State-of-the-art in business document standards

Philipp Liegl, Marco Zapletal
Vienna University of Technology
Email: {firstname.lastname}@tuwien.ac.at

Christian Pichler, Michael Strommer
Research Studios Austria
Email: {firstname.lastname}@researchstudio.at

***Abstract*—With the raising significance of electronic commerce and e-government applications the need for standardized business documents has emerged. Today, a business partner seeking to implement a new electronic commerce solution may choose from a multitude of different standards. Most of the standards are domain-specific and thus heterogeneous, since standards are developed out of diverse needs, motivations, and backgrounds. What is still missing is a thorough overview of business document standard families. Furthermore, a classification helping a business partner to choose a specific standard, based on his preferences is needed. In this paper we present the key results of our business document survey, where we made a classification of business document standards using standard categories. For each category we introduce a representative standard example. Based on our research results, a clear and precise classification of different business document standardization approaches is provided. Through the classification, the distinctive advantages and disadvantages of each standard become clear. Thus, the decision which business document standard to use is eased.**

I. Motivation

The importance of compatibility has already been recognized long before the advent of computer systems and networks. In physical networks such as railroads, telephone systems, and electronic distribution networks compatibility is a critical success factor. According to [1] it was Thomas Edison's ability to think in terms of whole systems rather than in terms of single generators, which led to his success in the electricity business. Similarly, successful electronic business transactions between arbitrary business partners can only be established if standardized and interoperable system interfaces are provided. Feng [2] identifies *standardization* as "the process by which explicit specifications for the form or function of a particular technology are created". The result of these specifications are called *standards*. In particular business document standardization efforts aim to overcome data definition heterogeneities between different business partners. In this paper we focus on electronic business document standards and their application in the field of electronic commerce.

Within the last four decades a multitude of different business document standardization initiatives have been found, leading to numerous business document standards. All of these standard definitions aim at facilitating electronic data interchange between business partners. A review of current literature shows, that several surveys on different business document standards have been published [3] [4]. However, several surveys are outdated or cover only a limited set of all available business document standards [5] [6]. Other surveys specifically focus on a defined set of specific standards, without considering categories of related standard families, e.g., publications in the health care domain mostly focus on the Health Level Seven (HL7) standard.

In this paper we give a contemporary overview of different business document standards used in today's electronic commerce applications. We do not concentrate on a specific set of standards, but introduce a well defined set of standard categories. For each standard category a representative example is introduced. Finally we provide a comparison of five selected clusters according to compatibility to business messaging, technical features, potential user groups as well as acceptance criteria. The authors have several years of extensive experience in different standardization committees and initiatives. Thus, we evaluated the results of our study based on our experiences and findings in the area of business document standardization.

The remainder of this paper is structured as follows: Section II introduces the defined business document standard clusters together with a representative example from each cluster. Section III provides an evaluation of each standard cluster in regard to technical implementation, business messaging conformance, as well as target groups and acceptance criteria. Section IV concludes the paper with an outlook towards our future research work in the domain of business document standards.

II. Overview of business document standards

As outlined in the introduction, several business document standardization initiatives have developed a multitude of different business document standard definitions. Figure 1 gives an overview together with a timescale of the most important standard definitions.

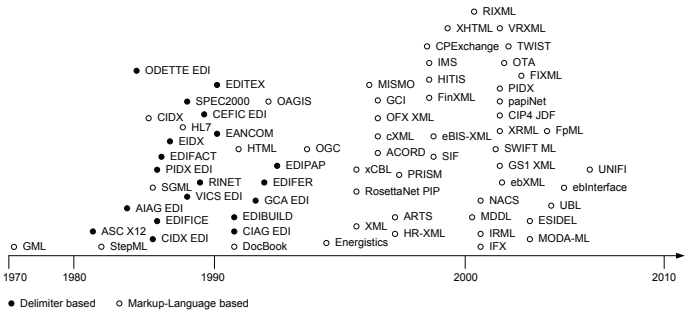

Fig. 1: Overview of different standards

We identify two major groups of business document standards: delimiter-based languages and markup-based languages.

Delimiter-based approaches use standard ASCII characters to separate different data elements, segments, and messages [7]. The two most important delimiter-based standard definitions are UN/EDIFACT and ANSI X12. EDIFACT-based standards were particularly developed in the eighties and early nineties of the last century as outlined by the cloud of black dots on the left hand side of Figure 1.

