Mapping between heterogeneous XML and OWL transaction representations in B2B integration

Jorge Cardoso [a,*], Christoph Bussler [b]

[a] CISUC/Departamento de Engenharia Informática, Universidade de Coimbra, Portugal
[b] Saba Software, Inc., USA

ARTICLE INFO

ABSTRACT

XML-based standards have been widely used to enable and ease Business-to-Business (B2B) integration. Examples of standards include cXML, CIDX and ebXML. While these XML-based standards are syntactic, contemporary organizations have available new means to structure their internal data representations using semantic descriptions, such as RDF(S) and OWL. This scenario poses an interesting challenge: "How to reconcile external XML-based standards and internal OWL-based representations in B2B integration scenarios?" In this paper, we present a conceptual approach, and its implementation, to integrate external syntactic data representations with organizational internal semantic data representations by using the notion of heterogeneous mappings which are established between the two types of representations. The application developed, B2BISS, enables an effective management of mappings. As the number of mappings stored in the repository increases over time, organizations can gradually rely on a semi-automatic to automatic B2B integration.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Global economies are increasingly becoming networked. The notion of value chains [39], value nets [35], b-Webs [45], and value networks [36] as concepts and tools have been used to understand and analyze networked industries. They are useful instruments for portraying the interconnection of operations, players and transactions. Many industries now exhibit strong co-operative behavior with inter-firm relationships having a significant role in strategic performance. The existence of business chains, nets, b-Webs or networks leads invariably to the necessity of developing Business-to-Business (B2B) solutions for integration [32]. The objective of the integration is to ease the management of transaction-based interactions between business players that are part of a network to cut costs, increase revenues, and improve time-to-market [11,5,48].

Nowadays, one simple solution that organizations are adopting to reach business network integration is to rely on the use of XML-based domain specific standards to exchange transactions [8]. Examples of well-known standards include cXML (Commerce eXtensible Markup Language), CIDX (Chemical Industry Data Exchange), ebXML (Electronic Business using eXtensible Markup Language), XML/EDIFACT, papiNet, PIDX (Petroleum Industry Data Exchange), and xCBL (XML Common Business Library) [33].

While XML-based standards allow data exchange between networked businesses, they do not guarantee the interoperability of systems. XML only provides syntax to structure the data exchanged in B2B settings. If on one hand external XML-based standards are syntactic [3], with the adoption of semantics to make data explicit, organizations are considering shifting from a syntactic representation level to a semantic one [49,10]. By using semantic domain models based on ontologies (e.g. RDF(S) or OWL [1]) enterprises acquire several benefits, such as the ability to perform inference on knowledge bases and the capacity to share domain models to easily exchange and integrate information.

---

* Corresponding author. Tel.: + 351 239 790 000; fax: + 351 239 701 266.
E-mail addresses: jcardoso@dei.uc.pt (J. Cardoso), chbussler@aol.com (C. Bussler).

The networked economy [23] requires accounting for the nature of alliances, the technical infrastructures of players, and the data exchanged in business networks. Once a network is designed, it is possible to analyze the model to identify gaps. Once gaps are known, implementation plans can be prepared to close them. One important gap to close for achieving a stronger B2B integration is the lack of a common understanding between external standards and internal organizational data representations [30]. Developers are still faced with the problem of understanding the meaning of the information represented in XML-based standards and establish a correspondence or mapping to the internal OWL-based data representation of organizations [34,48]. The manual software coding of mappings between data models is a time consuming task with high costs.

