Business documents in a service oriented context

Philipp Liegl*
Institute of Software Technology and Interactive Systems
Business Informatics Group
Favoritenstrasse 9-11/188
Vienna, Austria
liegl@big.tuwien.ac.at

ABSTRACT

In a SOA based inter-organizational business process different business documents are exchanged in an agreed order between business partners. Although several standards for the definition of a process choreography exist nowadays, the precise and unambiguous definition of the exchanged business documents is still an open research task. However, if the business partners do not commonly agree on a business document format, interoperability is unlikely to be achieved. A solution can be provided by defining a UML profile, integrating concepts from UN/CEFACT's Core Components standard and industry specific requirements and best practices. The UML profile allows to define business document models on a conceptual level. These conceptual level models can be used to exchange business document information between software developers and more important to automatically derive logical level models e.g. XML schema. These logical level artifacts can be automatically deployed in IT systems of a service oriented architecture.

1. MOTIVATION

With the introduction of Electronic Data Interchange (EDI) initiatives the need for a standardization of the exchanged information became apparent. One of the best known EDI standardization initiatives is the UN/EDIFACT [20] [1] standard. Typical EDI agreements involved a well defined set of business partners collaborating with a predefined set of business documents over a longer time-frame. Thus adaptations of the exchanged business documents were rather rare and if they occurred the changes were mostly minor ones. An early and interesting survey on the determinants of EDI diffusion has been published in [16]. Since the first EDI initiatives in the early sixties which were reserved for large companies and industries, the IT landscape has changed significantly. Today service oriented technologies enable small and medium sized enterprises to participate in electronic collaborations as well. The before rather hard-wired processes between enterprises have been replaced by loosely-coupled and service oriented ones. Service orientation allows for faster changes in B2B processes and thus also in the exchanged business documents. The hard wired EDI approach is too inflexible for these new ad-hoc business processes. Several initiatives, mostly industry specific ones, have been found in recent years - some with notable, others with almost no success. Since almost all of these initiatives were focused on a certain domain or industry, a broad cross-industry acceptance could not be reached. As part of the UN/CEFACT (United Nations Center for Trade Facilitation and Electronic Business) initiative for document standardization the core component [21] standard has been developed, consisting of reusable building blocks for the definition of business documents. However, core components are a theoretical concept and no integration into a modeling tool exists until today.

In order to allow for a better understanding of the industry-specific business document modeling requirements a precise survey of the various business document standardization efforts is necessary. Thereby the specific advantages and disadvantages in regard to a use in a service oriented context have to be examined. Using the results of the survey and the concepts of the core component standardization, a common solution allowing for the integration into a UML modeling tool has to be found. Eventually the business document modeler should be able to construct business documents on a conceptual level with a UML modeling tool of choice. Finally another important question remains - how to uniquely derive deployment artifacts from conceptual data models which may then be employed in a service oriented context e.g. XML schema. Furthermore it must be examined how conceptual business document models can be integrated into business process models.

2. PROBLEM

If two businesses engage in an automated business process two major agreements are necessary. First, the order of the exchange of business documents has to be agreed upon - the so called business choreography. Using the global business document exchange choreography each business partner can configure his own IT system. Second, the business documents and their precise semantics have to be settled. In figure 1 an example business process `order from quote` between a *buyer* and a *seller* is shown. First the *buyer* requests a `quote` from the *seller* and submits a `purchase order` to the *seller*. The *seller* either replies with an `order acceptance` or with an `order rejection`. In a service ori-

*2nd year PhD student at Vienna University of Technology

ented context the interfaces (WSDL) [26] and the messaging (SOAP) [25] are well defined. However little or nothing is said about the actual workload being exchanged between the two companies.

