Unified Data Modelling and Document Standardization Using Core Components Technical Specification for Electronic Government Applications

Yannis Charalabidis[1], Fenareti Lampathaki[2] and Dimitris Askounis[3]

Decision Support Systems Laboratory, National Technical University of Athens
[1] yannisx@epu.ntua.gr, [2] flamp@epu.ntua.gr, [3] askous@epu.ntua.gr

Abstract

In the effort of Governments worldwide to effectively transform manual into electronic services, semantic interoperability issues pose as a key challenge: system-to-system interaction asks for standardized data definitions, codification of existing unstructured information and a framework for managing governmental data in a unified way. Integrating and extending recent developments in Germany, Hong-Kong, UK and US governments - but also eBusiness interoperability research results in Europe, the proposed approach is applied within the Greek e-Government Framework and bears the following steps: (a) Gathering and analysis of governmental service forms at field level, so that common structures can be identified, (b) Customization of UN/CEFACT Core Components, (c) Semiautomatic XML Schema Definition (XSD) files creation, using sets of naming, structuring and verification rules, (d) Adoption of international or development of country-specific Codelists, (e) Development of guidance material for the application of the approach by government officials at central or local level. Discussion on the results is targeting best practice examples that can be drawn at scientific and organizational level, but also the key difficulties that have to be tackled – in relevance with the administrations and industry readiness in each country. Future steps include enhancing the paradigm in order to cover vertical data standards and applying the approach to other EU or associated countries.

Key words: eGovernment, Semantic Interoperability, Data Standardization, UN/CEFACT CCTS, UBL, Core Components

Unified Data Modeling and Document Standardization Using Core Components Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

1   Introduction

During the last years public administrations are striving to leverage modern information and communication technologies, in an effort to improve the quality of their services towards citizens and businesses, to provide multiple communication channels and to make their internal or cross-organizational processes more efficient. Furthermore, the definition of cross-border public sector services in the European Union, also known as Pan-European Electronic Government Services (PEGS), appears as a necessary step in order to achieve truly interoperable, one-stop electronic service provision at a pan-European context.   However, the successful provision of eGovernment applications does not only require the progress of services architectures, but also implies fundamental knowledge sharing, transparency of public sector information and requires a careful rethinking of the information resources in order to ensure that information is interpreted in the same way by communicating systems.   And this is where semantic interoperability issues come into the foreground.

As identified in relevant literature [32], semantic interoperability is a systematic way of interrelating information resources, so that we can allow for interpretation of pervasive and ever-growing information elements. Semantic interoperability is thus obtained when actors, including organizations, people or systems, are able to share meaning, which is to understand each other [20]. This implies sharing some context (e.g. definitions of terms, examples and counter examples, translations, etc.), which enables common interpretation of the data that is exchanged. It also implies that actors have some joint objectives, which justify the data exchange, hence contributing to give it some additional contextual meaning [19]. Achieving semantic interoperability has been recognized as an important research field both in the eBusiness [10] and the eGovernment research domain [8], further stressing the need for joint practices and knowledge sharing between relevant approaches [7].

In the area of eGovernment, semantic interoperability has to phase a multi-faceted problem: the number of existing documents that take part in the service provision from administrations towards citizens and businesses is usually at the level of several hundred, the owners of those documents may be different organizations (ministries, prefectures and municipalities, public sector organizations) and there is a lack of common fields standardization and adherence to common definitions, due to the still non-electronic nature of many public sector services.   Moreover, when services are being made electronic, usually the existing, diverse documents and forms are just transferred to an electronic format, resulting into non-interoperable - hence electronic - artefacts.   Based on these observations, standardized modelling of governmental data is confronted with the following key challenges:

- Although a plethora of data modelling standardization attempts exist in eBusiness, a lack of relevant standardization efforts for governmental documents, addressing how the underlying governmental information must be modelled, named and structured, is detected.

- The public sector, by size and scope of activities, represents the biggest single information content resource for the creation of value-added information content and services [11]. Businesses conduct the full spectrum of their transactions with the help of a few typical business documents in contrast to governmental bodies that deal with numerous diverse documents on a daily basis.

- Governmental documents often differ, according to the public body that issues them, even at the same country or region. This effect is also observed in parts of the documents (e.g. the address definition in the Citizen ID and in the Marriage Certificate).

In this context, the need for a step-by-step modelling methodology that will standardize and harmonize governmental documents and will try to articulate semantic interoperability emerges. Such a methodology needs to follow a long-term perspective, encourage creation, dissemination and reuse of existing components and schemas and ensure a proper evolution path and the reuse of internationally accepted standards.