An important development for the domain of business document modeling was IBM's Generalized Markup Language (GML), developed in the 1960s. GML later was the basis for the development of the Standard General Markup Language (SGML), which was the basis for two other prominent markup languages - Hypertext Markup Language (HTML) and eXtensible Markup Language (XML). In particular the development of XML in the late 1990s revolutionized the way how business document standards were developed. From this point on almost all of the known business document definitions were based on XML. In Figure 1 markup-based document standards are denoted by white dots. White dots in the cloud of black dots on the left hand side of Figure 1 are either standards based on general markup languages (e.g., SGML) or standards which were initially defined in a delimiter-based manner, but were later transferred to an XML-based syntax, e.g., Health Level Seven (HL7) or the Chemical Industry Data Exchange Standard (CIDX).

In our paper we go beyond a simple classification of standards into delimiter and markup-based languages, but provide a classification of different standard clusters. For each cluster we evaluate one representative standard example. As shown in Table I we identify eight standard clusters along with a representative standard example.

(1)	Top-down standard approaches	UN/EDIFACT
(2)	Bottom-up standard definitions	ebInterface
(3)	Hybrid approaches	Universal Business Language (UBL)
(4)	Early markup adopters	Open Applications Group Integration Specification (OAGIS)
(5)	Integrated approaches	Electronic Business XML (ebXML)
(6)	Transitioned approaches	Health Level Seven (HL7)
(7)	Implementation neutral approaches	Core Component Technical Specification (CCTS)
(8)	Converging approaches	Universal Financial Industry Message Scheme (UNIFI)

TABLE I: Identified standard families

Each of the clusters represents a family of related business document standards, categorized according to their technical features. Business document standards, due to their diversity, may be included in several clusters at the same time. In the following the different clusters, together with an accompanying example are outlined.

a) Top-Down standardization approaches: Top-down standard definition approaches aim at the inclusion of as many different requirements as possible. Stakeholders of the standard such as industries, interest groups as well as individuals submit their requirements to a standardization body maintaining the standard definition. As shown in Figure 2, a top-down business document standard definition is a union of different requirements. These requirements are represented in core standard documents, which essentially are a superset of all the stakeholders' different requirements. Due to the inclusion of a multitude of different requirements, top-down document definitions contain a lot of different, often optional elements and are thus complex.

Fig. 2: Top-Down standardization

UN/EDIFACT. The United Nations Electronic Data Interchange for Administration, Commerce, and Transport (UN/EDIFACT) is an international EDI standard developed by the United Nations. The standard itself consists of syntax rules, used to structure data (referred to as EDIFACT) and standard messages stored in directories, allowing cross-country and cross-industry data exchanges (referred to as UN/EDIFACT directories). Generally, the standard itself is referenced by the term UN/EDIFACT.

Top-down standards provide a generic concept for the representation of business documents. Thus, a potential user may easily find the required business messages and elements in the standard definition. Furthermore, top-down standards such as UN/EDIFACT are well accepted and have been tested and evaluated thoroughly during the last 25 years of application.

However, our research has also shown that top-down definitions are complex and business service interfaces capable of processing top-down business document messages are costly to implement. Thereby, we define a business service interface as a piece of software processing incoming business document messages and passing the required information to the internal software of a company. In most cases only big players can afford to establish such business service interfaces. Small and medium-sized enterprises often lack the architectural prerequisites to implement such solutions.

We conclude that in general all of the different EDIFACT-based standard definitions follow a top-down principle (ANSI X12, ODETTE etc.). Implementation neutral approaches such as the Core Components Technical Specification (CCTS) also follow a top-down approach.

Although they are powerful and generic, top-down standards may require a considerable effort for the realization of a compliant business service interface. In our study we observed an upcoming trend in recent years towards the use of bottom-up standard definitions.

b) Bottom-Up standard approaches: As shown in Figure 3, bottom-up approaches are defined by the inclusion of the most important elements, which may be used in an electronic business transaction. Instead of defining a superset of required elements, a bottom-up standard comprises a core set of elements. In order to provide support for elements which are not part of the core standard definition, the concept of extensions is used. Extensions are plugged into the core standard definition without altering the core standard. Depending on the specific

standard, the different extensions are either supervised by the standardization organization or it is left to the users to define their own customized extensions.

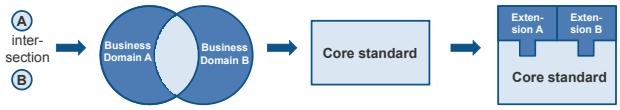

Fig. 3: Bottom-Up standardization

ebInterface. As an example for a bottom-up standard we introduce ebInterface, an XML-based standard for electronic invoices. The ebInterface standard is the result of a joint effort started by AustriaPRO, an association affiliated with the Austrian Federal Economic Chamber. Several Austrian ERP vendors and other stakeholders agreed upon a common electronic invoice standard.

As a bottom-up approach ebInterface is less overloaded and ambiguous than top-down approaches. Consequently, implementation as well as communication costs are low. Thus, in particular small and medium-sized enterprises are the major target group of bottom-up standards.