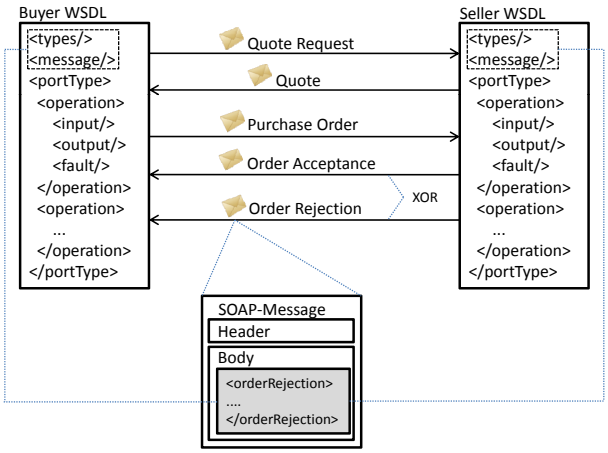

Figure 1: Business Documents in a SOA context

On the lower side of figure 1 an example SOAP message is shown. Whereas header and body of a SOAP message are well defined, the actual document format carried in the body remains undefined. However, in order to allow the two WSDL interfaces to be compliant the message types have to be defined as well. Standardization efforts in recent years have brought up a multitude of different standards and approaches for the definition of business data. Unfortunately the different standards have a set of shortcomings.

First, the multiple initiatives being based on XML are mostly incompatible to each other. Since the different standards are not based on a common semantic base difficult mappings between different standards are necessary. However, data mappers are expensive to implement and inflexible to changes of the structure of the exchanged business documents.

Second, a lot of the different standards aim at the inclusion of every possible element in a business document. This results in a strong overhead of a standard. Furthermore, such an overhead makes a standard difficult to implement - in particular for small and medium sized enterprises. Whereas for instance a regular invoice just requires a small set of elements and attributes an invoice standardized by UN/EDIFACT contains a multitude of possible elements not needed in a regular business transaction.

Third, most of the standards are transfer syntax specific. Standards such as UN/EDIFACT are tightly bound to the implementation syntax. Instead of defining a business document on a conceptual level most of the standards are based on a logical level e.g. XML schema. Changes in the transfer syntax therefore require a complicated reengineering of the entire standard. Thus future adaptations of the standard are difficult and time consuming to implement.

Finally the exchange of a logical level business document definition between two business partners is complicated and error prone during the development phase of the B2B process. Instead of exchanging e.g. XML schema definitions, a conceptual business document model is easier to understand for the developers on each business partner's side.

To sum up, a new methodology for the definition of business documents is needed in particular to meet the needs of a service oriented context. A SOA context requires regular changes in the exchanged document structure and the automatic derivation of deployment artifacts in order to quickly meet the changed business document requirements.

3. HYPOTHESIS

In order to allow two companies to agree upon a common document semantic a methodology for defining business documents is needed. UN/CEFACT's Core Components [21] is a methodology to uniquely define business documents. The current core component version is 2.01 with the development of 3.0 [24] currently going on. The development of core components is backed up by the United Nations and can therefore be seen as the most promising global approach for business document standardization. However, several other industry initiatives and standardization bodies have developed business document standards as well. Nevertheless, the different standards are largely incompatible to each other, thus hampering a broad proliferation of business document standardization.

Hypothesis 1. It is necessary to find a common base between different industry specific requirements and the global core components initiative. Since core components are a generic concept the different industry requirements can be reflected. However core components are a theoretical concept and therefore no real integration into a modeling tool is possible. Recent years have shown, that the Unified Modeling Language (UML) is becoming a very promising technology for the modeling of structural and dynamic behavior. If the core components technology together with the identified industry requirements could be integrated into a UML modeling tool, the proliferation of the technology would be accelerated, thus leading to a broader acceptance. A UML tool of choice could be used to model conceptual business document models. Therefore a solution for the integration of the core component concept into a UML modeling tool must be found [7].

Hypothesis 2. If core components are modeled using a UML tool of choice, the result would be conceptual business document models based on UML. These models can be used to exchange structural business document information between business analysts, software developers and other stakeholders. Although a conceptual model is appropriate for the exchange of structural information, it cannot be used in a real-world IT system. In order to allow for the integration of a business document in a SOA context it must first be represented on a logical level e.g. XML schema. Therefore a solution for the derivation of logical level business documents (e.g. XML schema) from their conceptual level core component representation must be found. However, not only the forward engineering (UML to XML) must be addressed, but also the reverse engineering (XML to UML).

