

# iSURF eDoCreator: e-Business Document Design and Customization Environment

*Fulya Tuncer<sup>1</sup>, Asuman Doğaç<sup>2</sup>, Yıldıray Kabak<sup>2</sup>, Şenan Postacı<sup>1</sup>, Suat Gönül<sup>1</sup>, Erdem Alpay<sup>1</sup>*

<sup>1</sup> *Dept. of Computer Engineering, Middle East Technical University, İnönü Bulvari, Ankara, 06531, Turkey*

*Tel: +90 312 2102076, Fax: + 90 312 2105572, {fulya,senan,suat,erdem}@srdc.metu.edu.tr*

<sup>2</sup> *Software Research, Development and Consultancy Ltd., METU-KOSGEB Tekmer, Ankara, 06531, Turkey*

*Tel: +90 312 2105598, Fax: + 90 312 2101259, {asuman,yildiray}@srdc.com.tr*

**Abstract.** The leading effort for creating a standard semantic basis for business documents to solve electronic business document interoperability problem came from the UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) Core Components Technical Specification (CCTS) through a document modeling methodology. Currently, the main challenge in using UN/CEFACT CCTS based approaches is that the document artifacts are stored in spreadsheets and this makes it very difficult to discover the previously defined components and to check their consistency. Furthermore, businesses need to customize standard documents according to their specific needs. The first XML implementation of UN/CEFACT CCTS, namely, Universal Business Language (UBL) provides detailed text-based descriptions of customization mechanisms. However without automated tool support, it is difficult to apply the customization and to maintain the consistency of the customizations. iSURF eDoCreator tool addresses these problems by providing an online common UN/CEFACT CCTS based document component modeling repository which integrates the machine processable version of paper-based UBL customization guidelines. In this way, iSURF eDoCreator environment aims to maximize re-use of available document building blocks and minimize the tedious document design and customization efforts. The environment also proved to be very useful in performing the gap analysis between different customizations of UBL.

## 1. Introduction

Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged [1]. With the recent technological developments such as Web services, a degree of interoperability at the transport and the communication layer has been achieved. Yet the interoperability of exchanged documents is still a difficult problem. Several standard interfaces have been developed to provide the interoperability of Electronic Business Documents [2]. However, having more than one standard hampers interoperability because the business organizations using different standards still suffer from the interoperability problem. Furthermore, even the businesses that use the same standard tailor it for their specific needs because they have different rules and requirements for the information they exchange according to the industry domain they operate and their geopolitical and regulatory contexts. As a result, there is an interoperability challenge even for the communities that use the different customizations of the same document standard.

The leading effort for creating a standard semantic basis for business documents to solve electronic business document interoperability problem came from the UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) Core Components Technical Specification (CCTS) [3] which provides a document modeling methodology. The ultimate

aim is to derive all electronic documents from common building blocks with well-defined rules. This implies creating core document building blocks with common semantics and then specializing them to the contexts. Additionally, both the core components and the specialized document building blocks must be discoverable from a common repository to be able to re-use them. In other words, the well-defined semantics for the core document components together with a discovery facility will improve their reusability and enhance interoperability.

The UN/CEFACT CCTS is syntax independent. One of its first syntax dependent implementations in XML is UBL [4]. Currently, the approved version of UBL is 2.0 and there are thirty one XML schemas for common business documents like “order” and “invoice”. In addition to the document definitions, UBL 2.0 provides a library of XML schemas (XSDs) for reusable common data components like “Address”, “Item”, and “Payment” from which the context specific documents are customized.

Currently the main challenge in using these UN/CEFACT CCTS based approaches like UBL is the fact that the document artifacts are stored in spreadsheets and this makes it very difficult, if not impossible, to discover the previously defined components and to check their consistency. Furthermore, as already mentioned, businesses need to customize standard documents according to their specific needs. The Universal Business Language (UBL) provides detailed text-based descriptions of customization mechanisms. However without an automated tool support, it is difficult to apply the customization and maintain the consistency of the customizations.

iSURF eDoCreator addresses these problems by providing a document modeling environment with a common online UN/CEFACT CCTS based component repository which integrates the machine processable version of the paper-based UBL customization guidelines to be applied on the document components. In this way, it aims to maximize the re-use of available document building blocks and minimize the duplicative efforts of document designers in customizing the document schemas.