The ebInterface standard covers only a rudimentary set of elements and thus needs to provide some extension mechanism to support additional use cases as well. Currently, the standard contains a custom section where any user-specific content may be added. However, since this extension point is not standardized, interoperability issues as well as adaptation problems may occur. Consequently, current research focuses on the definition of well defined extension mechanisms to overcome uncontrolled growth of extensions.

Another example for a bottom-up approach is the electronic payment standard (EPS), which is used as a simple and secure payment method in online transactions. The basis for this standard is formed by the Electronic Payment Initiator standard (ePI) [8], specified by the European Committee for Banking Standards. Note that EPS comprises only those elements needed for national transactions and thus uses a subset of ePI.

Bottom-up standards are in particular useful for smaller companies, since they require significantly less implementation effort than top-down approaches. In general bottom-up approaches are a rather new concept and thus not so many standards are available at the moment.

c) Hybrid standardization approaches: Hybrid business document standards have been developed in order to merge the advantages of top-down and bottom-up standardization approaches. On the one side they provide the generic standard base of top-down standards, since multiple requirements of different stakeholders are included in the standard definition. On the other side they have extension points like bottom-up standards and thus domain-specific amendments may be added to the standard. Well known examples for hybrid standardization approaches are the Universal Business Language (UBL) and Human Resources XML (HR-XML).

Universal Business Language (UBL). In our survey we introduce UBL as an example for a hybrid standard. The goal of UBL is the definition of a free library of standard electronic XML business documents. Currently the UBL standard contains 31 standardized business documents. UBL pursues a top-down definition approach, i.e., an inclusion of as many different elements as possible. Additionally, UBL contains a mechanism for defining domain-specific standard extensions through a container element named *UBLExtensions*. Within this optional container any non-UBL element may be incorporated.

UBL has a set of advantages emanating from the fact, that the standard specification has been based on the data definition approach from the Electronic Business XML (ebXML) initiative. UN/CEFACT (United Nations Center for Trade Facilitation and Electronic Business), one of the co-founders of the ebXML initiative, and the UBL technical committee agreed on a common strategy, where the UBL standard will eventually be integrated into UN/CEFACT's standard family. Thus, in the long-term UBL provides a robust foundation of predefined core component elements, based on UN/CEFACT specifications. Core components will be elaborated in the Section on implementation neutral approaches. Additionally, in case user-specific requirements have to be integrated into a UBL-based message, UBL's flexible extension mechanism may be used. Thereby, UBL manages to overcome the often criticized shortcoming of top-down definitions, where no user-specific extensions are possible.

However, our research has shown that UBL has a set of shortcomings as well. As Ken Holman pointed out in an email to the UBL developers list [9], the current version of the UBL purchase order covers 830,338 different elements in context and 2,171,455 attributes when flattening the document structure and taking the combinatorial issues of qualified elements into account. Thus, the standard in general has a quite overloaded structure, similar to top-down standards.

Other standard definitions such as Health Level Seven (HL7) and Human Resources XML (HR-XML) may also be accounted as being hybrid standardization approaches. Both standards aim at a union of requirements, which is clearly a top-down feature. Additionally, the standards provide extension points for user-specific amendments, which is a bottom-up feature.

For this standard cluster we conclude that hybrid business document standards provide a good compromise if a generic standard is required which still allows domain-specific amendments. Nevertheless, the implementation of interfaces for hybrid standards may cause considerable coding efforts.

d) Early markup adopters: As shown in Figure 1 the 1990s brought a strong transition from EDIFACT-based standards to markup-based standard solutions. Of particular interest to our survey are the early adopters of the markup technology - those who started to use XML for their business document standard definitions first. Since the early markup adopters were among the first to employ the newly created XML specification, we are going to examine what kind of leverage effects the early adoption had.

OAGIS. OAGIS is developed by the Open Applications

Group (OAGi) and was inaugurated by multiple major companies, most of them from the IT sector. It was one of the first approaches using XML for defining business document standards. The main target of the approach was an optimization of the integration of applications, both inside of a company and between different enterprises. This was achieved by crafting standards where necessary and by recommending standards where they already existed.

According to the OAGIS standard, all participating applications communicate by sharing Business Object Documents (BOD). In addition, the specification defines certain business scenarios, identifying the integrated business applications and components and the BODs that are used.

One major advantage of the OAGIS standard is the variety of companies that are members of the OAGi. In fact, in some business sectors OAGIS has reached a high level of adoption. Since OAGIS was among the first markup language adopters, the standard has considerable popularity among several industry users. For example, in the automotive industry OAGIS has become an important business language, since it is used and promoted by many leading companies in this sector in the US.