Hypothesis 3. As already outlined in section 2 business process modeling and business document modeling are strongly

interlinked. First, a business process model indentifying the exact exchange order of business document is constructed. Second, a business document model defining the exact outline of the exchanged business documents is created. In recent years UN/CEFACT's Modeling Methodology (UMM) [22] has established itself as a very promising approach for the definition of inter-organizational business process model. In order to allow for a seamless integration into business process models, an integration approach for the core components technology into the UMM has to be found. A UMM model defining the process perspective together with a core components model defining the document perspective provides a holistic B2B methodology for the definition of SOA requirements.

4. METHODOLOGY

In order to provide a solution to the current problems defined above a specific methodology will be employed which consists of five distinctive steps. Figure 2 gives an overview about the used approach.

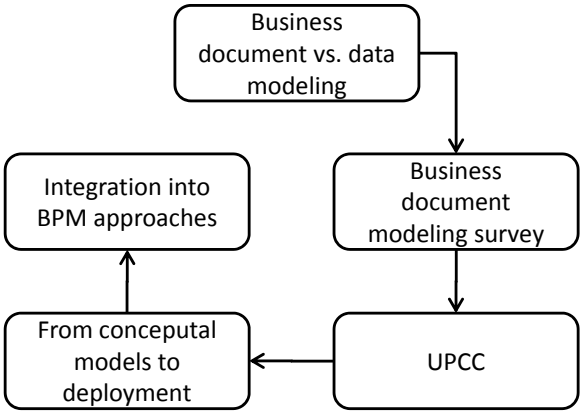

Figure 2: Dissertation approach

Business document modeling vs. regular data modeling. In the first step the specific differences between regular data modeling techniques such as the ER-model [3] from the relational database management systems (RDBMS) design and business document model requirements will be examined. In particular the specific requirements necessary for business document modeling are going to be revealed. Since the overall goal is to develop a holistic business document methodology a profound knowledge of the different modeling techniques is necessary. Furthermore it provides a good introduction into the domain of document modeling per se.

Business document modeling survey. The second step will comprise a survey of current business document modeling approaches, their advantages, disadvantages and application scenarios. Using the survey a precise overview about current industry approaches and best practices will be given. Furthermore a classification of the business document approaches in regard to their applicability in a service oriented context will be defined. The specific requirements of the different industry solutions will be reflected to the maximum extend possible in the overall core component methodology. Since core components provide a generic concept, the integration of domain and industry specific requirements is easily feasible.

UPCC - A UML Profile for Core Components. As outlined before, core components are reusable building blocks for the definition of business documents. However, core components are a theoretical concept and no integration into a UML modeling tool is possible. In order to merge the industry specific requirements identified in the step before and the core component technology a common solution based on UML must be found. However, UML is a very generic concept allowing the construction of a multitude of different structural and behavioral models. In order to restrict the UML meta model to the specific needs of business document modeling, a UML model will be defined. A UML profile tailors the UML meta model using stereotypes, tagged values and OCL constraints. The profile can be imported into any UML modeling tool of choice and provides the business modeler with the necessary artifacts to define business documents. If both modelers use the same profile, the conceptual business document models are compatible to each other, overcoming the interoperability limitations of the different standardization initiatives. We have already started to address this issue in [9] and also co-authored the first experimental UML profile for the core components standard 2.01 [23]. With the advent of the new core components standard 3.0 [24] the old profile is obsolete and a new profile has to be constructed. The definition of a new profile, reflecting industry specfic requirements will be an integral part of this dissertation.

From conceptual models to deployment artifacts. A UML profile for core components allows the definition of conceptual business document models. Since the core components standard is employed, the different business document models share the same semantic base, thus guaranteeing interoperability. However, a conceptual model cannot be deployed in a real world IT system. Thus a derivation mechanism will be developed, allowing the generation of deployment artifacts such as XML schema from conceptual business document models. The logical level XML artifacts can be deployed in a IT system of choice. However, the generation of deployment artifacts must not necessarily be restricted to XML schema - the generation of UBL [14], Relax NG [13] or any other appropriate artifacts is also possible since the conceptual core component model does not mandate any specific implementation technology. As a proof-of-concept a generation of deployment XML schema artifacts will be developed on top of the UML modeling tool Enterprise Architect [18].

