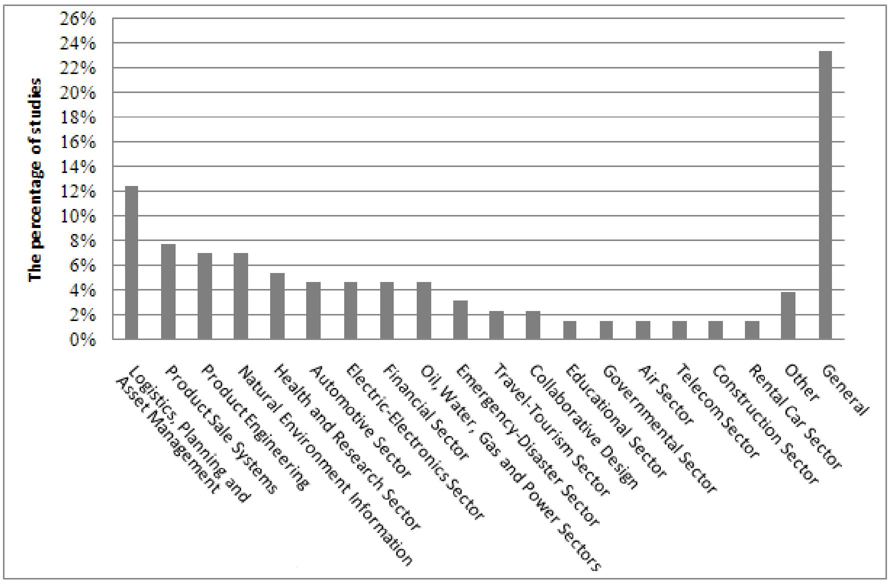

Fig. 2. The percentage of the selected studies per business application domains

Considering the distribution of studies per specific business application domain, we can notice that the "Logistics, Planning and Asset Management" domain has the largest representativeness (12.5%). It stands out, mainly because it involves supply chain initiatives, being characterized by intensive interaction between suppliers and consumers. Besides that, business application domains with representativeness between 7.8% to 5.5% include: "Product Sale Systems" (purchase order in general, and online shopping), "Product Engineering" (industrial automation technology, which requires integration and management of product life-cycle), "Natural Environment Information" (initiatives about geographic location, geographic information systems, meteorological and oceanographic information), and "Health and Research Sector" (pharmaceutical industry, health care, bio-informatics and research organizations). The other categories, although with smaller percentage of studies, still represent important numbers, if we consider that almost 23.4% of the selected studies do not make reference to any specific application domain (General).

Focus on the Integration Layers (RQ2). The studies were classified as promoting semantic EAI on data layer, message/service layer, process layer, or any combination of them. The Fig. 3 presents the percentage of studies per integration layer.

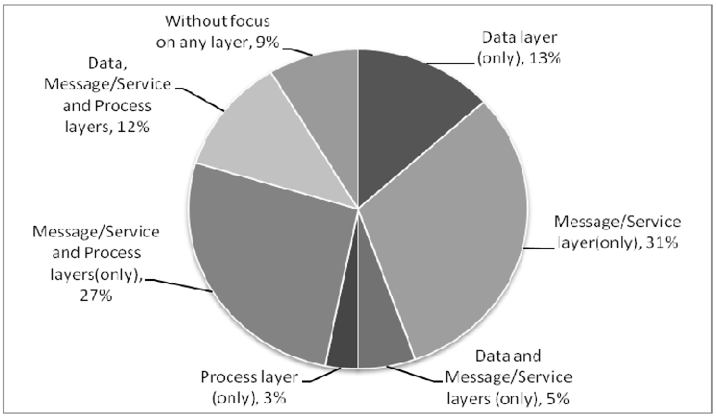

Fig. 3. Distribution of the selected studies per the focus on the integration layers

Some studies focus only on one layer: data layer (13%), message/service layer (31%), and process layer (3%). Others propose integration solutions by addressing two integration layers: data and message/service layers (5%), and message/service and process layers (27%). And, finally, there are studies that address the three layers: data, message/service and process layers (12%). Finally, when considered in isolation or when considered in tandem with other layers, the data layer is addressed by 30% of the studies, the message/service layer is addressed by 75% of studies, and process layer by 42% of them (again either solely or in tandem with other layers).

The studies that address data and message/service layers together are characterized by approaches that define data source integration solutions besides considering direct interactions (by message, service, etc.) among applications. The studies that address message/service layer together with process layer presents initiatives related to service orchestration, workflow definition, as well as business process-driven enterprise application integration initiatives. In this way, the studies that establish integration on data, message/service, and process layers together are characterized by proposing architectures, frameworks and integration approaches related to business process-driven enterprise application integration. The proposed solutions range from data source integration to application interaction driven by business processes. In this context, it is important to remark that no study focused on data and process layers without considering the message/service layer, which reflects the mediation role that the message/service layer plays.