Although OAGIS is very popular in certain sectors, it is still not widely accepted in other business sectors. Furthermore, the implementation of OAGIS technology, in particular in regard to Business Object Documents (BOD), requires substantial efforts. Thus, we conclude that in the field of business document standardization an early appearance on the market does not necessarily lead to a high rate of adoption. Nevertheless, some benefits may be leveraged as the example of OAGIS in the automotive sector has shown.

Another initiative which used XML at first was the XML Common Business Library (xCBL). Meanwhile, the initiative has been ceased and the xCBL efforts have been integrated into UBL. Furthermore, the Chemical Industry Data Exchange Standard (CIDX) and the Health Level Seven (HL7) initiative were among the first users of XML.

We conclude that there is no significant gap in regard to the acceptance of a standard between standards which adopted XML very early and recently developed standards.

e) Integrated standardization approaches: One of the lessons learned from the EDIFACT initiatives was the fact that finding an agreement on the exchanged data only, is not sufficient to establish automated electronic transactions between enterprises. Businesses must also agree on a common process choreography, before they are able to engage in automated business transactions. Thereby, a process choreography comprises the exact order in which electronic business transactions are executed.

Several standardization efforts consider business processes as well. The Universal Business Language (UBL) provides conceptual models, outlining the potential application scenarios of the defined business document definitions. Similarly, the current version of the Financial product Markup Language (FpML) contains a dedicated part for business process architecture. However, the most promising approach still remains the Electronic Business XML (ebXML) standard, released in 2001.

ebXML. The ebXML initiative was a joint effort between the two standardization organizations UN/CEFACT and OASIS with the goal to overcome the known problems of traditional EDI standards like EDIFACT. The decision was to define an integrated B2B framework having a strong business process focus.

The ebXML framework consists of five complementary components: registry, messaging, collaboration protocol profiles and agreements, business process specifications, and core components (CC).

In a nutshell, the five pillars of ebXML aim at providing an overall B2B infrastructure. Unlike other B2B standards, focusing on business document types only, the framework also deals with aspects like commonly agreed collaborative business process models, business partner profiles, business partner discovery, messaging infrastructure, etc. On the one hand, this can be considered as a distinctive advantage of the ebXML framework compared to other approaches. On the other hand, the broad focus of ebXML requires relatively high implementation efforts. Thus, it was one of the goals - and also a critical success factor - of the ebXML initiative to gain broad support by tool vendors. It was envisioned that tool vendors provide commercial-of-the-shelf software (COTS) for ebXML, allowing also small and medium-sized companies (SME) to buy affordable e-business solutions. This vision has not turned into reality. Until today, tool support for ebXML has remained rather low, resulting in a general low acceptance of ebXML. A detailed overview of ebXML's history, providing also a critical evaluation, is given in [10].

Comparable efforts include RosettaNet which is considered to follow a similar approach as ebXML in terms of its focus on providing an overall B2B infrastructure. However, RosettaNet's field of application is limited to the domains of computer electronics, electronic components, semiconductor manufacturing, and telecommunications. Another well known, but more lightweight approach is the already introduced OAGIS.

We conclude that integrated approaches, due to their complexity, have a rather low acceptance in the industry. However, if they are realized in a successful manner, their mightiness goes beyond approaches considering only document standardization.

f) Transitioned standardization approaches: The introduction of XML was a salvation for the ones and a plague for the others. New standard definitions could easily be created based on XML. However, existing standards, based on EDIFACT, either had to provide an XML equivalent for their standard or transition their entire standard definition to XML, in order to keep up with the pace of XML. While several of the EDIFACT-based standards are still in use today, others successfully transitioned their standard definitions to XML. We identify, that the current architectural style of service-oriented architectures (SOA) generally prefers XML-based standards, since they are easier to integrate in existing

solutions. However, companies which already have EDIFACT-based interfaces are reluctant to abandon their stable and tested EDI interfaces. Thus, several standardization organizations provide their standard definitions in both formats - EDIFACT and XML. This standard duality imposes several additional challenges on a standardization organization. On the one hand they still have to provide updates for the EDIFACT definitions in order to support legacy implementations. On the other hand new standard amendments have to be integrated in both formats - EDIFACT and XML - likewise. Additional maintenance effort is the result.