The repository, available at [5], currently contains all of the BIEs in the common library of UBL 2.0; all the BIEs of the UN/CEFACT Cross Industry Electronic Invoice (CII) [6]; all the BIEs of NES [11] and UBLTR [12]. The repository is gradually evolving to include document components from other standards such as the Core Components (CCs) and the Business Information Entities (BIEs) of UN/CEFACT CCL 08B [13]. As the new document models are created or document building blocks are customized and committed to the repository, its impact on the data interoperability will increase by making it possible to share all these with a wider audience.

## **2. Objectives**

iSURF eDoCreator is a part of a larger effort, namely, iSURF Interoperability Service Utility. The main objective of iSURF Interoperability Service Utility is to provide electronic business document interoperability. For this purpose two parallel challenges are being addressed: One is to semantically annotate the document schemas as described in [7] and the other is to provide a graphical environment for the customization and re-use of UN/CEFACT CCTS based document schemas, as described in this paper.

Creating, extending, customizing document schemas conforming to UN/CEFACT CCTS methodology are tedious, labour intensive and time-consuming processes requiring (1) analysis of available component interfaces (2) design of spreadsheet model of the document (3) creation of XSD files and finally (4) creation of genericcode files for each of the coded attributes.

Although UN/CEFACT CCTS and UBL provide guidelines for document modeling and document customization respectively, there is no machine processable process implemented to help the designers. iSURF eDoCreator environment converts the UN/CEFACT CCTS

modeling methodology into a machine processable process to execute on the document building blocks in the online repository and implements the UBL Customization guidelines to provide common and publicly available document modeling services. The tool also generates the spreadsheet model of the document schema and the XSD files along with the generic code files.

### **3. Enabling Technologies and Standards**

#### *3.1 UN/CEFACT Core Component Technical Specification (CCTS)*

UN/CEFACT Core Component Technical Specification (CCTS) provides a methodology for semantic data modeling on a syntax independent level. It achieves a common understanding of data structures and message types in order to provide interoperability at the data level among e-Business applications. Furthermore, by setting Naming and Design Rules (NDR), which define how to name, structure and assemble the components, it establishes a common understanding among business partners eliminating ambiguities arising from misinterpretations.

The key concepts of CCTS are based on Core Components (CC) and the Business Information Entities (BIE): Core Components are building blocks with generic semantics and purpose; they are context-neutral templates so that they can later be adapted to different contexts and reused. Once the business contexts are declared on the CCs, they become BIEs and reflect the requirements of a given business context based on that specific CC.

Both CCs and BIEs have three different types: Aggregated entities (ACCs and ABIEs), which consist of one or more Basic entities (BCCs and BBIEs) as atomic attributes, and Associated entities (ASCCs and ASBIEs) as complex attributes associating other Aggregate Entities. Furthermore, a Basic Entity is based on a Data Type which can be either a Qualified Data Type or an Unqualified Data Type. All of these components are termed as “document building blocks” in this paper.

#### *3.2 OASIS Universal Business Language (UBL)*

UBL is an OASIS standard which adopts the UN/CEFACT CCTS approach and defines a library of XML schemas (XSDs) for re-usable components and common business documents. Unlike to UN/CEFACT CCTS, it includes only Business Information Entities by excluding Core Components. Furthermore, UBL 2.0 reuses Core Component Type and Data Type definitions from UN/CEFACT CCTS.

There are two types of customizations specified in UBL 2.0 [8]: Conformant customization and Compatible customization. The conformant customization allows the UBL instances in the customized implementation to also conform to the original standard UBL 2.0 schemas.

