A BUSINESS ONTOLOGY FOR SUPPORTING CROSS BORDER COOPERATION BETWEEN EUROPEAN CHAMBERS OF COMMERCE

Alessandro D’Atri, Research Centre on Information Systems (CeRSI), Luiss University, IT
datri@luiss.it

Antonio De Nicola, Research Centre on Information Systems (CeRSI), Luiss University, IT / LEKS, IASI- CNR, IT
denicola@iasi.cnr.it

Oleg Missikoff, Research Centre on Information Systems (CeRSI), Luiss University, IT
omissikoff@luiss.it

Abstract

The recent EU enlargement opens up new opportunities, but poses new issues to be addressed. In particular, in order to enable and support cooperation between firms from different countries, it is necessary to address interoperability issues. The LD-CAST project aims at enabling cross border cooperation between European chambers of commerce (CCs) for supporting the development of private company initiatives. The project objective is to build a European network of portals that will enable end users (mainly private companies) to access in a seamless mode services provided by public organizations registered in each portal. This paper briefly presents a cooperation framework for semantic interoperability mainly based on the following semantic technologies: ontology management, semantic annotation, and semantic search and discovery. Finally the business ontology produced in the course of the project is presented.

Keywords: Ontology, Interoperability, Semantic Annotation.

1 INTRODUCTION

The purpose of this paper is to describe the LDCAST Business Ontology that is the precondition of a cooperation framework for enabling semantic and process interoperability among institutions and enterprises located in old and newly accessed European countries. Besides technological issues, as it has been highlighted by Parissa Haghirian, cultural differences can indeed strongly influence the knowledge transfer process, sometimes even between different units of a single multinational corporation [Haghirian 2003].

More specifically, the aim is to present an ontology-based solution to support cooperative business processes, developed in the context of LD-CAST\(^1\).

The objective of the LD-CAST project is to address the semantic interoperability requirements emerged from a study on European Chambers of Commerce. In particular, it aims at specifying modelling methodologies and functionalities of a supporting platform flexible enough for meeting a wide range of system requirements.

---

\(^1\) LD-CAST (Local Development Cooperation Actions Enabled by Semantic Technology): a 6\textsuperscript{th} Framework Program - IST Project, funded by the European Commission
The LD-CAST Semantic Architecture is mainly based on a set of knowledge repositories (e.g., a business processes catalogue, a services registry) and semantic tools (e.g., an ontology management system, a semantic annotation tool.)

The main goals of the LD-CAST semantic solutions are firstly to model and store the knowledge required for enabling cross-border business cooperation, and then to support the search and discovery of knowledge resources (e.g., services, rules, and regulations) needed for building executable business processes.

In particular these goals consist in the definition of:

- an appropriate representation model (meta-model) for service, process, and request ontologies;
- generic templates for defining services, processes and requests;
- a set of specific representation models for business process modeling;
- a methodology for semantic annotation of the different kinds of resources involved, i.e. services and business processes;
- a methodology for service discovery.

As stated, the LD-CAST framework aims at enabling European chambers of commerce (CCs) to support cooperation between private companies in the development of cross border initiatives. The scientific literature on this topic cannot provide a substantial help, as it is already difficult finding researches addressing the management of knowledge even across organizational borders only [Kafentzis et al. 2004].

More concretely, the project will build a European network of LD-CAST portals for allowing end users (mainly private companies) to access, in a seamless mode, the services and any useful information provided by public organizations registered in each LD-CAST portal.

This network will be accompanied by a consistent set of guidelines for designing and implementing one-stop model platforms at all levels (European, national, regional and local), compliant with the European Interoperability Framework (EIF) guidelines. The solutions will be conceived in a way that does not require any specific change in working procedures and/or systems of the participating organizations.

In order to facilitate interoperability between CCs, acting as service providers, semantics-based techniques and tools will be provided and a “common ontology” will be defined, to help model processes, services and different types of requests.

Interoperability between different CCs and easy access to large data repositories are among the key aspects of the LD-CAST project, as supporting end users can require processing large amounts of information. Such information can be heterogeneous, since it concerns different entities (i.e., Business Services, Business Processes, Service Providers, Business Object Documents), but at the same time tightly related, since it belongs to the same application domain and is used for achieving similar objectives.