HL7. Health Level Seven (HL7) is a standard from the health care domain and developed and maintained by a not-for-profit organization named Health Level Seven.
The first version of the HL7 standard, referred to as version 2, followed the paradigm of a delimiter-based approach to encode health care information. With the availability of XML it was desirable to utilize XML as the new format for information exchange. Therefore, an XML representation for the delimiter-based HL7 standard was developed. Currently, two different formats (delimiter-based and XML-based) representing equivalent information are available. The core of the new version 3 is the Reference Information Model (RIM), serving as the basis from which all of HL7's information models for specific clinical situations are derived from.
Having the RIM of HL7 at hand, a formal methodology is provided, allowing to model elements and messages in a precise manner. Furthermore, using the RIM as a source for deriving other information models allows keeping consistency. One consequence of the paradigm switch from a delimiter-based format (version 2) to a common information model (version 3) is the incompatibility of version 2 and version 3 standards. If both approaches are to be supported by an application, additional implementation effort is necessary.
Another prominent example for a transitioned standard is provided by the Chemical Industry Data Exchange (CIDX) standard. However, CIDX has already abandoned the EDIFACT-based syntax and focuses entirely on XML. A similar scenario was pursued by XML/EDI, trying to provide EDIFACT-based standards in terms of XML. However, the project failed due to a lack of acceptance in the industry. Specialized standards, developed entirely either on XML or EDIFACT and focusing on certain industries and application domains, prevailed.
We conclude that transitioned standardization approaches are facing additional maintenance efforts, since they have to maintain two separate standard definitions in different representation formats. In the long-term, transitioned standardization approaches will abandon the EDIFACT-based syntax and focus entirely on XML, as the example of CIDX has already shown.

g) Implementation neutral standardization approaches: The following cluster focuses on implementation neutral standard definitions. In principle all business document standard definitions are bound to a specific syntax. In most cases this syntax is either some sort of delimiter-based approach or involves markup in order to separate different elements. Although the common syntax is in most cases XML, the underlying XML schema of each standard is different and thus no standard is like the other. Since no common base is provided for all of these standards, incompatibilities are inevitable. Implementation neutral standardization approaches aim at defining a common document definition on a generic and conceptual level without considering any specific syntax. In a model-driven approach conceptual business document models may be used in order to derive specific implementations of the standard, e.g., XML schema artifacts.

The advantages of such a model-driven approach are manifold. The conceptual model may be used to derive arbitrary implementation-specific artifacts. Since all artifacts are derived from a single conceptual model, the different artifacts share a common semantic basis. Based on this semantic basis, mapping mechanisms between different standard definitions may be implemented in a reusable and scalable manner.

CCTS. As a representative example we introduce a well known implementation neutral standardization approach named UN/CEFACT's Core Components Technical Specification (CCTS). Core components are reusable building blocks for assembling business documents. The core component standard family comprises three distinctive parts.
The Core Components Technical Specification (CCTS) defines the meta-model of the core component approach. Thereby, CCTS distinguishes between core components, which are context independent, and business information entities, which are context-specific. The idea is to define the basic building blocks of business documents first on a context independent level. If a certain industry wants to use a core component in order to assemble a business document with it, the core component is taken and tailored to the context-specific needs of the industry. In terms of CCTS, a core component thereby becomes a business information entity. Note that a business information entity may only be derived from a core component by restriction. Therefore, a business information entity only contains elements, which have previously been defined in the underlying core component. Since all business information entities are derived from the same core components, a common semantic basis between all business information entities is given at any time. The core components, which form the basis for business information entities are standardized by UN/CEFACT as well. In the *Core Components Library (CCL)* the reusable building blocks (core components) are defined in an unambiguous manner. Finally, the *UML Profile for Core Components (UPCC)* provides a UML representation of implementation neutral core components.

The major advantage of the core component approach is the fact that the concepts are defined on an implementation neutral level. In principle any appropriate representation may be used to build core component compliant business document models. The CCTS standard already contains an example implementation with the UML Profile for Core Components. However, other technologies such as the Web Ontology Language (OWL) have also been used in order to provide a core component compliant business document model representation.

The shortcomings of the CCTS do not lie in the specification itself, but rather in the library concept which is used for core components. Since any business document definition must be based on a core component, the existence of the appropriate core component in the UN/CEFACT maintained library is a prerequisite. If a necessary core component is not available in the library, a core component user may submit a new proposal for inclusion of a core component to UN/CEFACT. The harmonization process, however, takes some time and a business partner might not want to wait for so long. In order to overcome this limitation, core component registries may not only be established on a global level, but on a country or industry-specific level as well. Thus, if adherence to the global core component library is not desired, a country or industry-specific core component library may be set up. The price to be paid are incompatibilities between the different libraries.
Another implementation neutral standard definition approach is the Context Inspired Component Architecture (CICA). CICA aims at defining a collection of reusable components designed to fulfill cross-domain and cross-country business document requirements.
Implementation neutral standard definitions are very promising for achieving a common semantic data model on which different document definitions may be based on. The concept of business information entities, which are based on generic core components, helps to overcome the problem of overloaded top-down definitions. Only those elements are included in an industry-specific message, which are really needed and adherence to the generic base is provided. We conclude that currently the approach pursued by the UN/CEFACT's Core Components initiative is the most advanced implementation neutral technology available.

