

An Environment for a Powerful Business Document Schema Profiling

Nicola GESSA¹, Arianna BRUTTI¹, Valentina CERMINARA²
Piero DE SABBATA¹, Angelo FRASCELLA¹, Cristiano NOVELLI¹

¹ENEA XLab, via Martiri di Monte Sole, Bologna, 40129, Italy

Tel: +39 051 6098 111, Fax: + 390516098084,

Email: nicola.gessa@enea.it, arianna.brutti@enea.it, piero.desabbata@enea.it,
angelo.frascella@enea.it, cristiano.novelli@enea.it

²FTI-Forum per la Tecnologia dell'Informazione, via Martiri di Monte Sole, Bologna, 40129, Italy, Email: valentina.cerminara@gmail.com

Abstract: In this paper we discuss some aspects in business document customisation, and we present a suite of applications, the Document Customisation Suite, that can help the final users of a business standard in its adoption. This work starts from our studies and experiences on setting up business collaborations among Small and Medium Enterprises (that have specific exigency and characteristics) and from various activities performed in the context of the European project. It bases on the peculiarities of existing business standards and schema languages for document template definition. The paper is moreover strictly related with a set of initiatives for the definition of an open framework that supports the adoption of common specifications for business collaboration setting up. The context of this work is the definition of common frameworks to enhance the interoperability of enterprise systems in performing business transactions and collaborations.

1. Introduction

In the scenario of eBusiness collaborations one of the possible key points to improve system interoperability is the definition and the adoption of common data formats that can satisfy the need of the enterprises to have a shared business language for the business data exchange. Then, standard definition and adoption should represent the obvious approach to face and solve, at least, one of the dimensions of the interoperability problem, the document interoperability. A standard specification, in fact, provides both a common syntax to represent the data and, more important, a related semantic that is supposed to be unambiguous and accepted by all the potential business partners.

The relevance of the role of the standards has been recognized in the history of communication framework definition and it is witnessed not only by the presence of many official national and international standardisation bodies, but also by the huge number of private consortia and initiatives that try to solve communication problems in business communities. On the other hand the resulting specifications are often felt as constraints and their adoption in some cases clashes with the characteristics of the real business scenarios: this leads to poor usage of the standards themselves.

It is worth to observe the different behaviour of technological versus applicative standards: technological standards (for example related to a communication protocol like HTTP) are quite stable once they have been released by a standardisation organisation; on the contrary an applicative standard for eBusiness (for example ebusiness document specifications like UBL) is continuously required to be improved and enlarged to cover the variety of real life businesses.

This behaviour has led to a standard life cycle more complex than in the past, where many actors in parallel continuously try to update and extend the content of the standardised specifications and, acting at different levels, provide customisations for a specific applicative domain or for an individual case.

This paper presents a set of applications designed to support a new approach to foster the adoption of standardised specifications in the field of the data models. In section 2, we will describe our vision about the mechanisms that regulate and drive the definition and adoption of standards, in order to highlight the key-points and the issues in the standard definition process. In this section we want to make a distinction about different phases and roles, and consequently different needs that constitute the standard definition and adoption process

In section 3, we want to briefly describe the state of art about the main technologies for an enhanced data format definition, and to discuss some recent applications that tackle the problem of business document standard management. Section 4 presents the Document Customisation Suite, describing its final aim, its characteristics and functionalities, providing a view of the novelties of our approach. Finally some conclusions will follow.

2. Criticalities in Standard Definition and Adoption

A list of standards for eBusiness would result to be incredibly long, many of them responding to a specific approach and standardisation model, with very different results in terms of adoption and impact in the business. On the other way, it is possible to provide a set of criticalities that regards more or less every business standard, from the enterprises perspectives, in the adoption of standards. These criticalities are:

- Time: the standardisation life cycle often results to be misaligned respect to the business activities.
- Resources: time and economic resource needed by the enterprises to participate in standardisation processes.
- Usability of the specification of the standards.
- Adoption: cost required to adopt and integrate standard specifications in legacy systems or ERP.
- Return of Investment: the advantage perceived from the enterprises with the adoption of the standards.
- Reliability of standards: how much the standards are reliable or how much they rapidly change in time?

If we consider all these points we can image the difficulties, especially for the SME, to evaluate the opportunities provided by a standard, and to undertake the set-up of an existing system versus standardisation initiatives. On the other hand, standards have being recognized as a key factor for the overcoming of those barriers that still exist for data exchange and enterprise system interoperability. A way to face these criticalities consists for sure in making standardisation processes more simple and efficient, and to provide tools to improve standard flexibility. Analysing a standard specification, two key aspects should be considered:

1. Standard management processes should be simplified facing the complexity of the standard, which is related with redundancy and uncertainty in its implementation [7].
2. Standard flexibility can be improved with proper tools for an easy and ad hoc (from an enterprise perspective) customisation.