An ontology-based approach has been chosen for improving the management of the involved resources in terms of their categorization and organization, and for permitting an intelligent retrieval. The ontology-based annotation (semantic annotation) will in fact contribute to identify a precise meaning of such resources and the relationships among them in order to allow an easier and more effective access.

In particular, this paper addresses the Chambers of Commerce semantic interoperability issues and presents the business ontology on which the semantics-based techniques and tools will be based.

The rest of the paper is organised as follows: in chapter 2 a state of the art is outlined, chapter 3 describes the semantic technologies, and chapter 4 presents the two LD-CAST ontology; conclusions and future works are presented in the final section.
2 STATE OF THE ART

As stated by Telesko and Karagiannis, European projects provide a fresh source for harvesting innovative methodologies and field data [Telesko, Karagiannis 2002]. Several address specifically eGovernment issues, and some do it by adopting semantics-based solutions. Among them three have been selected and analysed in closer detail: OntoGov\(^2\) [Apostolou et al. 2005], QUALEG\(^3\) [QUALEG 2006], and Terregov\(^4\) [TERREGOV 2006].

The OntoGov European Project addresses the problem of re-configuring, changing and composing eGovernment services provided by different providers. The proposed solution is a semantics-based software engineering environment for developing and managing the life-cycle of offered services. It is based on 7 small ontologies describing several resources concerning the eGovernment domain. These ontologies define concrete and abstract services, legal documents, organizational units, the information flow, the decision making process in the public administration, and all information needed to finalise the configuration of web services.

The QUALEG European Project addresses the problem of integrating eGovernment services. In particular, it proposes to use ontologies for different purposes: modelling activities performed in the Public Administration; automatically generating a central data repository, the QUALEG database; and managing workflows executed within the QUALEG system. The QUALEG solution is based on 7 ontologies. They concern: the Public Administration, including its organizational structure, the processes, the responsibilities and the roles that civil servants can assume in the activities; the data regarding a particular Public Administration; the database schema; the workflows; and the data required by the QUALEG system to execute workflows.

The Terregov European Project proposes a solution to support interoperability between services provided by local government agencies. Furthermore this project provides centrally controlled orchestrated procedures, involving multiple agencies, to promote transparency and responsibility in eGovernment. The Terregov solution is based on social care ontologies specific for the project pilot regions. Finally multilingualism support is provided by a multilingual reference ontology.

3 SEMANTIC TECHNOLOGIES

This chapter describes supporting the LD-CAST platform, with specific regard to the semantics-based tools that are part of the LD-CAST platform, should provide.

According to the main components of the semantics-based solutions, the chapter is organized into three main sections: ontology, semantic annotation and semantic search and discovery.

3.1 Ontology

An ontology entails some sort of world view with respect to a given domain. It gathers a set of concepts (concerning entities, attributes, processes, etc.), together with their definitions and inter-relationships [Uschold, Gruninger 1996].

More specifically an ontology is “a formal, explicit specification of a shared conceptualisation” [Gruber 1993] where formal refers to the fact that the ontology should be machine understandable; explicit means that the type of concepts used and the constraints on their use are explicitly defined; shared reflects the notion that ontology captures consensual knowledge, that is, it is not restricted to...
some individual, but accepted by a group; conceptualisation refers to an abstract model of some phenomena in the world, which identifies the relevant concepts of that phenomena. A more extensive description defines an ontology as “an explicit, agreed and shared definition of a portion of reality by means of a conceptual model. This model may exist in someone’s head or be embedded in a software or information system, in an object or in a process. The task of an ontology builder is to identify the model and make it explicit. This allows the model to be accessed by, or communicated to, a wider range of potential users, be they people, organisations or software agents”. With respect to a thesaurus, an ontology aims at describing concepts, whereas a thesaurus aims at describing terms. An ontology can be seen as an enriched thesaurus where, besides the definitions of, and relationships among, terms of a given domain, more conceptual knowledge is represented. With respect to a Knowledge Base (KB), an ontology can be seen as a KB whose goal is limited to the description of the concepts necessary for modelling domains. A KB, in addition, includes the knowledge needed to model and elaborate a problem, or to answer to queries about a domain [Missikoff 2005].