h) Converging approaches: In several industry domains such as the financial sector a multitude of different standards have been developed over the years. A lot of these standards cover the same problem domain and oftentimes concepts are defined in a redundant manner in different standard definitions. Thus, several industries started to converge existing approaches towards a single standard definition. One obstacle of converging standards is the fact that standards cannot be converged from one day to the other. Therefore, a convergence plan must be provided, allowing for the coexistence of the different standards which have to be converged, at least in the initial phase. Nevertheless, in the long-term all the different standard definitions under consideration in the given domain have to be converged to the newly introduced standard definition.

UNIFI. As an example for a converging standardization approach we introduce the Universal Financial Industry Message Scheme (UNIFI). The objective of the standardization committee was to enable communication interoperability among financial institutions, their market infrastructures, as well as end-user communities. However, within the financial sector a multitude of different, often overlapping standards have already been defined. UNIFI aims at a convergence of these different initiatives into one standard in the long term. However, in the short term the different standards need to coexist due to legacy and regulatory reasons. In order to allow for a coexistence of different standards together with UNIFI, a canonical message model is provided, to which the different standard definitions may be mapped to. The canonical model serves as an intermediate format for mapping between different standards. Thus, UNIFI aims at long term convergence, while facilitating short term coexistence.
The major advantage of the UNIFI initiative is the reduction of redundant business document standard definitions in the financial domain and the incorporation of distributed concepts into a single standard definition. Nevertheless, a converging standard such as UNIFI also has shortcomings as well. Several industry partners in the domain under consideration might not want to adapt to the new single converging standard definition due to several reasons. Apart from apparent criteria such as transition costs, it is often stated that a new converged standard definition is simply not needed because the old implementation is sufficient. While this argument may be true for short-term considerations, an adaptation of a converging standard provides strategic advantage in the long-term.
ebXML may also be seen as a converging standardization approach (if not as *the* converging standardization approach) because it aimed at a cross-industry and cross-border standard consolidation.
Our survey concludes that converging business document standardization initiatives are of particular importance, given the current abundance of business document standard definitions. Although convergence on a global level towards one single standard definition is still a huge challenge, there is great potential in industry-specific convergence. UNIFI provides a good example of a successful initiative and other industries are likely to follow.

III. A STANDARD COMPARISON

In Table II we present the key results of our business document standard survey. We analyzed our identified standard clusters according to different technical parameters with the overall goal to identify potential user groups and acceptance trends of a standard cluster. We do not further elaborate on early markup adopters, transitioned approaches, and converging approaches, since the standards in these clusters can be associated to one of the other clusters as well.

At first we examined the *business messaging compatibility* for each standard cluster using four parameters: representation, semantics, business process support, and specified transport protocol.
A business message must have an unambiguous *representation* format defined by a syntax (build on a grammar) and a vocabulary. All identified clusters obviously meet this requirement. Additionally, the *semantics* of different data elements and messages must be precisely defined, i.e., all parties must have the same interpretation of the exchanged information expressed by the business message representation. In principle

TABLE II: Standard cluster comparison - (1) - top-down approach (2) - bottom-up approach (3) - hybrid approach (5) - integrated approach (7) - implementation neutral approach

	(1)	(2)	(3)	(5)	(7)
Business Messaging Compatibility					
Representation	+	+	+	+	+
Semantics	+	+/-	+/-	+	+
Business Process	-	-	-	+	-
Transport	-	-	-	+	-
Technology features					
Used Syntax	EDI/XML	XML	XML	XML	-
Release iterations	$>$ 1 p.a.	$<$ 1 p.a.	$<$ 1 p.a.	$<$ 1 p.a.	$<$ 1 p.a.
Implement. complexity	+	+/-	+/-	+/-	+
Delta between releases	+/-	-	+	+/-	+
Backward compatibility	-	+	-	-	+
Extensibility	-	+	+	+	+
Concept. model avail.	+	+/-	+/-	+	+
Semantically unambiguous	+	-	-	+	+
COTS support	-	+	-	-	-
Standard maturity	+	+/-	+	+	+
Community size	+	-	+	-	+
Adoption	+	+/-	+	-	+
Potential User Groups					
Small enterprises	-	+	+/-	-	-
Medium-sized enterprises	-	+	+/-	-	+
Large enterprises	+	+	+	+	+
Acceptance					
Industry spec. accept.	+	+	+	+/-	+
National acceptance	+	+	+	+/-	+
Global acceptance	+	-	+	+/-	+

