ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2008, Vol. 37, No. 1

UBL: THE DNA OF NEXT GENERATION E-BUSINESS

Karsten Tolle

Institut für Informatik, Datenbanken und Informationssysteme (DBIS)
Johann Wolfgang Goethe-Universität Frankfurt am Main
D-60325 Frankfurt am Main

Abstract. This paper introduces into the evolution of Electronic Data Interchange (EDI) and the Universal Business Language (UBL), an OASIS standard to encode and customize business documents. It shows its peculiarities and also sets it into a broader picture showing where UBL is positioned in relationship to business processes and standards like BPEL and BPMN.

Keywords: Universal Business Language (UBL), Electronic Data Interchange (EDI), e-Business.

1. Introduction

Raising salaries and market demands are pushing more and more Small and Medium Enterprises (SMEs) to automate their business processes and to switch to e-Business solutions. For those SMEs complex EDI standards like X12 or EDIFACT are not sufficient. On top there are different other barriers that need to be overcome [TKJ+07]. With the Universal Business Language (UBL) a very simple and manageable set of business documents and the possibility to customize them to the SME needs is available.

This paper gives an introduction to UBL, its history and ancestor standards. Section 2 concentrates on the evolution of EDI since 1960 while section 3 introduces the new UBL 2.0 standard. UBL 2.0 can be viewed as the latest deployed EDI standard and some issues are still under development. In section 4 UBL is compared to DNA of living cells in order to explain its benefits. Section 5 places document standards like UBL in a broader scope including business process and standards for it. Finally section 6 closes this paper by mentioning some UBL projects and related work.

2. Evolution of EDI Standards

Some people only think of the older antiquated standards when they hear Electronic Data Interchange (EDI). This is not true. EDI[1] defines *an asynchronous exchange of structured information between computer applications that can be processed without human interaction*. This definition still fits to new approaches in this area.

However, the first approaches to EDI are going back to 1960. During that time the information structure and the way to exchange (including the infrastructure for it) needed to be set up by the parties themselves. Only in the late 70th first standards appeared and some are still used today. For example, *VDA 4905* which was first released in 1978 and is still used by Audi, BMW and Miele. It was designed to support just-in-time production and therefore the supply chain management. In 1982, the *American National Standards Institute Accredited Standards Committee* (ANSI ASC) ratified the first version of X12. X12 is still one of the leading standards for EDI in northern America. Based on X12, the United Nations within the CEFACT (Centre for Trade Facilitation and Electronic Business) developed the first version of UN/EDIFACT (*Electronic Data Interchange For Administration, Commerce and Transport*) in 1988. EDIFACT is still very successful, however, due to its complexity mainly big companies adopted it. To reduce this complexity and to adopt specific domain demands, also numerous domain specific subsets were generated.

EDI was supported by the evolving Internet. The Internet together with standardizes exchange protocols serves as an infrastructure layer. Cost and maintain intensive Value-Added-Networks (VANs) where no longer needed [EFS+05], which resulted in a higher distribution of EDI solutions. Today the highest cost factor is the implementation of interfaces and their maintenance between the different parties [KWH97, TK05].

Due to these reduced set-up costs, high salaries and therefore potential cost reductions by automation, the companies are trying to automate their complete

[1] Detailed definition of EDI by ANSI can be found at: *http://www.x12.org/x12org/about/faqs.cfm#a1*

business process cycles (including the processes between the companies). To support this e-Business Frameworks, like RosettaNet and ebXML appeared. Both frameworks define a set of standards and rules to describe Business Processes (BP) and Service Level Agreements (SLA). They also provide standards on technical level to ensure security and reliability. It is important to note that both frameworks are not exclusive. In fact, they can be used complementary. While on one hand side *Electronic Business using eXtensible Markup Language* (ebXML) concentrates horizontal on the different aspects of e-business, RosettaNet, on the other hand, concentrates vertical and can be used to solve domain specific problems [Ba06].

Both frameworks make use of the eXtensible Markup Language (XML). Since its specification in 1998 it penetrated many areas and sometimes it is even said: „XML is ASCII for the 21st century“ [Th01]. By the use of XML with EDI it obviously profits by better human readability (compared to X12 or EDIFACT documents) and the existence of many standard tools supporting XML. Especially the possibility to validate XML instances against Document Type Definition (DTD) or XML-Schemata (XSD) that define the structure of a document reduces the number of possible errors. There could other formats be used, however, XML was optimized for this purpose [FGN+02]. The only drawback of using XML is the document size that is relatively high – up to eight times – compared to X12 or EDIFACT documents. However, there are compression algorithms for XML around in case storage or bandwidth is critical[2].