During data extraction phase, we noted that some studies presented generic approaches, which did not make commitments to any integration layer, being classified as “Without focus on any layer” (9%). These studies are characterized by proposing conceptual or generic solutions, like reference models, standards, and metamodels, as well as technical guidance and recommendations, methodologies and life-cycle models, without focusing on any specific integration layer.

Ontologies in Semantic EAI: Adoption over the years (RQ3, RQ4), Kinds (RQ5), and Languages/Formalisms (RQ6). The adoption of ontologies in order to promote semantic EAI has grown over the years, as we can see in Fig. 4. The period from 2001 to 2003 reflects the initial phase of adoption, when the number of studies that did not use ontologies was greater or equal than the number of studies that used ontologies. From 2004, on the other hand, and, mainly, from 2007, the use of ontology became the principal means to promote semantic EAI, achieving more than 70% of the studies. Also, the set of all studies that use ontology represents about 71.8% of all the selected studies, indicating a high level of adoption. Petri nets, UML (Unified Modeling Language) models, standards for data exchange, formal languages for event composition, concept hierarchy, etc., were some of the other techniques used for addressing semantics in EAI. These techniques were used in the 28.2% studies that did not use ontologies, although some have appeared in studies that used ontologies.

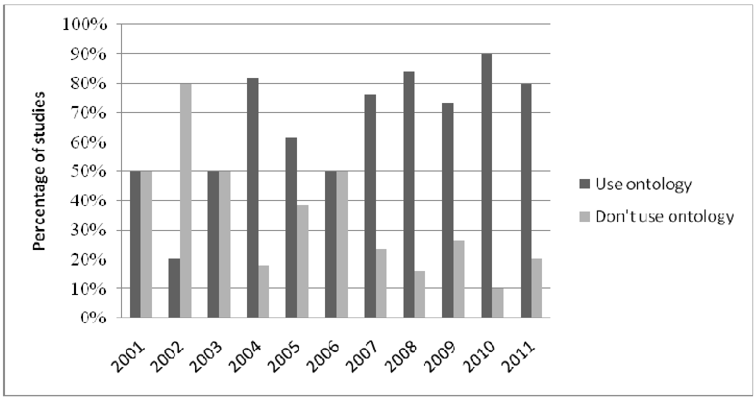

Fig. 4. Adoption of ontologies in semantic EAI along the years

Table 3 presents the percentage of studies that use ontologies per integration layer, and the numbers reflect some equivalence. However, we have two exceptions: (i) none (0%) of the studies that focus only on Process layer uses ontology; and (ii) there is a balance regarding the use of ontologies in studies that do not focus on any layer.

Table 3. Percentage of studies that use ontology per integration layers.

Integration layer	**Studies that use ontology (%)**
Data layer (only)	71%
Message/Service layer (only)	75%
Process layer (only)	0%
Data and Message/Service layers	86%
Message/Service and Process layers	76%
Data, Message/Service, and Process layers	87%
Without focus on any layer	45%

Besides analyzing the adoption of ontologies along the years, we aimed at identifying the kinds of ontologies that have been used. We identified 5 studies that use Top-level ontologies, which represent 5.4% of the studies that use ontologies. Table 4 presents these studies and the respective top-level ontologies they use.

Table 4. Studies that use top-level ontologies.

Study	Publication year	Top-layer ontology
[9]	2006	PSL (Process Specification Language) Ontology
[10]	2007	DOLCE – SUMO alignment
[11]	2007	DOLCE – SUMO alignment
[12]	2010	DOLCE
[13]	2011	DOLCE

The various studies claim to represent ontologies using a variety of formalisms and techniques, ranging from Semantic Web languages to more simplistic data representation techniques. Based on this aspect, we identified ten categories: "OWL", "RDF and RDFS", "XML", "OIL, DAML and DAML+OIL", "OWL-S", "WSMO", "Knowledge Representation", "Own language", "Other", and "None".

The first six categories refer directly to a specific technology. The "Knowledge Representation" category represents languages or formalisms associated to knowledge representation languages (Description logic, First-order logic, Frames, etc.) and graphical representations such as UML and Conceptual Maps, among others. The "Own language" category represents languages or formalisms that were proposed in the context of the corresponding work itself. The "Other" category groups technologies that did not appear in a representative number (three studies or less), including KIF, F-Logic, OCML, Common Lisp, Relational database schema and RDF4S. The "None" category groups studies that only propose the use of ontologies, but do not make commitment to any specific language/formalism. The Fig. 5 presents the percentage of studies per category (a study can fit in more than one category).