Integration into BPM approaches. Since business documents in a SOA context are exchanged in a collaborative process, the exemplary integration of a conceptual business document model into inter-organizational business process technologies will be subject to the last step of the thesis. As a process methodology of choice UN/CEFACT's Modeling Methodology (UMM), which we have co-authored, will be chosen. UMM itself is also defined as a UML profile on top of UML 2.0. Since the core component profile will also be defined on top of UML 2.0, a seamless integration is feasible. Of particular interest will be the derivation of artifacts from both methodologies - UMM on the process side and core components on the business document side. Given this

holistic approach, the easy configuration of a SOA system will be possible.

5. GOAL

In figure 3 the overall goal of the dissertation is shown. First two business partners have to agree upon a common process choreography. The process choreography defines in which order the different business documents are exchanged between the two business partners. In the context of the thesis the global choreography will be defined using UN/CEFACT's Modeling Methodology. After having found a common agreement on the business document exchange order, the business documents per se have to be defined. Using the UML Profile for core components, which will be developed in this dissertation, the two business partners uniquely define the document semantic. In the next step the common business document model is used in order to derive XML schema deployment artifacts. The XML schema is then fed into the IT systems of the two business partners. Since both IT systems are configured on the same basis, a common agreement on the interface definitions of the respective WSDL is found.

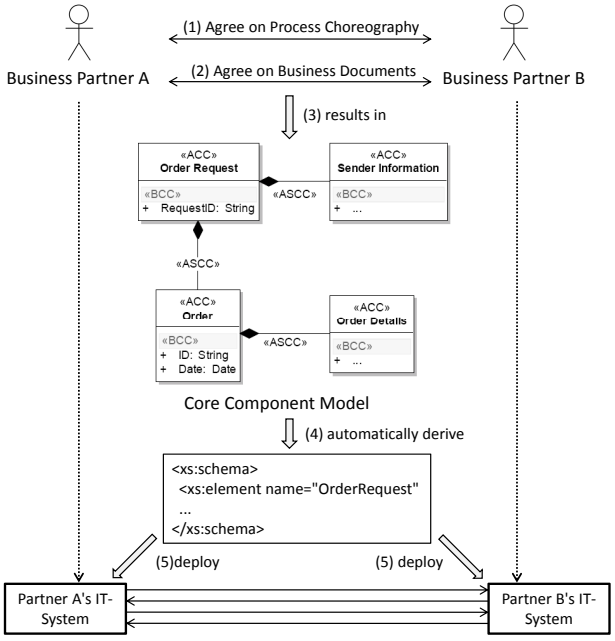

Figure 3: From a conceptual model to deployment

By setting the business document semantic on a conceptual level using the core components semantic, a quick agreement between the two business partners is possible. Since the derivation of the XML artifacts is done automatically, a fast deployment is possible after the requirements of the business documents have changed.

6. RELATED WORK

For the interested reader a general introduction into the field of business document engineering is given in [6]. In the field of business document standardization several initiatives have been found in recent years. Mostly the efforts are industry specific and bound to the XML standard. The most important standards include CIDX [4], SWIFT [19], HL7 [8], and PapiNet [15] just to name a few. A lot of research has been conducted in the field of XML artifact generation out of UML artifacts. Generally a distinction is made between forward engineering (XML artifact generation from UML models) and reverse engineering (UML model generation from XML artifacts). Whereas the first approach is often straight-forward the second one contains some obstacles. E.g. a derivation by restriction cannot be modeled with the Unified Modeling Language. A profound overview on the reverse engineering of XML schema to conceptual models is given in [27]. Thereby the authors examine several reverse engineering techniques and assess their applicability. In the field of forward engineering several research efforts can be named e.g. [17] and [5]. Several research efforts have also been conducted in the field of business process and business document modeling alignment e.g. [10] [11]. As outlined before core components define the semantics on a conceptual level and the semantics are transformed to the logical level XML schema by derivation. Several other research efforts have already addressed the definition of semantics in XML schema e.g. [12] and [2]. Most of these approaches however are missing a holistic approach for the specific needs in regard to a use within a SOA context.

7. CONCLUSION

Common modeling approaches for B2B scenarios often only reflect a particular part of a business collaboration e.g. only the process perspective or only the data perspective. In order to allow for a holistic view on a B2B collaboration a solution embracing all necessary views and technologies is necessary. We have already actively developed the UN/CEFACT's Modeling Methodology, reflecting the process perspective of a business collaboration. This thesis will close the gap to a holistic B2B methodology by incorporating the business document modeling perspective as well. Thereby different business document modeling standards will be evaluated and their advantages and disadvantages will be assessed.