In contrast to ebXML, RosettaNet provides also business document descriptions. However, due to the absence of XML Schema at the time RosettaNet was created, these descriptions are based on the older and less flexible DTDs. Newer standards and approaches are using XML Schema, where they have support for various data types and the possibility to modularize documents using namespaces and *include*-instructions, just to mention the two main benefits of XML Schema against DTDs. One of these newer standards is UBL that will be explained in more detail within the next section.

3. Universal Business Language (UBL)

The Universal Business Language (UBL) was released as version 1.0 by OASIS in 2004 as a further development of the Common Business Language (CBL). In December 2006 version 2.0 was approved as a major release and is the current actual version. UBL 2.0 defines 31 business documents for the business areas: Sourcing, Ordering, Invoice and Fulfillment. These documents are built-up out of a library of XML schemas for reusable data components. Figure 1 shows the evolution of different EDI standards that were mentioned in the previous section together with the UBL history.

It is important to mention that UN/CEFACT supports UBL 2.0. In April 2006 they agreed with OASIS on a memorandum of understanding where CEFACT accepts UBL 2.0 as “first-generation XML documents for e-Business”. Future developments of EDIFACT will be based on UBL 2.0. In return, OASIS agreed to keep UBL stable for at least the next three years. This means there might be minor releases providing UBL 2.x, but no major release like an UBL 3.0. UBL will still officially be hosted by OASIS, however, UN/CEFACT will have the according rights to generate UBL extensions [Gr06]. It therefore can be assumed that future releases of UN/EDIFACT will be based or at least close to UBL and a switch between these standards should be easily possible.

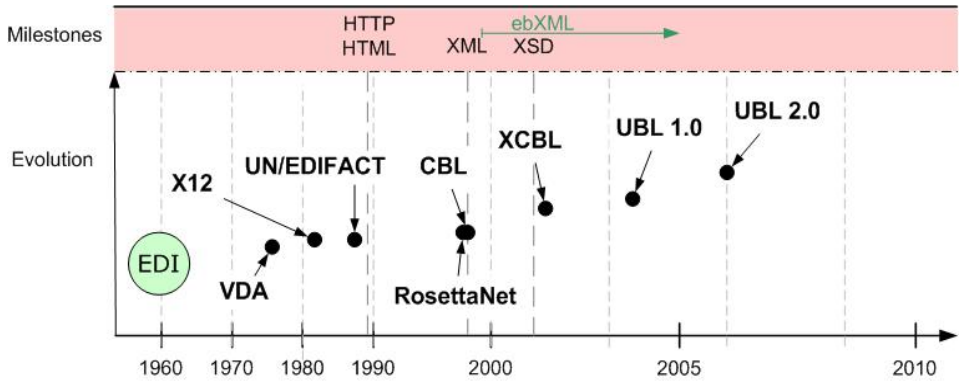

Figure 1. Evolution of B2B document standards

The key features of UBL are :

1. Simplicity – UBL provides the most important documents that should serve for 80% of all business cases and still remains manageable.
2. Customizability – UBL is generic. It therefore provides the possibility to confine, extend or even to create your own documents. This will be further explained below.

As mentioned, the UBL documents are built out of a library (called *Common Library*) that is also part of the UBL specification. This library consists of compo-

[2] See also Cover-Pages: XML and Compression at *http://xml.coverpages.org/xmlAndCompression.html*

nents called *Business Information Entities* (BIE). Some of them are aggregated (Aggregated BIE) out of others while others are atomic and are called *Basic BIE*. This way UBL is the first specification that implements the *Core Components Technical Specification* (CCTS) of ebXML.

Using these modules users have the opportunity to generate their own documents in case the existing ones are not sufficient. As long as the BIEs of the new document are derived from those existing ones, these documents are called *UBL Compatible*.