UBL Conformant Customization Guidelines allow defining restrictions on the original UBL schemas such as subsetting the document schema, defining constraints on document content or applying custom codes. In conformant design guidelines only optional elements may be excluded from the document schema. Furthermore, the cardinalities of the elements can be updated with the following restrictions: minimum cardinalities can be increased up to their maximum; maximum cardinalities can be decreased down to their minimum.

Compatible customization is used when conformant customization is not sufficient for a specific implementation. For example, users may need to extend an ABIE, create a new ABIE or create a new document from scratch. In compatible customization, the users modify an existing UBL 2.0 schema or create a new one by re-using the largest suitable aggregation from the UBL library. The key idea in UBL Compatible customization is to create new entities based on UBL artifacts in order to preserve the common understanding.

#### 4. iSURF eDoCreator System Architecture

The iSURF eDoCreator system architecture is composed of three main components as shown in Figure 1.

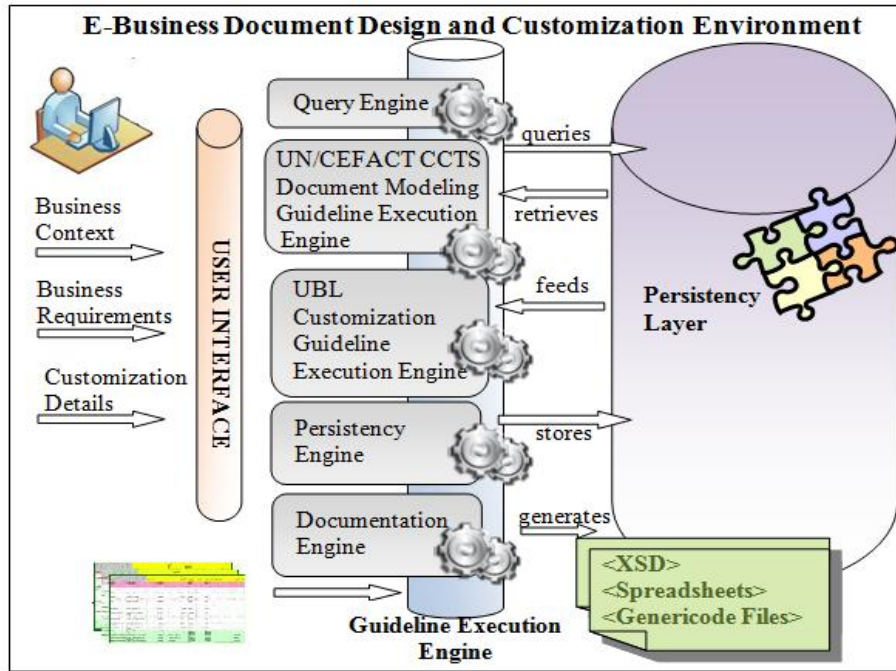


Figure 1 iSURF eDoCreator System Architecture Components

- The *Graphical User Interface* handles user interactions and helps to visualize UN/CEFACT CCTS based conceptual model in a hierarchical tree view and converts the conceptual model to logical model at the presentation layer. The tree view specifies document building blocks with their type and the Dictionary Entry Names. The tree icons identify type of the document building blocks, while the labels of the tree nodes specify their Dictionary Entry Names. The hierarchical organization of document components is shown through expandable tree interface. By clicking on the nodes of the tree, users can expand a tree node and navigate over the tree structure to see the composition details of document building blocks such as the Aggregate Business Information Entities, Basic Business Information Entities or Association Business Information Entities that are included in a Business Message Type, or the Data Types of the Basic Business Information Entities. The user interface of the modeling environment is implemented with Flex [9] since it provides Rich Internet Application API enabling generation of user-friendly and highly-interactive Web applications.
- The *Online Repository* component provides persistency and graphical access to the users over the Internet through its Web-based environment. The models generated by the users are persisted by using a relational database and retrieved through graphical queries. The wiki-based collaborative nature of the repository enables users to participate in collaborations that it is accessible by other participants at the time of submission.
- The *Guideline Processing Engine* is the main controller of the modeling environment. It guides the users in following the UN/CEFACT modeling and UBL customization guidelines during the generation of standard-based but at the same time