A very relevant consideration can be done at this point: standard definition and customisation can be performed at different moments during the life cycle of a standard: in fact two well distinct types of users need to manage standard specifications:

1. That responsible in the definition of the standard specification. In this case the aim is to evaluate, harmonize and integrate all the different requirements that reflect the different needs coming from different enterprise contexts. For this reason it is very difficult to obtain a simple and nimble set of specification since the complexity of the standards should be able to support all the actors in a specific business sector: a too simple standard probably would result in a naïve framework completely unsuited for real data exchange. A very relevant consideration can be done at this stage. As highlighted in [7], standards may have different levels of complexity, described in terms of uncertainty and redundancy, but although the standards should be simplified, it not feasible to avoid a certain level of complexity. More a standard is general, more it is complex, cumbersome and hard to understand and to adopt. For example, in case of UBL[1] some relevant documents, like Order, are constituted in principle of million of elements that allow on one hand to cover a wide range of requirements, but in practice result very hard to adopt.
2. That interested in the personalisation of the specification. In this case the aim is to tailor the standard specification (especially in terms of data formats) on the very specific requirements of an enterprise. This activity is strongly related to all the ICT infrastructures upon which is based the functioning of the enterprise data exchange and management. Many times database, software, mapping mechanisms, ERP systems represents a relevant counterpart that must be taken in account during the customization of the standard.

In order to respond to specific exigency, these two different figures need different abstraction levels, interfaces and tools to work on the specification. Also the input of these figures are different: for the responsible of the standards, the main input is basically an abstract and general vision, while for the responsible of the customisation the main input are the software systems that will be interfaced with the standard.

3. Document Schema Customisation, State of Art

The status of art will regard two different aspects related to our approach in the definition of a customisation environment:

- The technologies available to allow document schema customisation and management, and the solutions available to produce flexible schemas with the use of co-constraints, and to implement dynamic applications that could support the management of a schema document.
- Existing approaches and applications, with their specificities, in supporting document schema definition and management [8][2].

Nowadays many proposals and specifications exist that define schema languages. DTD, Relax NG[10], Schematron[3], XML Schema[4] are the most famous, each of them with its syntax and characteristics.

These languages allow defining schema documents or, in other worlds, allow providing the description of the structure of a document, or a set of validation rules that must be fulfilled writing a document that want to be correct for a specific use case. These languages are very useful in those contexts where there is the need to define a common structure and a shared vision about format for data exchange. On the other hands, these languages have different level of expressiveness; moreover, during these last years the requirements expressed by the actors for the definition of schemas and rules have becoming more and more complex.

There are basically two types of constraints that can be imposed in a schema documents:

1. Simple constraints: with these types of constraints it is possible, for a specific element in a document, to set the minimum and the maximum number of occurrences, to define

its data type (and characteristics like length or number of decimal), default values, enumeration values, value boundaries and so on.

2. Co-constraints: these are constraints dependent on the context, since they represent constraints on elements related with the presence and the values of other elements in the same documents. Some example of co-constraints are:
 - mutual dependence constraints: linking the presence/absence of an element/attribute to the presence/absence of another element/attribute;
 - restrictions on the open content: to assess the presence/absence of specific elements/attributes in a content model that is not precisely defined;
 - arithmetic and mathematical constraints: allow you to verify that the value of an element or an attribute satisfies a mathematics condition calculated on other parts of the document;
 - content dependent on the context: bind content model allowed for an element/attribute to the position or the presence/absence of elements/attributes in the rest of the document;
 - content dependent on the value: bind content model allowed for an element/attribute to the values of other elements/attributes.

The term “co-constraint” is often used to refer to those constraints that cannot be expressed using XML Schema 1.0.

It is relevant to note that whereas simple constraints are basically related with the structure of the document, that can be defined generally for a wide set of use contexts and applications, co-constraints are more related with the content of the document and are tailored for more specific use domains. For example, using simple constraints it is possible to define the structure of an Order document that can be used in various and very different industrial domains. On the other hand, co-constraints can be used to define a set of rules specific for a business domain (that adopts specific list of values), but also to refine document schemas for more specific use, for example matching requirements coming from a single enterprise.