In case there are only small extensions to existing UBL documents needed one can also make use of the extension-element that is included to each UBL document at the very beginning (which is new to UBL 2.0). Within this extension-element the user is free to enter anything, as long as it is valid XML. This way a validation against the according UBL XSD can be performed. Any instance that passes the according UBL XSD validation is called *UBL Conformant*.[3]

There exist different UBL localization subcommittees for different nations. These subcommittees are working on customizing UBL documents reflecting the specific needs for their nation. This includes possible extensions but mainly confinements of the XSDs within the given borders of the UBL XSDs. For example, changing the cardinality of an element from 0..n to 0..5 due to the fact that in this country the maximum number of it can be five. By those confined XSDs errors can be easily detected by validation against it. However, any of these localized documents will be valid against the standard XSDs (but may be not the other way round). In addition to these customizations, the localization subcommittees provide a translation for the different UBL components and their description that will be collected into one so called International Data Dictionary (IDD).

With these two features, being simple and customizable, UBL is extremely expressible. However, for interoperability, especially in the business area where contracts are made and money is involved, it is important also to understand the semantic behind the documents and their fields. This is something that is not fully provided and also not intended to be provided by the UBL TC. Of course, there is a rudimentary description given but the clear semantic determination is up the context of the document usage. This semantic determination may depend on the nation, domain or special agreements of the trading parties. Even though there is this interoperability problem, we try to explain in the next section why UBL is a step in the right direction.

[3] Note: This terminology is currently under discussion at the UBL TC. Changes might happen.

4. Comparing UBL 2.0 and DNA

For the developers of UBL under the lead of Jon Bosak, who is one of the fathers of XML, it was clear that there can not be one solution that fits any scenario and they did not try to develop the philosopher's stone. With UBL they tried to do one step in the right direction in order to achieve some improvements. In this section, we will try to explain the design goals and intended improvements as we see them of UBL by comparing it with deoxyribonucleic acid (DNA) that can be found inside living cells and even in some viruses. The DNA builds the genes and chromosomes that made their way through the evolution, comparable to business documents that already existed during and before the times of the old Egyptians.

In nature we can find an impressive variation of creatures and plants. However, all of them are using DNA to encode information based on the bases Adenine, Cytosine, Guanine and Thymine. The same techniques are used to decode or replicate them in all living cells. This can be compared to the usage of XML that is used to encode information.

Bigger information blocks are packed into chromosomes. By outside effects like UV light there are up to one million molecular lesions (small structural errors). There are different processes running within a cell repairing these errors. In the same manner XML Schema structures information and helps to avoid errors by validating instances against it.

When we look only at genes and just concentrate on one group, e.g. mammals, there is a very high level of accordance. One can compare DNA by bases Adenine, Cytosine, Guanine and Thymine and there will be only very few differences between humans and primates. Funny wise human genes are more equal to genes of a rat or a mouse than to dogs [TTB+03]. However, a molecular biologist nowadays can exchange gene parts of one species with another one. This is only possible because the gene/DNA structure is comparable including the semantic level. The same accordance we would have between different UBL customizations.

As a molecular biologist knows the DNA building bricks, UBL experts know the building bricks of a UBL Compatible instance and even better the structure of UBL Compliant instances. Therefore, the same way as a molecular biologist expert for chimpanzees, would understand the genes of a human and probably also the genes of a mouse, the same way UBL experts could change between the domains.

Furthermore any biologist knows the basics about DNA and its mechanisms. In the same manner computer science alumni know about XML and XML Schema. They can easily put their hands on UBL. Therefore, a high number of experts will be available for UBL in future. It will also be much easier for them to change between different domains than it is nowadays when changing between EDIFACT subsets.

However, we also need to accept that business processes are complex and have as many facets as we also find them in nature. We currently – and probably also in future – can not expect to find solutions and tools fitting all existing situations. UBL is customizable and can therefore be adopted to the needs and to changes. As Charles Darwin said: "*It is not the strongest of the species that survives,nor the most intelligent that survives. It is the one that is the most adaptable to change*."

5. UBL in broader context

In Section 2 already the two e-Business Frameworks ebXML and RosettaNet were mentioned dealing with many issues outside the document encoding itself. A more advanced introduction to such frameworks and their evolution can be found in [SJS+06]. In contrast to these frameworks, UBL only deals with the encoding of the business documents and their information. However, there is a strong relationship of business documents sent and the corresponding BP. For this reason, one will find various business processes defined as example usages for the document within the UBL specification[4]. In fact, the *collaborative BP* – hence the BP that is between the parties – can be viewed as being represented by the possible documents that can be exchanged. To my own experience, this view is very intuitive and easy to follow.

ebXML can be used to handle and define these processes and is specialized for the e-Business area. In addition, there is the *Business Process Execution Language* (BPEL) that is more generic and, in fact, was generated to orchestrate Web Services. Even though there are certain drawbacks of BPEL compared to ebXML [Du04], the current hype on Web Services and the support by big companies pushes BPEL. The funny thing about it is that both standards as well as UBL are hosted by OASIS (ebXML was created together by UN/CEFACT and OASIS).