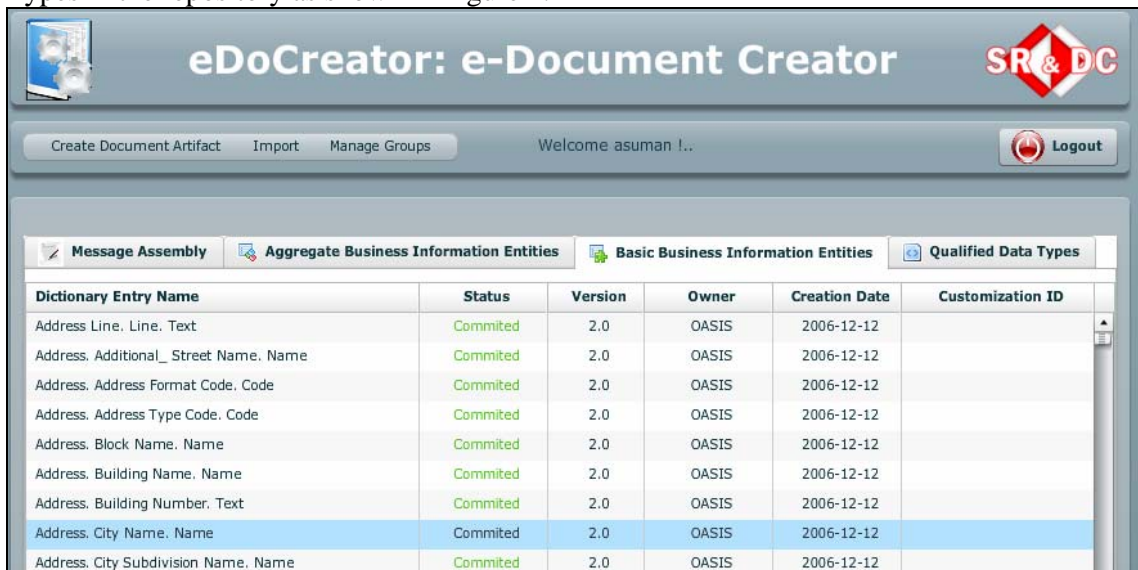
customized business documents. The engine checks the consistency between user actions and the guidelines at all steps. Furthermore, it handles interactions between the graphical user interface and the repository and implements a number of key features of the environment. The Persistency Engine mediates the interaction between user interface and the persistency layer. It serializes document building block models into an intermediate XML format in order to enable visualization of document building blocks in graphical user interface and deserializes customized document building block models into the intermediate XML format and then to the object model in order to persist it into the repository. Moreover, it processes both Microsoft Excel or Open Office Calc spreadsheets formats and loads document building blocks into the repository. Query engine forms SQL queries according the keywords entered by users and executes the query on the repository. Documentation Engine converts logical models into physical model and generates XML Schema Definition (XSD) files (XSD), spreadsheet models and generic code files in UBL directory format.

## 5. e-Business Document Design and Customization Environment

The iSURF eDoCreator tool provides online and 7/24 accessible environment for the graphical modeling of business documents by integrating UN/CEFACT CCTS modeling and UBL Customization guidelines. It is designed to aid the user in following these guidelines during the generation of standard-based and customized business documents.

### 5.1. Providing the graphical view of the document building blocks in the repository and querying the repository

Currently iSURF eDoCreator repository [5] contains all of the BIEs in the common library of UBL 2.0; all the BIEs of the UN/CEFACT Cross Industry Electronic Invoice (CII) [6]; all the BIEs of NES [11] and UBLTR [12]. The first functionality of the tool is to provide a graphical view of the coarse-grained document building blocks such as Message Assembly, Aggregate Business Information Entities, Basic Business Information and Qualified Data Types in the repository as shown in Figure 2.