An ontology is the pre-requisite for the development and application of the semantics-based solution proposed in the LD-CAST project.

3.2 Semantic Annotation

Semantic Annotation is the activity that allows describing in a non ambiguous way a certain resource. Semantic Annotation is a well known technique that has been proposed in literature originally to annotate documents and web pages. Recently, Semantic Annotation has attracted much attention for its great potentiality in the Semantic Web. The most recent proposals are aimed at the creation of additional structures that represent (in a formal, controlled way) semantic content of a web resource (e.g., a document, a business process or an eService).

Among typical applications of semantic annotation, we can find:
- **Semantic Retrieval**, i.e., the possibility of retrieving digital resources (not only documents) on the basis of their semantic content. This service is at the basis of the following, more specific applications;
- **Document Management**, for the semantics-based organization and retrieval of digital documents;
- **Knowledge Management**, for organization and retrieval of enterprise knowledge;
- **Web Services publishing and discovery**, with semantic matchmaking of requested and offered services;
- **Semantic Interoperability**, by annotating local Enterprise Software Application resources (information and processes) to support business cooperation among enterprise software applications.

The use of annotation addressed in the LD-CAST project intends to cover great part of these application areas, with particular regard to search and retrieval of resources.

3.3 Semantic search and discovery

Semantic search and discovery should allow retrieving searched resources previously annotated. Respect to the LD-CAST objectives, such step is fundamental because it allows, once that knowledge resources (i.e., business processes, service, service providers) have been annotated, to perform searching on the annotations and via the link to the real resource, to retrieve them.

4 LD-CAST ONTOLOGIES

In the LD-CAST project two different ontologies are needed in order to exploit semantic solutions that will be developed during the project: the Business Ontology (BO) and the Quality of Service (QoS) Ontology.

Alessandro D’Atri, Antonio De Nicola and Oleg Missikoff

A Business Ontology for Supporting Cross-Border Cooperation Between European Chambers of Commerce
The BO includes concepts related to the eGovernment domain. In particular it provides a formal and shared specification of the actual services, the service providers, and the corresponding administrative business processes. The BO can be considered as an ontological basis, including the main concepts of the domain of interest, that can be easily extended by domain experts. The objectives of this BO are to increase the understanding among business partners, and to use it as a basis for semantic search and retrieval of annotated resources.

The QoS ontology includes concepts related to the service, user, and process quality. The QoS ontology can be used to provide a shared specification of service and process quality, to discovery services by considering QoS values, and to monitor service and process quality after or during service and process execution.

In the next sections we focus on the Business Ontology.

4.1 LD-CAST Business Ontology

Methodology to build the Business Ontology

In order to build the Business Ontology we adopted UPON (Unified Process for ONtology building) [De Nicola, Missikoff, Navigli, 2005], an incremental methodology for ontology building. The methodology stems its characteristics from the Software Development Unified Process [Ambler, Nalbone, Vizdos 2005], one of the most widespread and accepted methods in the software engineering community, and uses the Unified Modeling Language (UML) [http://www.uml.org] to support the requirements collection, the analysis of the scenario, and the preparation of all the blueprints of the ontology project.

What distinguishes the UP and UPON from other methodologies for software and ontology engineering, respectively, is their use-case driven, iterative and incremental nature.

Another characterizing issue of UPON is the distinction of the application knowledge vs. the domain knowledge. The latter represents the context in which the former is placed. The former takes into consideration the latter but focusing on the application issues. Domain knowledge is mainly acquired through pre-existing resources having a general scope (glossaries, technical documents, manual and technical guides, etc.) while application knowledge is provided ad hoc by the involved domain experts.