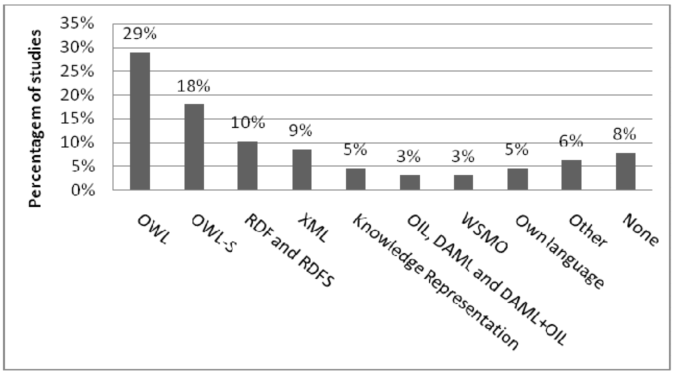

Fig. 5. The percentage of studies per category of ontology languages

We can notice a trend in using Semantic Web technologies, mainly OWL (29%), OWL-S (18%), and RDF/RDF-S (10%). Concerning ontology-based languages for service description, OWL-S (18%) and WSMO (3%) stand out. Despite that WSMO can be used in association with OWL, the largest number of studies used OWL-S instead of WSMO due to a closer relation between OWL and OWL-S.

The other categories do not represent, individually, a high number of studies. However they reflect a diversity of ontology representation languages used in the semantic EAI initiatives. It is worthwhile to point out that 8% of the studies do not address any aspect of formalization/implementation, i.e., they just suggest the use of ontologies by proposing general architectures, life-cycle models, guidelines, etc.

5 Discussion

Based on results presented in the previous section, in this section, we discuss some important findings and limitations of this mapping.

Semantic EAI Efforts over the Years. We consider that the distribution of studies along the years reflects the research efforts in semantic EAI, which suffer influence of the adoption of semantic technologies, mainly ontologies. In our view, the chart shown in Fig. 1 can be analyzed roughly according to the Gartner Hype Cycles [14]. The period between 2001 and 2003 corresponds to the "Technology Trigger" phase. The year of 2008 corresponds to the "Peak of Inflated Expectations". The years of 2009 and 2010 correspond to the "Trough of Disillusionment". The lack of change from 2010 to 2011 suggests that we are aimed towards the remaining phases: "Slope of Enlightenment" and "Plateau of Productivity".

Business Application Domains in Semantic EAI. The identified diversity of business application domains reflects the coverage of the EAI research area, and, therefore, its relevance. Moreover, we notice that, although traditional business application domains are still the most exploited, EAI initiatives span several niche application domains although in lower rate, characterizing a Long Tail-like [15] distribution (cf. Fig. 2). The domain of "Logistics, Planning and Asset Management" has had the largest representativeness, possibly due to the focus on integration that drives this kind of business, which is founded on interoperation in supply chains.

Focus on the Integration Layers. We have observed a predominant number of studies addressing the message/service layer. We believe that this can be justified by the role that functionalities (represented by the message/service layer) play in order to promote the link between data sources and business processes, and the increasing interest in service-oriented architectures in the past decade. We have observed that many of the integration solutions at the message/service layer also consider process technology, which has been seen as a clear trend in EAI. Furthermore, we have observed a low number of studies that focus only on the process layer (3%), suggesting that process layer integration depends on message/service layer integration. Moreover, a considerable number of studies (44%) focus on more than one layer, indicating that integration initiatives have established relations between integration layers to achieve interoperability.

Ontologies in Semantic EAI. We have observed that, in the past decade ontologies have become predominant in the semantic approaches to EAI. Ontologies have been used by the solution approaches in order to achieve integration through the various integration layers (data, message/service and process). Regarding the languages and formalisms used to build ontologies in the context of EAI initiatives, we have observed a predominance of Semantic Web languages, leading to ontologies which should be characterized as lightweight ontologies [16]. We have also noted that a number of *data representation techniques* have been referred to by the studies as ontology representation techniques, indicating a rather permissive use of the term ontology in the literature and a wide variation in what is considered an ontology. Considering the kinds ontologies employed, we can conclude that the use of top-level ontologies in EAI initiatives is relatively underexplored. Nevertheless, these ontologies have gained some attention in the latest years (see Table 4).

Limitations of this Mapping. Due to the fact that some stages were performed by only one of the authors, some subjectivity may have been introduced. To reduce this subjectivity, a second author was responsible for defining a random sample (about 35% of the studies) and performing the same stages. The results of each reviewer were then compared in order to detect possible bias. Moreover, terminological problems in the search strings may have led to missing some primary studies. Thus, we performed simulations in the selected databases and included a large number of synonyms in the search string. We decided not to search specific (non-indexed) conference proceedings, journals, or the grey literature (technical reports and works in progress), having worked with studies indexed by the selected electronic databases only. The exclusion of these other sources makes the mapping more repeatable, but with the consequence that we cannot rule out that some valuable studies may have been excluded from our analysis. Finally, the classification of studies regarding their focus on data, message/service and process layers is not straightforward, due to variety of possible approaches and irregularity of use of terminology in the literature. For achieving a more consistent analysis, some studies classifications were discussed in meetings. Thus, we cannot ensure that the results concerning the layers are fully repeatable, due to some level of subjectivity in this classification.