Legend: (+) Fully meets the criteria (+/-) Partly meets the criteria (-) Does not meet the criteria

all standard clusters fulfill this criteria. However, in bottom-up standard definitions and hybrid standard definitions user-specific extensions may be defined, where the agreement of different semantics is out of reach of a standardization organization. Thus, bottom-up standards and hybrid standards only partially meet this requirement. A *business process* defines the exact exchange order of business documents and ensures that appropriate responses and acknowledgments are sent. However, only integrated approaches such as ebXML, OAGIS, or RosettaNet consider a holistic approach towards business document standardization, where business processes are considered as well. Thus, we conclude that all other standard clusters only consider the data definition perspective, but do not take business processes into account. Furthermore, the participating parties in an electronic business transaction must agree on a *transport* protocol to interconnect their businesses. In the early days of EDI Value-Added Networks (VAN) were used to interchange business-to-business messages due to the absence of the Internet. With the emergence of the Internet the concept of VAN vanished and related technologies such as HTTP (Hyptertext Transfer Protocol) and SOAP (Simple Object Access Protocol) became popular for document exchange. Today, most standards leave it up to the implementer what protocol to choose for the exchange of a business document. Of all examined standard clusters only integrated approaches specifically recommend a certain technology, e.g., ebXML messaging is built on SOAP.

Finally, we conclude that only integrated approaches fulfill the entire needs of business messaging functionality. In fact ebXML is the only representative standard available, which covers business messaging in a holistic manner. However, the acceptance of integrated approaches such as ebXML is still very low. On the one hand this is due to the inherent complexity of these standards and unfortunately also due to low vendor support for compliant interfaces.

In the following we elaborate on the technological features of each standard cluster.

As already outlined in the introduction, traditional EDIFACT-based *implementation syntax* definitions have been superseded by XML-based standards. Today, XML represents state-of-the-art in business document standardization. However, top-down approaches and transitioned approaches still use the EDIFACT syntax. Naturally, implementation neutral standards do not use a specific syntax, but are defined on a conceptual level (e.g., using UML models).

In regard to *release iterations* we identified that top-down standard definitions are the only ones where more than one standard definition is released per year. In particular EDIFACT-based standards provide new standard releases on a regular basis (e.g., UN/EDIFACT releases new UN/EDIFACT directory versions twice a year). All other standard clusters have longer release cycles.

A critical factor when choosing a business document standard is the effort in regard to *implementation complexity*. Our research has shown that in particular top-down, hybrid, and integrated approaches require the most effort in establishing compliant business service interfaces. This is particularly contradicting for integrated standardization approaches, which claim to be suitable for small and medium-sized enterprises (SME) as well. Nevertheless, most SMEs cannot afford to implement complex and costly business service interfaces, since they do not have customizable ERP software available, but rely on customizable COTS. However, we found out that support for integrated approaches in COTS distributions is unfortunately very low.

Of particular importance in regard to the customization of business service interfaces is the Δ *between releases* of a certain standard. The higher the Δ, the higher is the effort for a potential customization. Our standard evaluation has shown that the Δ is the lowest in the cluster of bottom-up standards. Since bottom-up standards are defined as an intersection of requirements, where only the most important elements are considered in the core standard definition, only small changes occur between the different releases. In the cluster of top-down standards the Δ is low in regard to EDIFACT-based standards. These standard families are well tested and maintained and only little changes happen from release to release. However, other non-EDIFACT-based top-down standard definitions may include significant changes from release to release. Hybrid approaches and implementation neutral approaches reflect extensive changes in each release as our examination of UBL and CCTS has shown. In regard to integrated approaches no clear answer can be given, since some parts of, e.g., ebXML remain rather stable (e.g., business process specification scheme) whereas other parts have undergone significant changes (e.g., core components).

Another important factor in regard to the adaptation of business service interfaces is *backward compatibility*. If backward compatibility is provided, little to no adaptations have to be made to business service interfaces. We conclude that only standards from the bottom-up and implementation neutral cluster meet the requirement of high backward compatibility. Since bottom-up standards are specifically focused on not

changing the core standard definition too much, high backward compatibility is provided. We examined that implementation neutral standards such as the Core Components Technical Specification also try no to alter the basic concepts in order to foster reuse and backward compatibility between different versions.

In regard to *extensibility* only standards from the top-down cluster do not meet the required criteria. All other standard clusters provide necessary extension concepts for user or domain-specific standard amendments.

In particular for the communication between software architects and programmers a *conceptual model* representation of a business document standard definition is useful. We found out that currently only the bottom-up cluster misses a conceptual representation mechanism. Although the core standard definitions of bottom-up standards may be represented using a conceptual model, currently no appropriate approach for the representation of the different standard extensions exists.