In UPON there are cycles, phases, iterations and workflows. Each cycle, consisting of four phases (inception, elaboration, construction and transition), results in the release of a new version of the ontology. Each phase is further subdivided into iterations. During each iteration, five workflows take place: requirements, analysis, design, implementation, and test. The requirements workflow is the process of specifying the semantic needs and the knowledge to be encoded in the ontology. The analysis workflow concerns the refinement and structuring of the ontology requirements identified in the requirements workflow. The design workflow concerns refinement of entities, actors and processes identified in the analysis workflow, as well as the identification of their relationships. In the LD CAST project, concept refinement is achieved according to the OPAL meta model [Missikoff, Taglino 2002], an ontology representation methodology based on UML [UML] and OWL [Antoniou G. and at., 2004], developed at LEKS, IASI-CNR. The implementation workflow concerns formalization of the ontology in a suitable ontology language (e.g., OWL [Antoniou, Harmelen 2004]). The test workflow allows to verify the consistency of the ontology and the correct implementation of its requirements.

Workflows and phases are orthogonal in that the contribution of each workflow to an iteration of a phase can be more or less significant: early phases are mostly concerned with establishing the requirements, whereas later iterations result in additive increments that eventually bring to the final release of the ontology. Notice that more than one iteration may be required to complete each of the four phases. This approach follows faithfully the Unified Process.
At each iteration different workflows come into play and a richer and more complete version of the target ontology is produced. The incremental nature of UPON requires first the identification of the relevant terms in the domain, gathered in a lexicon; then the latter is progressively enriched with definitions, yielding a glossary; adding to it the basic ontological relationships allows a semantic network to be produced, until further enrichments and a final formalization produces the sought reference ontology. Furthermore, the progressive enrichment from a lexicon to an ontology is a peculiarity of UPON that makes it an original work with respect to UP.

As already mentioned, each concept of the BO is categorized by associating a “kind” to it according to OPAL. Such kinds include the major ontological categories, according to proposals of top ontologies, such as [Sowa, 1999], or meta-ontologies [Uschold and King, 1995]. OPAL organises concepts in three primary and some complementary categories. The primary categories are:

- **business actors**: entities of domain that are able to activate or perform a business process. A business actor is an active element of a business domain. Domain experts, in analyzing the reality, are asked to identify relevant actors that operate producing, updating or consuming business objects. An actor can be a Company, a Chamber of Commerce, and an Employee.
- **business objects** (and **business object documents** [BODs]): entities on which a business process operates. A BOD is a further refinement that represents a type of document in the business domain (e.g., Company History Certificate, Certificate of Existence.)
- **business processes**: business activities aimed at the satisfaction of a business goal, operating on a set of business objects (e.g. Releasing Fiscal Verification Document, Paying for Service.) It can be rather simple, with a limited duration in time, or complex, with parallel branches and phases that last for a long time span.

Some complementary categories are:

- **complex/atomic attributes**: in modeling the properties of a concept, we distinguish between structured information, such as Address, and elementary information, such as Street Name. Essentially, a (structured) complex attribute is defined as an aggregation of lower level complex and/or atomic attributes.

Furthermore, concepts are linked together by some ontological relations that represent well known modeling notions common to (the meta-model of) the majority of Knowledge Representation Languages. They can be organised into two groups: vertical and horizontal relations. The vertical relations are:

- **ISA** (refinement): relation among concepts. For instance, Italian VISURA is a Legal Verification Document.
- Decomposition: part-of relationship among concepts. For instance, Italian Issuing Legal Verification Document is a part of Legal Verification.

The horizontal relations are:

- **Predication**: it relates attributes to a concept. For instance, an Invoice is in predication relation with Date, Amount, and Recipient.
- **Relatedness**: domain specific relationships (named or unnamed) among concepts. For instance, an invoice is related to Customer (unnamed), Customer buys Service (named).

The Business Ontology

The scope of the LD-CAST Business Ontology is eGoverment. The proposed BO includes all relevant administrative domain concepts and refers to an ontological framework, the so-called ontology chestnut [De Nicola, Misceo, Missikoff 2005] (figure 1.) According to this framework, we can structure business ontology in three different hierarchical layers:

- **Upper Domain Ontology**: UDO. It is positioned at the top of the ontology chestnut and concerns generic business concepts (common to all countries), such as service Fiscal Verification, Certificate, or Company.
• **Application Domain Ontology: ADO.** It is positioned in the middle of the ontology chestnut and concerns structured concepts that relate only to a specific country, such as Italy, Poland, and Bulgaria.