To this direction, the present paper proposes a framework for managing and transforming governmental information elements that take part in service provision (application forms, documents, certificates, etc) in a unified way, standardizing data definitions and codifying existing unstructured information.   This framework is supported by specifically developed or customised existing tools and repositories, in an effort to provide an end-to-end solution for achieving semantic interoperability in large scale projects.   The presented methodology and toolset has been tested within the environment of the Greek eGovernment Interoperability Framework (eGIF) [13], a project that aims at effectively supporting the provision of one-stop shop, interoperable electronic services from National and Municipal organizations in Greece.

In the second section of this paper, the current state of the art in semantic interoperability in eGovernment is presented, identifying the novelty of the proposed approach. Section 3 introduces the proposed Governmental Data Modelling Methodology, while section 4 proceeds with the Greek eGIF XML Schema Library and Tools. Finally, conclusions and further research directions are provided in section 5.

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

2   Current State of the Art

Semantic, data and documents interoperability constitutes a research area that has attracted the interest of standardization bodies and initiatives carried out by e-government agencies in the interoperability arena and having produced corresponding interoperability frameworks.

As far as the standardization bodies are concerned, a state of the art approach promising to achieve a semantically unambiguous representation and usage of business information has been presented by the UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) with the release of the Core Component Technical Specification (CCTS), also known as ISO 15000-5 [37]. It includes a methodology for developing a common set of semantic building blocks that represent the general types of business data in use today and can be understood and interpreted by humans and machines in the same way while it provides for the creation of new business vocabularies and restructuring of existing ones. The UN/CEFACT Naming and Design Rules (NDR) [38] that accompany the CCTS specification define a set of guidelines for transforming CCTS based artefacts into XML Schema and XML based instances, while the UN/CEFACT Core Component Library (UN/CCL) [36] represents the repository for generic business data components, the so called Core Components. Based on the experiences gained in previous data standardization efforts, the CCL does not provide pre-determined, static or industry-specific data definitions, but comprises a huge set of context-agnostic, generally valid data templates (e.g. postal address, personal information) that are syntax-independent and represent the general business data entities which are commonly used in today's business processes and are the basis for creating contextualised entities (Business Information Entities) on eight unique and orthogonal context categories – business process, product classification, industry classification, geopolitical, official constraints, business process role, supporting role, system capabilities.

The Organization for the Advancement of Structured Information Standards (OASIS) has been the first to incorporate the CCTS methodology in its Universal Business Language (UBL) standard. UBL (Version 2.0 – December 2006) [31] is a royalty-free library of standard electronic XML business documents, designed to provide a universally understood and recognized commercial syntax for legally binding business documents. To this end, it provides a library of reusable data components based on an abstract Ur-Schema and a set of XML Schemas for 29 business documents. UBL has also defined guidelines for performing a compatible customization of UBL schemas together with suggestions for how to proceed when a compatible customization is not possible [30].

Implementing eGovernment Services is further specified and guided by eGovernment Interoperability Frameworks (eGIFs) that aim to facilitate the seamless exchange of information and the deployment of application-to-application interoperable systems in Central and Municipal Government, including service composition, security standards, certification of public sites and practitioners, as well as unified governmental data models for achieving a common understanding of semantics at a syntax-independent level. Upon consulting relevant literature [5], [6], [14], [28], [42] that summarizes the eGovernment interoperability approaches internationally, current national frameworks that have issued XML Schemas, guidelines for their design and customization or ontologies as a first step towards semantic interoperability are:

- In the United Kingdom, the e-Government Interoperability Framework [1] and its relevant specifications (e.g. the e-Government Metadata Standard [3] and the Schema Guidelines [2]) as issued by the e-Government Unit. The GovTalk XML Schema Library [4] contains approximately 78 XML Schemas (coming up to 102 with all versions taken into account). However, the majority of those Schemas are peripheral, do not refer to "core" governmental documents, like the Automated Registration of Title to Land (ARTL) Case Management Interface Schema and Schema for CON29 Land Searches, or map to simple components like address and person.

- In Germany, the Standards and Architectures for e-Government Applications (SAGA) [25] by the KBSt will be accompanied by the end of May 2008 by a XRepository that will contain a small number of XML Schemas for documents and a set of more than 50 codelists. An XGenerator tool that will automatically transform UML class diagrams into XML Schemas is also under way.