Based on these results and using UN/CEFACT's core components technology, a UML profile will be developed. The UML profile will allow to easily assemble business documents on a common semantic basis. The conceptual business document models will then serve as the basis for the derivation of logical document models e.g. XML schema artifacts. The XML schema artifacts can be fed into IT systems of participating business partners and guarantee, that all partners have the same understanding of the exchanged information. Finally the integration into the process modeling methodology UMM will be examined. After the finalization of the thesis a business modeler will be given a holistic methodology for the modeling of B2B scenarios.

8. REFERENCES

[1] J. Berge. *The EDIFACT Standards*. Blackwell Publishers, Cambridge, MA, USA, 2 edition, 1994.

[2] M. Bernauer, G. Kappel, and G. Kramler. Approaches to implementing active semantics with XML schema. In *Proceedings of the 14th International Workshop on Database and Expert Systems Applications*, pages 559–565. 2003.

[3] P. P.-S. Chen. The Entity Relationship Model: Towards a unified view of data. 1:9–36, 1976.

[4] CIDX. *Chemical Industry Data Exchange Standard*,

2007.

[5] C. Combi and B. Oliboni. Conceptual modeling of XML data. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, pages 467–473. ACM, New York, NY, USA, 2006.

[6] R. Glushko and T. McGrath. *Document Engineering*. Massachusetts Institute of Technology, United States, 2nd edition, 2005.

[7] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design Science in Information Systems Research. 28:75–105, 2004.

[8] HL7. *Health Level Seven*, 2007.

[9] C. Huemer and P. Liegl. A UML Profile for Core Components and their Transformation to XSD. In *ICDE Workshops*, pages 298–306, 2007.

[10] C. Janiesch, A. Dreiling, U. Greiner, and S. Lippe. Configuring processes and business documents - an integrated approach to enterprise systems collaboration. In *ICEBE*, pages 516–521, 2006.

[11] C. Janiesch, A. Dreiling, U. Greiner, and S. Lippe. Integrated configuration of enterprise systems for interoperability – towards process model and business document specification alignment. In *EDOC*, pages 445–448, 2006.

[12] J. B. Li and J. Miller. Testing the semantics of W3C XML schema. In *Proceedings of the 29th Annual International Computer Software and Applications Conference*, pages 443–448. 2005.

[13] OASIS. *RELAX NG Specification*. OASIS, December 2001. Committee Specification.

[14] OASIS. *Universal Business Language v2.0*. OASIS, 2006.

[15] papiNet. *papiNet*, 2007.

[16] K. Ramamurthy and G. Premkumar. Determinants and Outcomes of Electronic Data Interchange Diffusion. 42:332–351, 1995.

[17] D. Skogan. UML a Schema Language for XML based Data Interchange. In *Proceedings of the Second International Conference on the Unified Modeling Language*, volume 2, pages 211–220. Proceedings, United States, 1999.

[18] Sparx Systems. *Enterprise Architect*, 2008.

[19] SWIFT. *Society for Worldwide Interbank Financial Telecommunication*, 2007.

[20] UN/CEFACT. *UN/EDIFACT D.07B*, 2007.

[21] UN/CEFACT TMG. *Core Components Technical Specification - Part 8 of the ebXML Framework*, 2003.

[22] UN/CEFACT TMG. *UN/CEFACT's Modeling Methodology (UMM), UMM Meta Model - Foundation Module*, 2006.

[23] UN/CEFACT TMG. *UPCC - UML Profile for Core Components based on CCTS 2.01*. United Nations Center For Trade Facilitation and Electronic Business, October 2006. Candidate for version 1.0.

[24] UN/CEFACT TMG. *Core Components Technical Specification 3.0 - draft*, 2008.

[25] W3C. *Simple Object Access Protocol*, 2007.

[26] W3C. *Web Services Description Language*, 2007.

[27] A. Yu and R. Steele. An overview of research on reverse engineering XML schemas into UML diagrams. In *Proceedings of the Third International Conference on Information Technology and Applications ICITA 2005*, volume 2, pages 772–777. ICITA Proceedings, Australia, 2005.