• **Lower Domain Ontology: LDO.** It is positioned at the bottom of the ontology chestnut and concerns elementary concepts (essentially attributes and elementary activities), such as *Price*, *Street Number*, or *Company Name*.

Following this approach, the Core Business Ontology is the union of the set of concepts coming from the UDO and the LDO.

We remark that, to produce an effective ontology, a community of domain experts (in our case represented by the LD CAST Consortium) has been involved in the building process, and an agreement on the meaning of concepts has been reached.

![Figure 1. The Ontology Chestnut](image)

The business ontology, built according to this framework, is an open document. The present version represents a starting point and the ontology will be enriched as the project proceeds. Currently, the ontology consists of 326 concepts, 183 of which belong to the LDO, 75 to the ADO and 68 to the UDO. It is reasonable to estimate that this number will increase considerably by the end of the project.

Concerning the methodology for extracting terms, a first set has been produced using a prototypical software called “TermoStat” [Drouin 2003]. This software conducts automatic term acquisition in specialized corpora using comparisons with the behaviour of the lexicon in non-technical corpora. Contrary to many other term-extraction techniques, this system is able to retrieve single-word terms.

To perform this extraction, TermoStat computes the “specificities” of words occurring in a specialized corpus by comparing their frequency in the corpus and in a general-language corpus. Basically, the higher the specificity of a word, the more likely it is to be a term of the domain. Conversely, a word with a negative specificity coefficient certainly belongs to the general language.

The first set of terms automatically produced by TermoStat has been manually refined to finally produce the version of the business ontology (structured in LDO and UDO, both composing the Core Business Ontology, and the ADO.)

**Upper Domain Ontology**

For extracting the concepts of the UDO it has been used as a base a deliverable of the project describing the scenario, where a first automatic extraction with TermoStat has been followed by a
manual refinement. In particular the sections of the deliverable analysed were contained in the “As-is scenario” and, more specifically, in the “Supply sub-contract” section. The parts analysed in detail have been the following: Legal Verification, Fiscal Verifications, Technical & Quality standards Verification, General Requirements, Agreement, Registration, Mediation and Arbitration.

As stated above, this analysis has yielded 68 concepts, but this number is expected to grow considerably in future phases of the project. In the following table 1, examples of concepts of the UDO, categorized according to the corresponding ontological category, are presented.

<table>
<thead>
<tr>
<th>Ontological Category</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business Actor</strong></td>
<td>Chamber Of Commerce, Company, Customer, District Court, Trade Register</td>
</tr>
<tr>
<td><strong>Business Object</strong></td>
<td>Annual Turnover, Arbitration Clause, Business Ownership Structure, Corporate Form, Environment Quality, eSignature, Liability, Smart Card, Social Asset, Social Pacts</td>
</tr>
<tr>
<td><strong>Business Process</strong></td>
<td>Arbitrating, Conciliation, Fiscal Verification, Issuing Notification, Legal Verification, Licensing, Partner Search, Paying, Releasing Document</td>
</tr>
</tbody>
</table>

Table 1. Examples of concepts of the UDO categorized according to the corresponding ontology category

Lower Domain Ontology

The LDO contains elementary concepts, 101 elementary activities, and 82 attributes.

Concerning attributes, to achieve higher interoperability features, it was built by considering reuse of existing resources. In particular, to populate the LDO, we referred to some business initiatives and XML-based specifications for internet transactions. The following standards were considered: ebXML, RosettaNET, and OAGIS. We analyzed 140 elements coming from ebXML Core Components Dictionary and ebXML Naming Conventions for Core Components; 1873 from RosettaNet Business Dictionary v2.1 31_Oct_2002; 600 from OAGIS specification 8.0.

Considering that the objective of LDO is to be generic, in order to use it in every specific business application domain, we omitted many elements peculiar to a particular application domain (e.g. elements coming from RosettaNet that is a consortium operating in the high-tech industry). For this reason, we reduced the bulk of elements to be considered to approximately 200. Then a statistical analysis was performed on them aimed at identifying LDO elements. Finally, we selected 82 attributes.