6 Conclusions

This paper presented a systematic mapping in the context of semantic EAI. Six research questions were defined and addressed investigating the following aspects: (i) business application domains in semantic EAI initiatives; (ii) focus on the various integration layers; and (iii) the adoption of ontologies in semantic EAI.

The contributions of this work are on making evident some aspects associated to semantic EAI research efforts that can drive future research. In this context, we highlight the following conclusions: (i) Most studies in semantic EAI (75%) address message/service layer integration; (ii) Ontologies have became predominant in semantic approaches to EAI; (iii) Semantic Web technologies have been widely adopted by semantic EAI efforts (with OWL being the most common language for ontology representation in the sampled studies); and (iv) The use of top-level (foundational) ontologies, although not expressive yet, has emerged as a new trend in the second half of the period investigated.

As future work, we plan to perform deepen our analysis on the use of ontologies in semantic EAI. In particular, we intend to explore how ontologies have been used in semantic EAI, focusing on the role of ontologies in the integration approach. Further, we intend to investigate how the languages/formalisms used to represent ontologies influence the integration solutions.

Acknowledgments. This research is funded by the Brazilian Research Funding Agencies FAPES (Grant 52272362/11), CNPq (Grants 483383/2010-4 and 310634/2011-3) and PRONEX (Grant 52272362/2011).

References

1. Vernadat, F.B.: Interoperable enterprise systems: Principles, concepts, and methods. Annual Reviews in Control. 31, 137–145 (2007).
2. Gacitua-Decar, V., Pahl, C.: Ontology-based Patterns for the Integration of Business Processes and Enterprise Application Architectures. In: Mentzas, G. et al. (ed.) Semantic Enterprise Application Integration for Business Processes: Service-Oriented Frameworks. IGI Pub (2009).
3. Izza, S.: Integration of industrial information systems: from syntactic to semantic integration approaches. Enterprise Information Systems. 3, 1–57 (2009).
4. Bussler, C.: The Role of Semantic Web Technology in Enterprise Application Integration. IEEE Bulletin of the Technical Committee on Data Engineering. 26, 62–68 (2003).
5. Kitchenham, B., Charters, S.: Guidelines for performing Systematic Literature Reviews in Software Engineering (Version 2.3) - EBSE Technical Report. EBSE-2007-01. (2007).
6. Endrei, M., Alari, G., Ambati, A., Krogdahl, P., Martin, C.: Patterns: Serial Process Flows for Intra- and Inter-enterprise. IBM Redbooks (2004).
7. Guarino, N.: Formal Ontology and Information Systems. Formal Ontology and Information Systems. 3, (1998).
8. Gòmez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag (2004).
9. Martín-Recuerda, F.: Application integration using Conceptual Spaces (CSpaces). Lecture Notes in Computer Science (LNCS). 4185, 234–248 (2006).
10. Alazeib, A., Balogh, A., Bauer, M., Bouras, A., Friesen, A., Gouvas, P., Mentzas, G., Pace, A.: Towards semantically-assisted design of collaborative business processes in EAI scenarios. 5th IEEE International Conference on Industrial Informatics. pp. 779–784. , Vienna-Austria (2007).
11. Bouras, A., Gouvas, P., Kourtesis, D., Mentzas, G.: Semantic integration of business applications across collaborative value networks. International Federation for Information Processing (IFIP). 243, 539–546 (2007).
12. Paulheim, H., Probst, F.: Application integration on the user interface level: An ontology-based approach. Data & Knowledge Engineering. 69, 1103–1116 (2010).
13. Treiblmayr, M., Scheider, S., Krüger, A., Von der Linden, M.: Integrating GI with non-GI services – showcasing interoperability in a heterogeneous service-oriented architecture. GeoInformatica. 16, 207–220 (2011).
14. Gartner Hype Cycles, http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp.
15. Levine, D.M., Stephan, D., Krehbiel, T.C., Berenson, M.L.: Statistics for Managers using Microsoft Excel. Prentice Hall (2002).
16. Guizzardi, G.: On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models. In: Vasilecas, O., Edler, J., and Caplinskas, A. (eds.) Frontiers in Artificial Intelligence and Applications, Databases and Information Systems IV. pp. 18–39. IOS Press, Amsterdã (2007).