In order to write and understand BPs there also exist a standard notation for them called *Business Process Modeling Notation* (BPMN) that is hosted by the Object Management Group (OMG). For BPMN there exists a mapping into BPEL. However, BPMN is independent of it and has a wider scope and expressiveness. Therefore, one can not translate any BPMN into an executable BPEL. Especially regarding human interactions BPEL is limited, since it depends on callable services. For BPMN, in fact, the creators of it stated in their FAGs that some features of BPMN might not be able to be implemented by any technology.

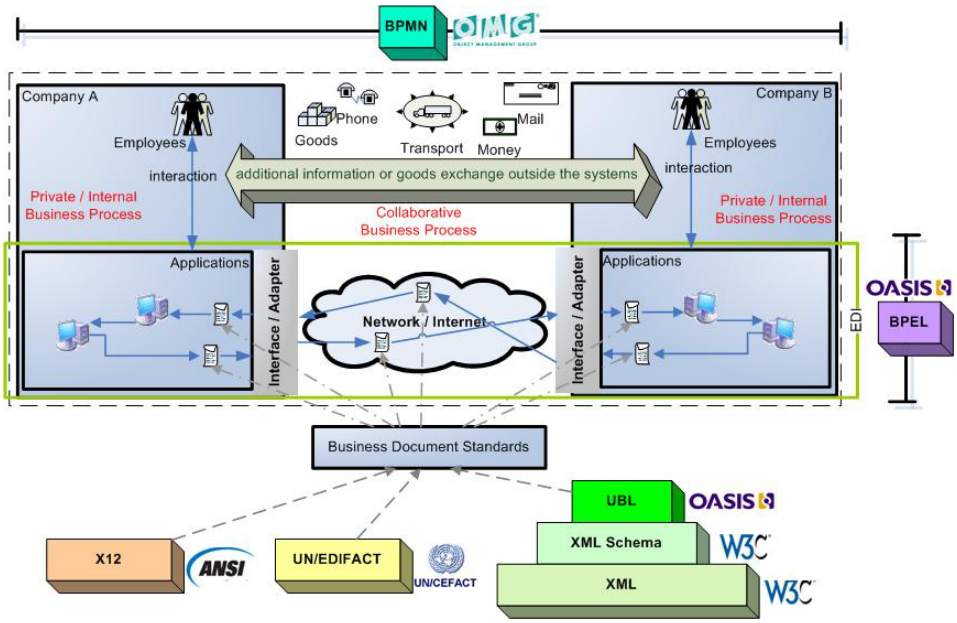

Figure 2. Top level view showing different BP types, their relatinships and related standards

In order to implement an interface for a company that allows the automatic trade with another one, the collaborative BPs need to be matched with the private internal BPs and with the applications of the company. Therefore interface implementation enforces the understanding of all related components: a) to be exchanged document by structure and semantic, b) collaborative BP and c) private internal BPs.

One of the very few free tools supporting the understanding of the documents that are exchanged is called Jdeluxe [Me07], an Open Source Eclipse Plugin for exploring and document XML B2B standards. This way also the semantic determination of the UBL fields that is not given by the standard itself, can be agreed on and understood between the parties in a collaborative way.

[4] UBL 2.0 specification can be downloaded at: http://docs.oasis-open.org/ubl/os-UBL-2.0.zip

6. Related Work, Conclusion and Outlook

One of the driving examples for UBL usage is the Danish Government e-Invoice system and their new approaches with the so called OIO Service Oriented Infrastructure (OIOSOI) [OIO06, BL06]. The Danish Government needs to take SMEs into account and special projects like the EU-Projects ABILITIES [TKJ+07] or GENESIS [SPJ07] even focus on SMEs and try to reduce entry barriers by taking advantage of UBL's simplicity, combined with different approaches. However, it will take some time until big vendors like SAP, BEA or IBM will start to concentrate on SMEs. For them the needed effort to handle the peculiarities of SMEs is too high and their return of investment on big companies is still more lucrative. Hopefully, this will change by the efforts and results generated by the mentioned projects.