In order to write document schemas, many commercial applications are available on the market. In this analysis we want to consider two applications that are specific for the customisation of business document standard:

- eDoCreator[9]. This application provides a graphical environment for the modification and customisation of document schemas based on UN/CEFACT CCTS [11]. This application, developed within the iSURF project, implements the UN/CEFACT CCTS methodology and the UBL Customisation guide lines, upon an online repository of documents. The tool produces both the spreadsheet and the XML Schema files corresponding to the defined document. The main functionalities of the eDoCreator are:

1. Visualisation of the elements in the repository.
2. Query on the repository.
3. Creation of new documents.
4. Support for conformant and compatible customisation.
5. Data type refinement.

The main aim is to support the creation of new documents, exploiting and maximising the use of the UBL library of common components. In this sense eDoCreator is thought to support the definition and management on new business documents and, considering the standard management life-cycle, it faces the first procedures in the definition of a common interoperability business standard.

- CAM, Content Assembly Mechanism. CAM provide an open source system based on XML that allows the definition and validation of documents through the reuse and composition of general validation rules defined in general documents. These rules are collected in a CAM template that represents the final desired document structure. The CAM

architecture allows the definition of a very specific, personal document structure, but it is not based on a schema language like the aforementioned ones.

4. A Customisation Environment for Standardised Data Models

We prefigured in [7] an approach for standard definition and adoption that consists basically in three steps:

1. The definition of general specification
2. The definition, starting from the general specification, of a sectorial specification
3. The customisation of sectorial specification to satisfy enterprise peculiarities

Respect to this approach, our aim is to complete and to integrate the ending phases of this process adding a customisation environment that, exploiting co-constraint definition, can definitively guarantee a complete specification of all the data format constraints that reflect the enterprise informative system needs. In this perspective, the idea is principally to support the final users of a standard in its adoption and customisation.

In this activity we focus on XML standard documents that exploit XML Schema for the definition of document templates; in particular we adopt the XML Schema 1.1 specification, that represent a powerful, promising and wide accepted schema language.

On the other hands, since there are many design mechanisms for a XML Schema template, the architecture and structure of the template are very important for the its management. In this paper we present two different applications for the customisation of XML Schema document.

1. The Custom XSD[13], that can manage a specific set of XML Schema
2. The Custom UBL[12], that has been implemented to support the management of the UBL Schema documents and to face different and very relevant kinds of problems related with the peculiarities and complexities of this standard. The structure of an UBL schema document will be discussed in order to highlight some problems and to describe the adopted solutions.

CustomXSD and Custom UBL are both web applications for the consultation, the exhaustive navigation and customization of XML Schema documents through the definition of simple constraints and co-constraints (XML Schema 1.1). These applications allow to obtain from an initial document schema another restricted document schema, but consistent with the initial document. It is therefore particularly useful for defining customized documents without losing the conformity with the standard. CustomXSD and Custom UBL allow non-expert users to generate custom XML Schema documents (which can be created from existing XML Schema representing industrial standards) in a way that is independent from the syntax of the language.

The constraints that can be added through the two applications are restrictions than those already existing in the initial document. This means that an XML document valid for the XML Schema created by CustomXSD or CustomUBL remains valid according to the initial XML Schema. This characteristic is important to not generate documents inconsistent with the original standard documents.

The users of the application must have a good knowledge about the concepts related with the domain of their expertise, but they could not have knowledge about the structure, syntax and features of XML Schema. The information are visualised, managed and edited in an abstract and user-friendly way. In comparison to existing applications, CustomXSD aims to improve the ease of use (even for the least experienced user), the method of visualization, navigation and manipulation of documents, the dynamic components, the high level of interactivity and especially the opportunity to express constraints and co-constraint through a simple interface.

4.1 Custom XSD

The application allows loading an XML Schema document (in Venetian Blind model) via URL. After the customization, it produces the following files that can be downloaded locally:

- A custom XML Schema 1.1;
- A XML document for documentation purpose. This document collects and describes the changes made to get the customized document from the base document;
- An HTML page which displays the information in the XML document through an organized and intuitive representation of them.

4.2 Custom UBL

The application allows loading an UBL Schema document (maindoc). After the customization, it produces the following files that can be downloaded locally (Figure 1):

- A custom UBL schema document (maindoc);
- The custom UBL schema documents (common) imported from the maindoc;
- A XML document for documentation purpose. This document collects and describes the changes made to get the customized documents from the base document;
- An HTML page which displays the information in the XML document through a organized and intuitive representation of them.