- In Denmark, the latest version of the Interoperability Framework [27] has been released by KIU since 2006. InfoStructureBase [26], the Danish collaboration tool promoting interoperability, includes an international standards repository containing business process descriptions, datamodel descriptions, interface descriptions, complex XML schemas and schema fragments (information object) from public and private organizations and an UDDI (Universal Description, Discovery and Integration) repository containing information on web services. The InfoStructureBase has launched a contract WSDL (Web Service Definition Language) tool that enables the user to create strong Web Service definitions based on new and existing Schema definitions found in the InfoStructureBase repository, without having to deal with the many complexities of using XML Schemas in WSDL files. In this context, the UBL artefacts (i.e. Business Information Entities and Documents) have been adopted as they are.

- In Hong Kong, the Information Technology Services Department (ITSD) (The Government of the Hong Kong Special Administrative Region (HKSAR)) has established the HKSARG Interoperability Framework (Version 5.1) [15]. The XML Co-ordination Group has developed an XML Schema Design and Management Guide while Common Schemas are published on the XML registry [Site 1]. Such schemas are in line with the Core

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

Components Technical Specification (CCTS) but incline more towards being characterized as codelists and core components and do not present any actual governmental documents.

- At a pan-European level, the European Interoperability Framework [20] which is currently being revised by IDABC [7] and the Architecture Guidelines for Trans-European Telematics Networks for Administrations (Version 7.1) [18] are met. As far as the Semantic Interoperability aspect is concerned, an EU-Project SEMIC.EU (Semantic Interoperability Centre Europe) [Site 3] has also been established in order to support the data exchange for pan-European eGovernment services, but has not released to date any interoperability assets.

- In the United States of America, the National Information Exchange Model (NIEM) [Site 2] is the outcome of the partnership between the U.S. Department of Justice and the Department of Homeland Security. It is designed to develop, disseminate and support enterprise-wide information exchange standards and processes that can enable jurisdictions to effectively share critical information in emergency situations, as well as support the day-to-day operations of agencies. NIEM has adopted an approach quite similar to the UN/CEFACT CCTS and has published today the broadest set of XML Schemas.

Further related work on semantic interoperability that has also been taken into account has been presented in [29], [32], [43], [44]. The gaps in the current state of the art that the proposed methodology is attempting to fill, lie in the following areas:

- Identification, reuse and extension of existing components in the eBusiness-oriented UN/CEFACT CCTS methodology, being one of the first large scale attempts internationally.

- Adoption of a service-driven approach, that implies binding the public administration documents with specific information exchanges among the stakeholders or systems, that take part during service provision to citizens and business – and not modelling of documents generally found within governmental systems.

- Development or customisation of tools that are needed in the modelling process, when large numbers of documents are to be modelled – including service and document registries, enterprise modelling tools, XML authorware and CCTS customisation tools.

3   Governmental Data Modelling Methodology

Within the computer science domain, Data Modelling is considered as the process of structuring and organizing data. While data analysis is a common term for data modelling, the activity actually has more in common with the ideas and methods of synthesis (inferring general concepts from particular instances) than it does with analysis (identifying component concepts from more general ones) [33]. Documents tend to obtain a broader scope and can be defined as any information exchange between stakeholders that take part in a service. They are conceived not only with the traditional meaning of the term (i.e. declarations, certificates, etc.), but contain queries and responses (acknowledgements) that are not typical documents as well.

The main objective of the proposed approach has been to enable mutual understanding and instant exchange of information between governmental entities, which have never collaborated directly before, through providing policies, practices and standards on electronic document modelling and formalization and by exploiting state-of-the-art data modelling tools and methodologies. This is to be accomplished through achieving the following goals:

- Creating a repository for semantics in terms of common data dictionaries and core components (based on the ISO 15000-5 standard - or UN/CEFACT CCTS - Methodology) for the most typically used structures, such as Address, Person and Organization.

- Developing semantically enriched Generic Governmental Documents, to be issued electronically and exchanged during the provision of:

  – The 20 basic public services towards citizens and businesses as defined by the i2010 Strategic Framework (e.g. Birth Certificate, Passport, VAT Statement)

  – The 100 core public services in Greece, stemming out of more than 9,000,000 actual service requests by citizens and businesses during the last 3 years and relating to almost 1,000 service types.

- Developing ontologies and metadata schemas for the semantic annotation of documents, so that all involved parties can easily locate and interpret the necessary information with the use of internet-based retrieval tools.

- Issuing guidelines for performing a compatible customization of the released XML Schemas to the specific public organization requirements together with suggestions for how to proceed when a compatible customization is not possible.

In this context, the adopted methodology runs over repeating cycles (spiral approach) in order to fully engage the public sector, bearing the following steps as depicted in Figure 1:

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

- *Step 1*: Selection of the documents to be modelled and documentation of their context, i.e. services that have a big impact and the potential to be provided as web services in the future. Modelling of the services during the execution of which each document is needed, between which stakeholders it is exchanged and if it contains any data fields that associate it with a specific public organization or geographic location.