Dictionary Entry Name	Status	Version	Owner	Creation Date	Customization ID
Address Line. Line. Text	Committed	2.0	OASIS	2006-12-12	
Address. Additional_Street Name. Name	Committed	2.0	OASIS	2006-12-12	
Address. Address Format Code. Code	Committed	2.0	OASIS	2006-12-12	
Address. Address Type Code. Code	Committed	2.0	OASIS	2006-12-12	
Address. Block Name. Name	Committed	2.0	OASIS	2006-12-12	
Address. Building Name. Name	Committed	2.0	OASIS	2006-12-12	
Address. Building Number. Text	Committed	2.0	OASIS	2006-12-12	
Address. City Name. Name	Committed	2.0	OASIS	2006-12-12	
Address. City Subdivision Name. Name	Committed	2.0	OASIS	2006-12-12	

Figure 2 iSURF eDoCreator displaying the coarse-grained document building blocks in the repository

The tool provides the details of properties of these coarse-grained document building blocks by clicking on the “properties” tab of the popup menu that appears when right clicking on the selected item. This expandable view of the document building blocks allows the user to see the whole data content of a component at a glance as shown in Figure 3. This is in contrast to spreadsheet format in which rows corresponds to document building blocks, and columns give the details of the component on that row. Furthermore, in the spreadsheet model, the properties of an encapsulated data type are presented in separate spread sheets, so in order to have a complete view of a document building block, it is necessary to explore more than one spreadsheets.

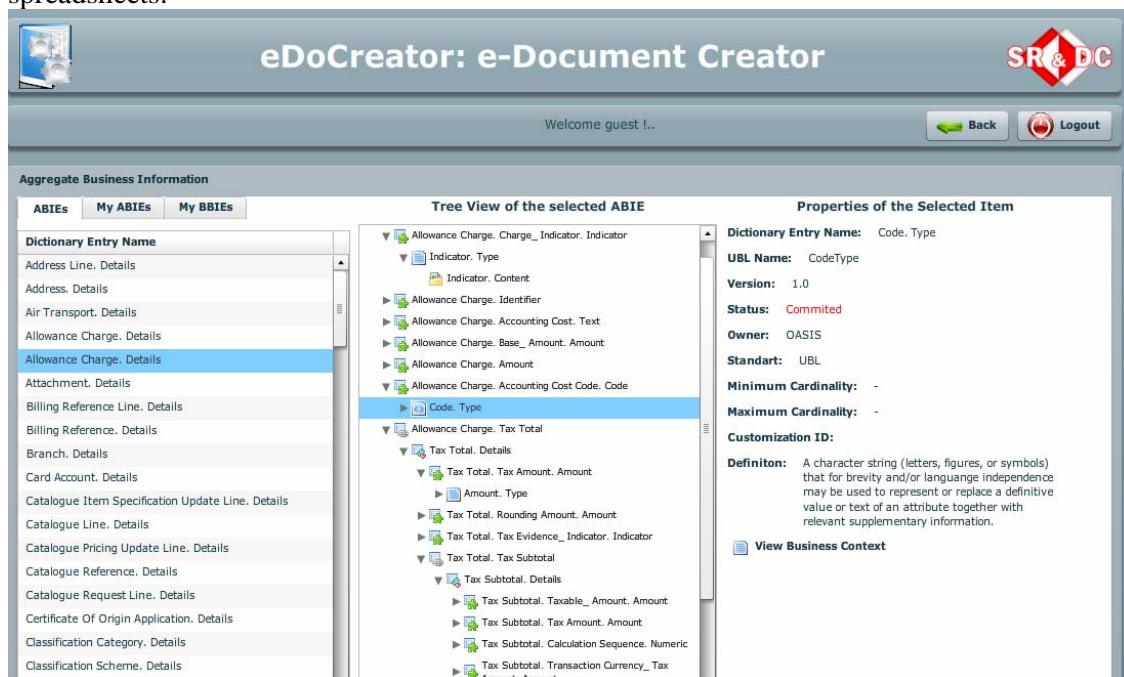


Figure 3 Tree view of the selected BBIE

The retrieval of the document artifacts are realized through graphical queries. The results list includes the standard document artifacts, the customized ones as well as the user created document components. The keywords to form the queries are gathered from users via a graphical interface. Furthermore, Boolean combinations of constraints can also be formed by connecting the constraints using logical Boolean “AND” and “OR” connectors.