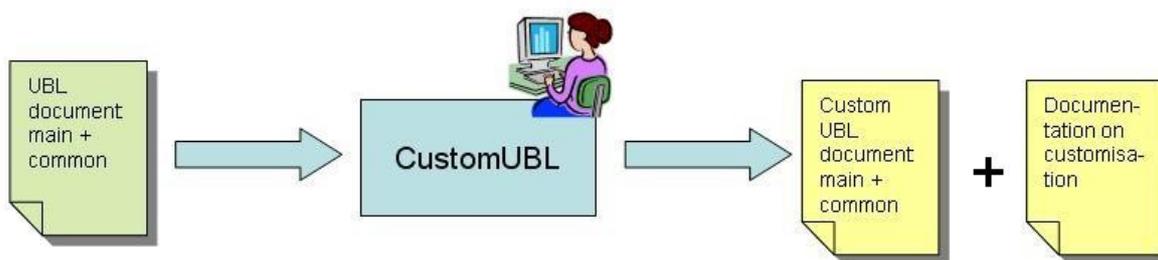


Figure 1

For example, an enterprises that wants to send and receive a business order in a digital format (toward a partner that already manages this kind of data exchange using the UBL standards), can with the CustomUBL easily analysis and customize the UBL order document, adapting it to its needs. The result will be a XML Schema that represents both the requirements and the ability of the enterprise in exchanging a digital order document, and that can be used to regulate and validate such exchange. From now on, the two enterprises can share a common vision of an UBL order documents that can adopt to start managing commercial orders with digital data exchange.

CustomUBL has efficient search functionality, essential for the navigation of UBL documents which are characterized by a very large number of elements. With this search functionality the user can get all the possible paths in which an element can appear and navigate the document using the path obtained through the research. You can also search all the items whose name contains a specific substring. For each item found is reported the number of occurrences, and you can get all the possible paths. Moreover, the search functionality exploits the semantic description that the UBL documents provide in the tag “xsd:documentation”, particularly in the tags “ccts:ObjectClass”, “ccts:PropertyTerm”, “ccts:RepresentationTerm”. With these information, the user does not need to know exactly the name of an element, but can also search in the document for an element knowing its way to use.

Figure 2 shows a screenshot taken from the CustomUBL: the interface presents on the left a navigation tree that reflects the XML structure defined by the selected XML Schema. When the user selects an element on the tree, the information about the element are visualised in a tab panel in the right part of the interface. The user can select as many elements as he needs, and all their information are loaded in tab panels. The panels are filled with the description of the elements, and the available parameters for the elements, with their values.

The user can than modify the selected element, fixing the conditions and the constraints for the element. All these operations are made without writing any XML Schema Code. The Custom XSD provides an interface very similar to that of the CustomUBL.

It is worth to note that, cause of the characteristics of the structure of an UBL template, the use of co-constraints is the only way to customize this kind of documents, in terms of data type characteristics. In fact UBL documents are written using the Garden Of Eden programming styles and then both elements and types are declared globally. With this structure every change on a particular element on the document tree impacts on every other element of the same type in the document. In other words, with the adoption of co-constraints, CustomUBL allows not only to modify global elements or types, but also to modify only a selected element in a specific, unique xpath. For example, with our approach, the application allows the user not only to fix the minOccurs and maxOccurs parameter for a single element (that are the only two parameters that are fixed in the tag <element>) but it is possible to modify the data type and to impose context constraints on it.