- *Step 2*: Gathering and analysis of governmental service forms at field level, so that common structures can be identified. Details about the metadata recorded for each document are depicted in the following figure and can be attributed to the eGIF Services and Data Documentation Model [13] and partly to the Dublin Core [9] and the United Kingdom's Metadata Standards [3].
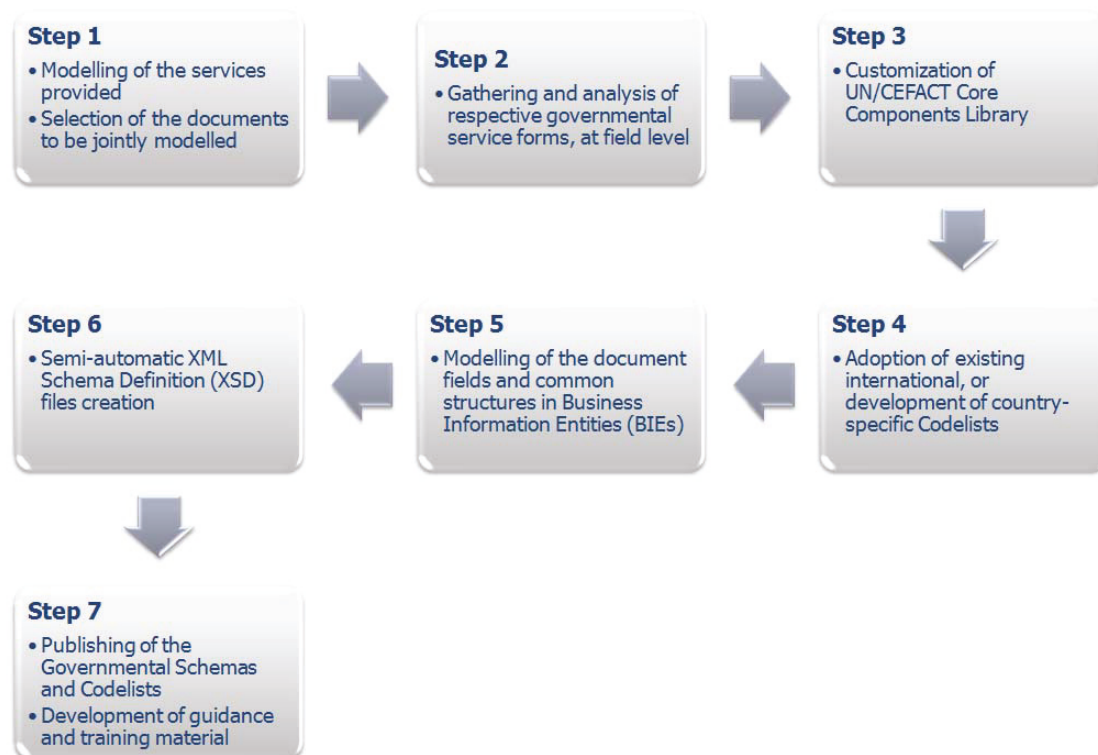

Figure 1: The 7 Steps of the Proposed Approach

- *Step 3*: Customization of UN/CEFACT Core Components in order to meet the needs of the public sector. Major changes that differentiate the UN/CEFACT CCL [36] and the Greek CCL are:

  – A set of Core Components, like Document, Query and Response, has been reinforced with additional fields, while specific components (i.e. Person, Organization, Contact, Event, Status, etc) have obtained a broader scope in order to cover any possible governmental information exchange

  – Components, such as Completed Work, Regulated or Hazardous Goods and Work Shift, that apply more in a business-to-business environment have not been reused

- *Step 4*: Adoption of international or development of country-specific Codelists that provide mappings between descriptions and codes for every field that has Code or – in certain cases – Identifier as its Representation Term. The following Codelists appear in the list of the approved international Codelists:

  – ISO-3166-1 for country codes [22]

  – ISO-4217 as the currency code list [23]

  – ISO 639-2 as the language code list [21]

  – UNECE Units of Measure used in International Trade [31]

  – UNECE Character Set Encoding Code [40] and IANA Character Set Code [16]

  – UNECE Agency Identification Code [39]

  – IANA MIME Media Type [17]

Additional adopted (custom) codelists indicatively refer to Gender, Profession Condition, Public Organization Type, Marital Condition and Vehicle Type.