## 5.2. Creating a new document artifact

The second functionality of the iSURF eDoCreator tool is the support it provides to create new document schemas by assembling available document building blocks by dragging and dropping components into target message schema as shown in Figure 4. The tool automatically locates the dragged component according to conventional ordering of UN/CEFACT CCTS: Basic Business Information Entities are listed before Associations Business Entities in the Message Assembly. Although the environment provides UN/CEFACT CCTS methodology steps, it does not mandate a strict sequence in applying these steps; rather it lets the users freely create the document models and guides them by notifying and reminding the steps that needs to be taken.

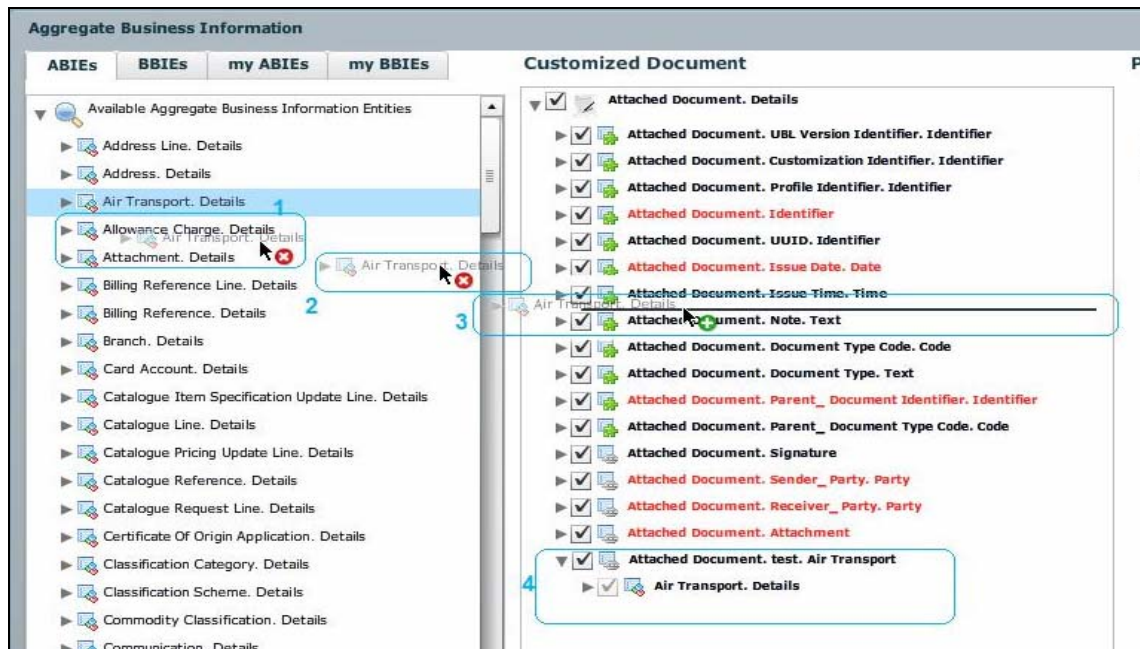


Figure 4 Creating new document artifact

### 5.3. iSURF eDoCreator Support for Conformant and Compatible Customizations

The modeling environment supports both the UBL Conformant Customization and the Compatible Customization. The customization starts by selecting a document artifact. It allows (1) subsetting source document model, (2) extending source document model, (3) constraining document artifacts and (4) creation of new document artifacts from scratch. In case an action invalidating the Conformance Design Rules is taken by the user, the tool warns the user that the design will not be conformant if the action is realized and lets the user withdraw the action if he wishes.

Subsetting source document model is done through visualization of document building block in a check tree. By selecting and deselecting tree nodes representing document building blocks, users can subset the document model. Extending source document is done by dragging and dropping available building blocks to assemble new building blocks. Constraining document artifacts can be done via subsetting the document model, or refining a data type. Creation of new document artifacts is described in Section 5.2.