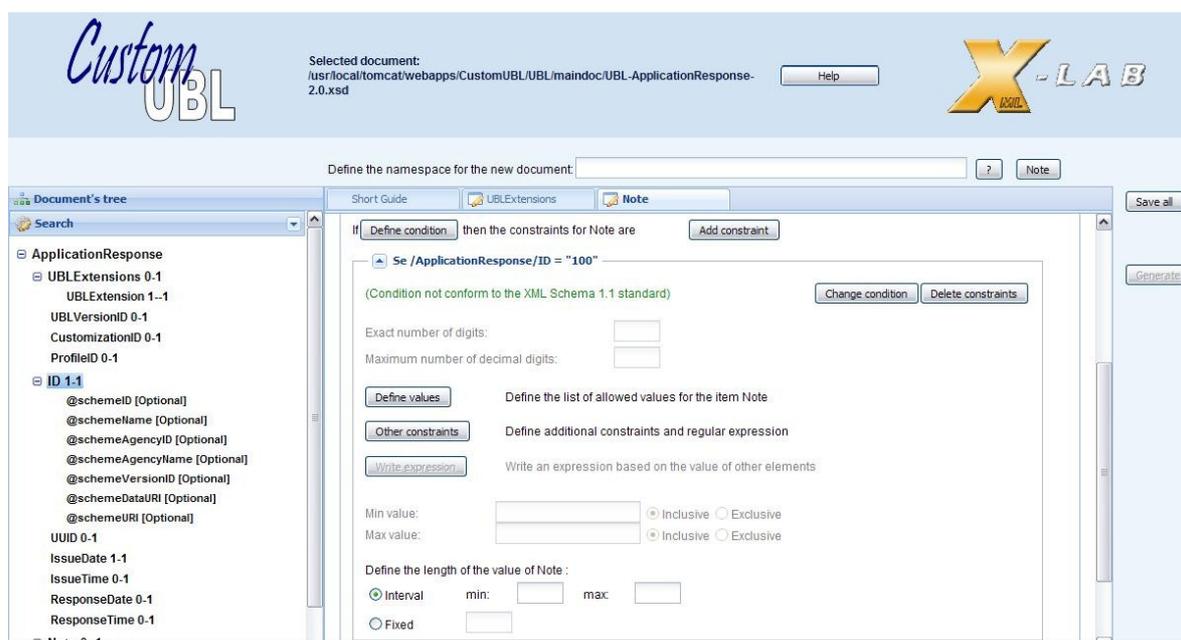


Figure 2- Interface of the CustomUBL

5. Conclusions

In this paper we present a suite of applications for the customisation of XML Schema documents, in order to simplify the building of enterprise document profiles, exploiting the expressivity of co-constraints (thus, beyond XML Schema 1.0 expressivity).

Our aim is to improve the possibilities, for a user who needs to adopt an e-Business standard, to customize and tailor on proper specific requirements a standard conformant data format that is suitable for a use context strictly related with enterprise characteristics and business peculiarities. This objective has clearly emerged from the experience of the

campaign of 17 pilot networks in the eBIZ-TCF project [5] and from the work presented in [6].

We have in particular tackled the complexity of UBL business documents, which are based on a huge set of components that represent an obstacle for non expert users in the setting up of interfaces between enterprise systems and the standardised documents.

Both for CustomXSD and CustomUBL, the benefits of these approaches appear to be strictly related with:

- greater expressiveness of the customisation.
- user orientation of the interface for document management, that presents documents in a simple readable structure that hides the complexity of the XML Schema language and the logical model that underlies the design of the documents.

Besides, the two applications support the users providing two key features for building conformant customisation:

- check the conformity of the documents
- provide a wide documentation about the modification performed.

References

- [1] Bosak J., McGrath T., Holman G.K., “Universal Business Language v2.0, Standard”, OASIS Open, 12 December 2006; <http://docs.oasis-open.org/ubl/os-UBL-2.0/>
- [2] OASIS Content Assembly Mechanism Technical Committee, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cam
- [3] Schematron: <http://www.schematron.com/>
- [4] XML Schema 1.1: <http://www.w3.org/TR/xmlschema11-1/>
- [5] eBIZ-tcf: <http://www.ebiz-tcf.eu>
- [6] N. Gessa, C. Novelli, M. Busuoli, F. Vitali, "Use and extension of ebXML business profiles for Textile/Clothing firms", proceeding of 5th International Conference on Electronic Commerce and Web Technologies, EC-Web 2004, p. 186-195, August 30 - September 3, 2004, Zaragoza, Spain, in LNCS 3182 Springer-Verlag, Berlin, ISBN 3-540-22917-5.
- [7] A. Brutti, V. Cerminara, S. Cristiani, P. De Sabbata, N. Gessa, “Use Profile Management for B2B Standard Adoption: The Case of eBIZ-TCF”, in proceedings of e-Challenges 2009 conference, Istanbul, Turkey, October 21-23 2009, edited by Paul Cunningham and Miriam Cunningham, IIMC International Information Management Corporation LTD, Dublin, Ireland, IOS PRESS, ISBN 978-1-905824-13-7
- [8] <http://ubl.xml.org/product/srdc-ltd-isurf-edocreator-10-beta>
- [9] F. Tuncer, A. Dogac, Y. Kabak, Ş. Postaci, S. Gonul and E. Alpay, “iSURF eDoCreator: e-Business Document Design and Customization Environment”, in proceedings of e-Challenges 2009 conference, Istanbul, Turkey, October 21-23 2009, edited by Paul Cunningham and Miriam Cunningham, IIMC International Information Management Corporation LTD, Dublin, Ireland, IOS PRESS, ISBN 978-1-905824-13-7
- [10] RelaxNG: <http://www.relaxng.org/>
- [11] UN/CEFACT Core Components Technical Specification, Part 8 of the ebXML Framework in Version 2.0. http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf
- [12] CustomUBL: <http://www.lisealab.it/?li=en&p=354>
- [13] CustomXSD: <http://www.lisealab.it/?li=en&p=322>