WSML, WSMO and WSMX [50] use semantically annotated services to promote B2B integration by using data and process integration. Transactions can be semantically enriched using the WSML ontology language. The infrastructure enables domain experts to create mappings between XML-based standards and WSMO ontologies. The mappings are represented in an abstract ontology mapping language. While this research provides an important theoretical contribution, it reveals a few limitations when it needs to be transposed to real-world industries settings. On the one hand, the use of non-standard languages to establish mappings may be considered a penalty for not allowing its sharing across an industry. The use of more well-established languages and tools (such as OWL and XSLT languages and parsers) enables an easier adoption by organizations. On the other hand, the created mappings cannot be reused. This means that for each standard, and its variations, used to represent transactions, a new set of mappings needs to be created. A repository of shared mappings that can be reused overtime would increase the adoption of the approach. Finally, a more responsive and customized system that would automatically classify incoming transactions and requested human involvement when new mappings were needed to fully complete a transformation from instances of one model to instance of the other model would be more adequate in real-world settings.

In this paper we describe a conceptual approach that allows organizations to participate in B2B networks using XML-based syntactic standards to support external transactions while structuring their internal data representations semantically using semantic languages such as OWL. In our approach, partners and suppliers can freely exchange syntactic XML-based transactions. Once an organization receives a syntactic external transaction it is allowed to create a heterogeneous mapping between its elements and the concepts of an internal semantic model (i.e., an ontology) that describes the organization's domain. The conceptual approach to manage heterogeneous mappings has been implemented with the B2BISS system. B2BISS tackles the problem of *information integration* since it fosters aspects such as linguistic and semantic differences to be reconciled among disparate data representations [38,11]. The system allows organizations to dramatically decrease the time and cost of integration by providing a flexible and easy to use graphical tool. Such a flexible infrastructure is instrumental for a rapid and cost-effective B2B integration [37].

The remainder of the paper is structured as follows. The second section presents a B2B integration scenario that illustrates the challenge of integrating external XML-based standards and internal OWL-based data representations. Section 3 enumerates the challenges, approaches and our solution to the problem of syntactic (external) to semantic (internal) B2B integration. Section 4 introduces the B2BISS system which enables a (semi-)automatic integration and relies on the notion of managing and reusing heterogeneous mappings. Section 5 describes five organizational scenarios involving the use of the B2BISS system. Section 6 presents a use case. It describes how B2BISS can be implemented and used within the SAP NetWeaver Process Integration infrastructure. Section 7 presents the related work in this area. Finally, Section 8 closes the manuscript with our conclusions.

2. B2B integration scenario

Organizations implement B2B infrastructures to allow a seamless interaction with several business partners that include suppliers, distributors, exporters, and retailers (see Fig. 1). B2B integration "is not just the buying and selling of goods and services, but also servicing customers, collaborating with business partners, and conducting electronic transactions within an organization" [47]. Drivers for B2B network integration include the need for disintermediation, to reduce inventory cycle time, to optimize business processes, and to use various distribution channels. On the other hand, barriers to integration are related to high technology costs, complex technology (OWL, cXML, ebXML, etc.), and confidentiality and privacy concerns.

One very important aspect that needs to be recognized is that different organizations use different internal semantics to describe their products, part numbers, invoices, and purchase orders. Therefore, when organizations are collaborating in business networks they have to understand the external data representation of the standards adopted to represent transactions. Then, when an organization receives a transaction, it needs to map its data to its own internal data representation. In Fig. 1 it is possible to find two types of data representations: external and internal. The central organization, named Hiltrix Corp., uses ontologies for its data model while it exchanges transactions with suppliers and customers using XML-based standards.

2.1. External data representations

External data representations are used with partners to exchange transactions. In general, every industry develops standards in order for companies to use transactions to communicate with each other. With the advent of the eXtensible Markup Language (XML), it became easier to define and standardize the contents of B2B transactions. XML is a highly flexible, ubiquitous data format, designed for multiple representations.

In progress of time, B2B solutions went through an evolutionary path from monolithic and proprietary standards (e.g. TRADACOMS in UK, ANSI ASC X.12 in U.S.A and UN/EDIFACT in the United Nations) toward flexible and standardized XML-based stacks covering the requirements from different industries [30]. Examples of existing standards to communicate with partners include cXML, CIDX, ebXML, XML/EDIFACT, papiNet, PIDX, and xCBL [8,6].