Computer Standards & Interfaces 31 (2009) 1045–1055

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Business to business interoperability: A current review of XML data integration standards

Fenareti Lampathaki *, Spiros Mouzakitis, George Gionis, Yannis Charalabidis, Dimitris Askounis

National Technical University of Athens, School of Electrical and Computer Engineering, Heroon Polytechniou 9, 15773 Zografou, Athens, Greece

ARTICLE INFO

Article history:
Received 6 April 2008
Received in revised form 26 October 2008
Accepted 7 December 2008
Available online 24 December 2008

Keywords:
Data integration
Standards evaluation taxonomy
UBL
OAGIS
CCTS

ABSTRACT

Despite the dawn of the XML era, semantic interoperability issues still remain unsolved. As various initiatives trying to address how the underlying business information should be modelled, named and structured are being realised throughout the world, the importance of moving towards a holistic approach in eBusiness magnifies. In this paper, an attempt to clarify between the standards prevailing in the area is performed and the XML Data Standards providing generic XML Schemas are presented. Based on this "XML Data Standards Map", a multi-faceted classification mechanism is proposed, leading to an extensible taxonomy of standards. A set of facets is analyzed for each standard, allowing for their classification based on their scope, completeness, compatibility with other standards, openness, ability to modify the schemas and maturity, to name a few. Through populating and querying this multi-faceted classification, a common understanding of Data Integration Standards can be ensured and the choice of a standard according to the requirements of each business can be systematically addressed.

Contents

1. Introduction

Data modelling issues have aroused the interest of the research community since the late 1960s when EDI (Electronic Data Interchange)

* Corresponding author. Tel.: +30 210 772 2078; fax: +30 210 772 3550.
E-mail address: flamp@epu.ntua.gr (F. Lampathaki).

0920-5489/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csi.2008.12.006

that facilitated the application-to-application exchange of standard business documents between companies, independently of software, hardware, and communication networks, was introduced. With the exponential growth of the Web that opened new opportunities for businesses to transact across all types of boundaries (geographical, national, business category, etc.), early research had focused on providing a lingua franca for B2B e-Commerce, XML, that went beyond HTML to reflect the richness of the data being published. In progress of time e-Business modelling frameworks went through an evolutionary path from monolithic and proprietary standards (e.g. TRADACOMS [32] for the UK retail industry, ANSI ASC X.12 [1] in North America and the United Nations-recommended UN/EDIFACT [33]) towards flexible and standardized XML-based stacks covering the requirements from different industries.

It had been a common belief that creating XML vocabularies was sufficient to achieve data interoperability, yet this assumption goes far beyond reality. XML by itself does not guarantee that XML expressed business information exchanged in the span of business processes across different enterprises will be understood equally well by all systems. This is because the XML syntax only provides for creating markup languages used as metadata, it does not address how the underlying business information must be modelled, named and structured. Semantics come to cover this gap by attaching meaning to data in a structured and technical way that both humans and machines can understand and process.

The difficulty however is that currently most modelling languages focus almost exclusively on the technical aspects of creating the model rather than the semantic aspects necessary for true interoperability. Many industry-specific consortia, like CIDX [47], PIDX [48], OAGi [23], and RosettaNet [49], have indeed solved major technical issues of traditional EDI, but appear insufficient to provide a common understanding of the underlying data and arrange the semantics of the business information. The prevalent "business standards dilemma", defined as the diversity of standards that address particular data requirements, but are designed on such a different basis that make the choice of a specific standard to be adopted a new challenge, is compounding the problem [15,29]. For example, trading partners have to deal with several standards at the same time and, since only recently a methodology for standardizing business semantics (ISO 15000-5, commonly known as UN/CEFACT Core Component Technical Specification) has emerged, careful analysis — on the part of the developer to ensure correct understanding and interpretation, mapping and integration between interfaces is required incurring great expense.

As long as the semantic discourse of data and standards exists, a holistic approach in e-Business cannot be achieved [27]. In this context, the present paper is oriented towards analyzing the Data Modelling State of the Art in terms of core technologies, like XML and UN/CEFACT CCTS (Core Components Technical Specification), and international standards and initiatives in Data Modelling, that have produced sets of XML Schemas, like xCBL, eBIS-XML, OAGIS, UBL, XBRL and xCBL. Such standards fall into the same jurisdiction and try to address data integration issues with the adoption of semantically-enabled XML Schemas.