As already outlined earlier, *semantic ambiguity* may occur with bottom-up and hybrid standardization approaches, since their extension mechanism allow any user-specific amendments. As a countermeasure a standardization organization may prevent user-specific extensions and provide well defined and standardized extension sets for certain domains (e.g., for the telecom industry).

An important part of our survey was to examine, whether a given standard cluster is supported by commercial-of-the-shelf software (COTS). In particular SMEs cannot afford costly ERP software, but rely on COTS. In fact currently only bottom-up standards are supported by COTS, since all other standard clusters are either too complex or their included standards are not pertinent for SMEs. Unfortunately support for ebXML in COTS is almost zero, although one of the main goals of ebXML was to support SMEs.

Concerning the *maturity of the different standard clusters* we found out that only bottom-up standards still require considerable consolidation and maintenance work. This is due to the fact, that bottom-up standards are the youngest standard family compared to all other clusters.

In regard to the *user community size* of a standard cluster, bottom-up and integrated approaches have the lowest community size. For bottom-up standards this is due to the adolescence of the standard. Integrated approaches unfortunately have never reached a critical mass of users.

In regard to the *adoption rate* of a standard cluster only integrated approaches suffer from a low acceptance rate. Taking ebXML as an example we conclude that although some parts of the standard are well accepted and in use in the industry (e.g., ebXML registry), little to no applications using the full range of ebXML exist.

Considering *potential user groups* per standard cluster we evaluated that in particular small and medium-sized enterprises are reluctant to adopt top-down or integrated approaches. Our evaluation has shown that acceptance of a standard by SMEs can only be guaranteed if appropriate tool support for handling documents is provided. Naturally, large enterprise are able to handle any of the presented standard clusters.

In regard to the *acceptance factor* in terms of industry, national, or global distinction all standards are equally accepted. Only standards from the bottom-up cluster are rather designed for a local use. Furthermore, standards from the integrated approach cluster lack acceptance on all levels.

IV. Conclusion and Outlook

In this paper we presented the major results of our business document standard survey, conducted during our research work on business document standardization. We have shown how the multitude of different business document standards may be classified using eight different categories. Each of the different categories contains a representative business document standard family. For each cluster we examined a representative standard in detail and gave examples of other contained standards. Due to space limitations not all standards contained in a category were introduced. However, it is rather the concept of the standard category itself which should help an implementer to classify a business document standard. Based on the presented findings per category an assessment of a standard should be possible even for a non-expert in the field of business document standardization. Finally, we showed our research results in regard to business message compatibility, technology features, potential user groups, and acceptance level of each standard.

We conclude that implementation neutral business document standards, and UN/CEFACT's Core Components Technical Specification in particular, are the best candidates for achieving a common business document exchange format. However, the biggest obstacle towards achieving seamless document interoperability does not lie on a syntactical or semantical level. It are political decisions which have to be made in order to achieve a common document standard. This is a challenge yet to be solved.

References

[1] T. Hughes, *Networks of Power: Electrification in Western Society, 1880-1930*. The Johns Hopkins University Press, 1983.

[2] P. Feng, "Studying standardization: a review of the literature," in *Proceedings of 3rd IEEE Conference on Standardization and Innovation in Information Technology*, 2003, pp. 99–112.

[3] D. J. Kim, M. Agrawal, B. Jayaraman, and H. R. Rao, "A Comparison of B2B e-Service Solutions," *Communications of the ACM*, vol. 46, no. 12, pp. 317–324, 2003.

[4] A. Dogac and I. Cingil, "A survey and comparison of business-to-business e-commerce frameworks," *SIGecom Exch.*, vol. 2, no. 2, pp. 16–27, 2001.

[5] J.-M. Nurmilaakso, P. Kotinurmi, and H. Laesvuori, "XML-based e-Business Frameworks and Standardization," *Computer Standards and Interfaces*, vol. 28, no. 12, pp. 585–599, 2006.

[6] H. Li, "XML and Industrial Standards for Electronic Commerce," *Knowledge and Information Systems*, vol. 2, no. 4, pp. 487–497, 2000.

[7] J. Berge, *The EDIFACT Standards*, 2nd ed. Cambridge, MA, USA: Blackwell Publishers, 1994.

[8] European Committee for Banking Standards, "Electronic Payment Initiator (ePI), V 1.1," 2003.

[9] K. Holman, "UBL Catalogue analysis," http://markmail.org/message/o3ra6fffiw6mu7jw.

[10] K.-D. Naujok and C. Huemer, "Case study: Designing ebXML: The work of UN/CEFACT," in *Ontologies-Based Business Integration*. Springer Berlin Heidelberg, 2008, pp. 79–93.