### 5.4. iSURF eDoCreator Support for Refining a data type

A Data Type can be refined by modifying its content components or supplementary components. The available built-in restriction types that can be defined on content components vary according to type of the content component. For example, Text.Content can be restricted by setting a value on its length, maximum length, and minimum length. Supplementary components can be subsetting or their default/fixed or optional/required attributes can be set. Users can either create a new Qualified Data Type from an Unqualified Data Type or customize an existing Qualified Data Type to create a Qualified Data Type tailored to their needs by refining a data type.

### 5.5. iSURF eDoCreator Support for Documentation of Document Model

The designed conceptual model of the document is converted to a physical model according to the UBL Naming and Design Rules. The XML schema generator automatically detects dependencies in BIEs and generates XSD files for both document schema and the common artifacts. Furthermore, the tool generates spreadsheet models and generic code files in UBL directory format.

## 6. Business Benefit and Conclusions

iSURF eDoCreator provides a document modeling environment for users to assemble their documents from common components according to UN/CEFACT CCTS document modeling methodology and customize these documents using UBL 2.0 customization methodology. When the document building blocks are derived from a common semantic specification with well-defined rules, it becomes possible to achieve electronic business document interoperability.

iSURF eDoCreator tool has been used to derive the UBL 2.0 conformant eInvoice, Turkey [12] from the standard UBL 2.0 eInvoice. Such customizations are becoming popular recently, especially within the scope of the large scale integration project, PEPPOL (Pan-European Public Procurement Online) [10] currently being implemented in EU. PEPPOL will be producing UBL 2.0 conformant invoice, order, virtual company dossier and catalog schemas to be customized to the Member States and we believe that the publicly accessible iSURF eDoCreator tool provides an opportunity to help with these customizations. Additionally, the tool proved to be very useful in performing the gap analysis between NES/UBL Invoice [11] and eInvoice, Turkey [12].

## Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 213031, the iSURF Project.

## References

- [1] Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York (1990).
- [2] Kabak Y., Dogac A., "A Survey and Analysis of Electronic Business Document Standards", <http://www.srdc.metu.edu.tr/webpage/publications>, Submitted to ACM Computing Surveys.
- [3] UN/CEFACT Core Components Technical Specification – Part 8 of the ebXML Framework in Version 2.01, which is available at [http://www.unece.org/cefact/ebxml/CCTS\\_V2-01\\_Final.pdf](http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf)
- [4] Universal Business Language (UBL) 2.0, [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=ubl](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl)
- [5] iSurf eDoCreator: e-Business Document Design and Customization Environment. [http://www.srdc.com.tr/index.php?option=com\\_content&task=view&id=229&Itemid=203](http://www.srdc.com.tr/index.php?option=com_content&task=view&id=229&Itemid=203).
- [6] UN/CEFACT Releases XML Schema for Cross Industry Electronic Invoice (CII).
- [7] OASIS SET TC Deliverable: Semantic Representations of the UN/CEFACT CCTS-based Electronic Business Document Artifacts (Editors: Dogac, Kabak), <http://www.oasis-open.org/apps/org/workgroup/set/>
- [8] UBL Guidelines for Customization Version 1.0, 30 September 2008, <http://xml.coverpages.org/ni2007-04-04-a.html>
- [9] Adobe Flex, <http://www.adobe.com/products/flex/>
- [10] PEPPOL (Pan-European Public Procurement Online), <http://www.peppol.eu/>
- [11] NES. UBL Northern European Subset. <http://www.nesubl.eu/>
- [12] OASIS UBL Turkish Localization Subcommittee, [http://www.oasis-open.org/committees/sc\\_home.php?wg\\_abbrev=ubl-trlsc](http://www.oasis-open.org/committees/sc_home.php?wg_abbrev=ubl-trlsc)
- [13] UN/CEFACT Core Component Library, <http://www.unece.org/cefact/codesfortrade/uncl/CCL08B.xls>