With regard to the fact that generic classifications of standards (e.g. International Classification of Standards [18]) do not serve the exact needs of the selection and evaluation of data modelling standards; they merely list different standards, a conceptual framework that takes the form of a multi-faceted taxonomy has been developed for the systematic evaluation of standards. The first version of the proposed evaluation framework was developed in the context of the EU-funded "GENESIS: Enterprise Application Interoperability via Internet-Integration for SMEs, Governmental Organizations and Intermediaries in the New European Union" Project [12]. Its aim is the research, development and pilot application of the needed methodologies, infrastructure and software components that will allow the typical, usually small and medium, European enterprise to conduct its Business transactions over Internet, by interconnecting its main transactional software applications and systems with those of collaborating enterprises (B2B transactions), governmental bodies (B2G transactions), banking and insurance institutions (BNK transactions) with respect to the EC current legal and regulatory status and the existing one in the new EU, candidate and associate countries. The development of the evaluation framework was due to a practical need faced in the project [11]: there were many potential data modelling standards and available specifications, and the need to evaluate and select those that would serve each integration need (i.e. B2B, B2G and BNK) emerged. In addition, it was not always clear, which transactions could be covered by the proposed models, standards and specifications and it quickly became evident that there was no evaluation model in the bibliography that would consider all the needed aspects.

In this context, a data modelling evaluation framework was developed and used both for a quick overview of several data modelling standards and a more thorough evaluation of a selected core related to the scope of the project. This paper proposes an extended version of the GENESIS-related evaluation framework that provides the means for the systematic analysis of any B2B data standard and for deducting conclusions regarding the most appropriate standard according to the weight each business poses to the criteria — facets. The set of parameters and characteristics of the standards indicatively include scope, completeness, openness, modularity, maturity, configuration support and modelling of messages and aim to provide a thorough understanding of the standard before implementation. Our approach also presents innovative aspects by incorporating for the first time facets, such as integrated management of enterprise and data models, cross-country support, support for rules modelling, workflow capabilities incorporated into the documents and compatibility with other standards, in the standards' evaluation phase.

The remainder of the paper is structured as follows: in the second section related work in literature is reviewed and discussed. An introduction on Data Modelling is given in Section 3. Section 4 proceeds with the core technologies in Data Modelling. The presentation of the XML Data Integration Standards providing generic (not industry-specific) B2B XML Schemas and the taxonomy of the standards follow in Sections 5 and 6 and lead to the evaluation of standards in Section 7. A short summary of our results and further research activities required towards the next generation of enterprise data integration complement this work.

2. Literature review

In the span of this work related research efforts which reference Business Information and Data Integration and Modelling [2–4,6,7,10,17,19–21,26,28,30,31,38,45,46,50] were examined and reviewed. The main findings upon which our approach builds originate from the following relevant work:

- In [20], a state of the art in e-Business Frameworks that are suitable for industrial procurement, design, production or distribution and were active in 2004 is presented. The frameworks under scrutiny are: BPEL, BPML, CIDX (Chemical Industry Data Exchange), cXML, ebXML, OAGIS, papiNet, PIDX (Petroleum Industry Data Exchange), RosettaNet, UBL, xCBL and XPDL (XML Process Definition Language). Key variables related to the properties and standardization of the e-business frameworks are also identified and depending on the values assigned to the variables a basis for analyzing the commonalities, differences and regularities between them has been created.
- In [38], XEDI, CBL, cXML, ebXML, RosettaNet, IFX, OFX, FPML, FinXML, IDEAlliance, HL7, HR-XML and OAGIS are listed as XML/EDI standards for business communication.
- The main issues and solutions to B2B e-Commerce interactions are surveyed and the dimensions for evaluating B2B interaction frameworks are proposed in [3]. A representative set of XML-based B2B interaction frameworks including xCBL, BizTalk, cXML, RosettaNet and ebXML is also presented.