Concerning elementary activities, the LDO contains 101 concepts belonging to four countries: Italy, Poland, Romania, and Bulgaria. The identification of concepts has been performed with the support of Chambers of Commerce through interviews and analysis of existing technical documentation. In the following table 2 examples of elementary activities, organized according to the corresponding countries, are presented.

<table>
<thead>
<tr>
<th>Country</th>
<th>Elementary activities</th>
</tr>
</thead>
</table>

Alessandro D’Atri, Antonio De Nicola and Oleg Missikoff

A Business Ontology for Supporting Cross-Border Cooperation Between European Chambers of Commerce
### Table 2. Example of elementary activities belonging to the LDO

**Application Domain Ontology**

The ADO concerns structured concepts related to specific countries. It has been built by specializing concepts of the UDO and by aggregating concepts of the LDO. An example is the concept *Italian Partner Search*. This concept is a specialization of the UDO concept *Partner Search* and it is composed by aggregating the following LDO concepts: *IT Paying Fee*, *IT Requesting Partner Search*, *IT Producing Partner Search Report*, *IT Issuing Partner Search Report*. In the following table 3 examples of the ADO concepts, categorized according to the corresponding ontology kind, are reported.

<table>
<thead>
<tr>
<th>Ontological Category</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Actor</td>
<td><em>RO Trade Registry</em>, <em>PL National Registration Court</em>, <em>PL National Criminal Registry</em>, <em>LSCC</em>, <em>IT Italian National Income Agency</em>, <em>IT INPS</em>, <em>IT INAIL</em>, <em>IT CC</em>, <em>CCIR</em>, <em>BCCI</em></td>
</tr>
<tr>
<td>Business Object Document</td>
<td><em>RO Trade Registry Application Form</em>, <em>RO Proxy Document</em>, <em>RO Payment Receipt</em>, <em>RO Partner Search Request</em>, <em>RO Partner Search Report</em>, <em>RO Partial Search Result</em>, <em>RO Order</em>, <em>RO Offer</em>, <em>RO Legal Verification Request</em></td>
</tr>
<tr>
<td>Business Process</td>
<td><em>BG Fiscal Verification</em>, <em>BG Legal Verification</em>, <em>BG Partner Search</em>, <em>IT Fiscal Verification</em>, <em>IT Legal Verification</em>, <em>IT Partner Search</em>, <em>PL Fiscal Verification</em>, <em>PL Legal Verification</em>, <em>PL Partner Search</em>, <em>RO Fiscal Verification</em>, <em>RO Legal Verification</em>, <em>RO Partner Search</em></td>
</tr>
</tbody>
</table>

**Table 3. Examples of concepts of the ADO categorized according to the corresponding ontology kind**

Alessandro D’Atri, Antonio De Nicola and Oleg Missikoff
A Business Ontology for Supporting Cross-Border Cooperation Between European Chambers of Commerce
5 CONCLUSIONS AND FUTURE WORKS

In this paper we presented a Business Ontology to support cross border cooperation developed in the context of the LDCAST Project. The developed ontology consists of 326 concepts (20 actors, 122 processes, 92 BODs, 10 objects, 82 attribute) and 692 relationships (112 specialisations, 104 decompositions and 476 relatedness.) Since the project aims at enabling cross border cooperation, it is extremely useful for final users to be able to find connections between domestic concepts with similar ones used in other countries. In performing this task the ontology could provide a strong support. Every concept would therefore be assigned a list of “similar concepts” each with a coefficient of similarity (from 0.4=very different to 1=synonym). This would provide a substantial support to the work of local service providers in helping enterprises to initiate a new activity in a foreign country. For instance, starting from a familiar concept, it would be easy to discover whether the country where a certain company wants to open a branch, provides a similar service even if differently called. This functionality would draw the ontology closer to an interoperability platform, and would provide a practical working tool for final users.

References

(Staab S, Studer R. eds.), Springer, pages 67-92
UML (Unified Modeling Language) - http://www.uml.org