This does not mean that existing and running solutions need to be changed. Think of VDA 4905 going back to 1978 still being in use. As long as the system is running and no change is needed, there would be no need to do so.

However, for the most SMEs that are starting now with EDI, the standards X12, EDIFACT or even older ones are no option due to missing or expensive experts, tools and the complexity behind – some might be forces by big partners they depend on.

The key step is now that communities around UBL determine the semantic of the UBL documents for certain trading networks, like it is currently done by the Northern European Subset (NES)[5].

7. Acknowledgements

The work presented in this paper is partly funded by the European Commission through the ABILITIES STREP (IST-FP6-027306). The author is solely responsible for the paper's content.

References

[Ba06] H. Badakhchani. Introduction to RosettaNet. *http://dev2dev. bea.com/lpt/a/10*, August 2006.

[BL06] M.H. Brun, C. Lanng. Reducing barriers for e-business in SME's through an open service oriented infrastructure. ICEC'06, *August* 2006, *Fredericton*, *Canada*.

[Du04] J.-J. Dubray. Where is BPEL 2.0 going? *June* 2004, *http://www.ebpml.org/bpel_2_0.htm.*

[EFS+05] W. Elgarah, N. Falaleeva, C. C. Saunders, V. Ilie, J.T. Shim, J.F. Courtney. Data exchange in interorganizational relationships: review through multiple conceptuallenses. *SIGMIS Database*, 36(1): 2005, 8–29.

[FGN+02] A.Fairbanks, J. Gribbons, E. Nybo, D. Pean, J. Wright. Research in XML (extensible markup language), *Journal of Computing Sciences in Colleges, Vol.*17, *Issue* 6, *May* 2002.

[Gr06] T. McGrath. UBL and UN/CEFACT – adding sanction to traction. *Presentation at UBL International* 2006, 13-17 *Nov. Copenhagen, Denmark*, *http://www.ublconference.com/200611/programme.html#ublcefac*.

[KWH97] K. Ketler, J.Willems, V.Hampton. The EDI implementation decision: a small business perspective. *In SIGCPR*'97: *Proceedings of the* 1997 *ACM SIGCPR conference on Computer personnel research*, *New York, NY, USA*, 1997, 70–76.

[Me07] D. Meyer. Jdeluxe. *http://www.jdeluxe.org/Joomla/index.php*, 2007.

[OIO06] The Ministry of Science (Denmark), Technology and Innovation: The Architecture of the Danish OIO Service Oriented Infrastructure. March 2006. *http://purl.oclc.ort/NET/SOI/OIOSIarchitecure*.

[SJS+06] C. Schroth, T. Janner, A. Schmidt, G. Stuhec. From EDI to UN/CEFACT: An Evolutionary Path Towards a Next Generation e-Business Framework. *The 5th International Conference on e-Business* (*NCEB* 2006), *Bangkok, Thailand, November* 2006.

[SPJ07] C. Schrot, G. Pemptroad, T. Janner. CCTS-based Business Information Modelling for Increasing Cross-Organizational Interoperability. *I-ESA* 2007, *March 28th - 30th, Madeira, Portugal*.

[Th01] H.S. Thompson. The XML MetaArchitecture. *http://www.w3.org/2001/Talks/0502-ht-XMLArch/, W3C Talk,* 2001.

[TK05] J.M. Tenenbaum, R.Khare. Business services networks: delivering the promises of b2b. *In BSN*'05: *Proceedings of the IEEE EEE05 international workshop on Business services networks,Piscataway, NJ, USA*, 2005, 8–8.

[TKJ+07] K. Tolle, V. Kiauleikis, K. Justinaviciene, G. Knoll, C. Guglielmina, A. Arezza. An Interoperable E-business platform towards better integration of New Member States Small and Medium Enterprises. *I-ESA* 2007, *March* 28*th* - 30*th*, *Madeira, Portugal*.

[TTB+03] J.W. Thomas, J.W. Touchman, R.W. Blackesley et. others. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, *Vol.*424, 14 *August* 2003, 788-793.

Received September 2007.

[5] Information about the NES can be found online under: *http://www.nesubl.eu/*