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

- *Step 5*: Modelling of the document fields in Business Information Entities based on the Core Components Library. The concepts involved and recognized in this step are:
    - *Business Information Entities* as pieces of business data or groups of pieces of business data with a unique Business Semantic definition
    - *Unqualified Data Types* defined for all approved CCTS primary and secondary representation terms. They contain no additional restrictions on their source CCTs other than those defined in CCTS and agreed upon best practices.
    - *Qualified Data Types* that apply additional restrictions, like length or enumeration, on the Unqualified Data Types

The relations between Core Components, Business Information Entities, XML Schemas, Qualified and Unqualified Data Types are elaborated in Figure 2.
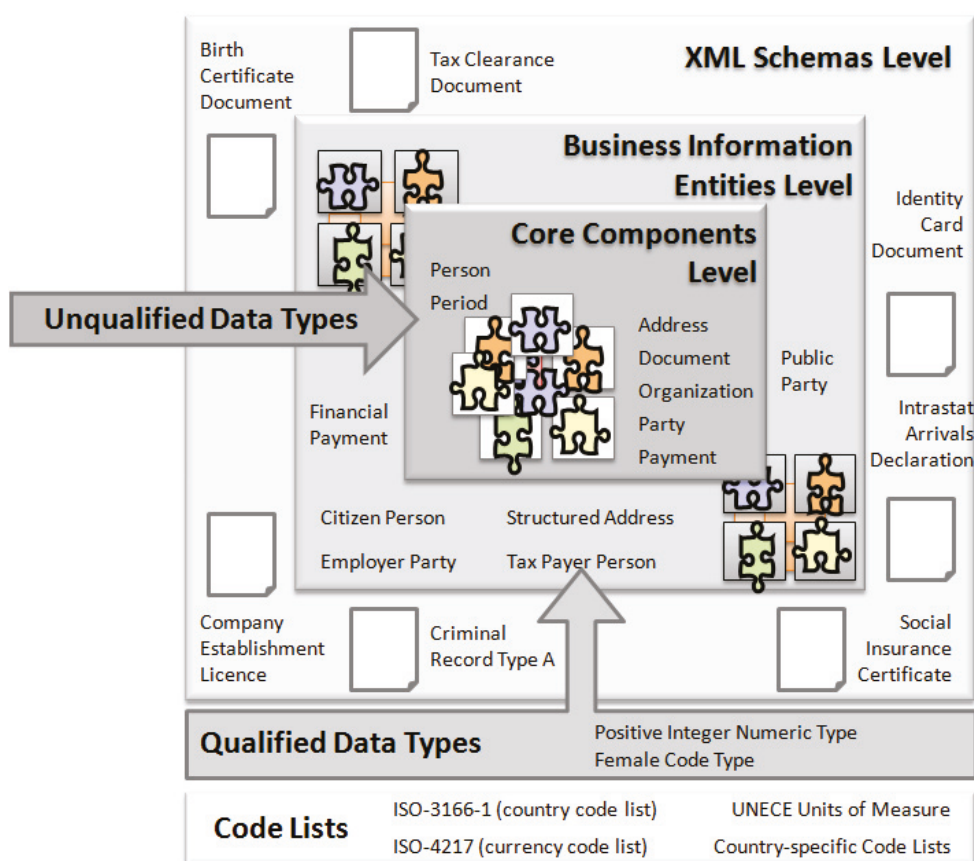

Figure 2: Basic Data Modelling Concepts and their Relations

- *Step 6*: Semiautomatic XML Schema Definition (XSD) files creation, using sets of naming, structuring and verification rules prescribed in the Services and Data Documentation Model of the Greek eGIF.

- *Step 7*: Publishing of the overall schema, after agreement of all relevant Public Administration Officials, development of guidance and training material for the application of the approach by government officials at central or local level.

The above steps of the methodology are to be performed in ever-enlarging sets of services and documents, starting from the core services and administrations (e.g. i2010 Services, relevant organizations and documents) and moving towards engaging services, administrations and documents of lesser importance or readiness.

4   The Greek eGIF XML Schema Library and Tools

4.1   Semantic Interoperability Toolset

In the case of the Greek eGIF, data modelling activities are supported with the help of a semantically-rich and context-aware Service and Documents Repository (the Interoperability Registry), going beyond the currently

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

available static data repositories and document management systems of public authorities and businesses, in their quest for interoperable, pan-European, standardized services. The overall Governmental Data Modelling Tools Framework extends over the following axes, as shown in Figure 3:

- The eGIF Interoperability Registry [34], an own-design database system implemented in Microsoft SQL Server and .net framework, comprising of:
    - A Services Repository with metadata definitions and detailed BPMN workflow diagrams.
    - A Document Repository containing documents and fields definitions and publishing the unified governmental data models (Core Components, Business Information Entities, XML Schemas) as well as their standard codelists.
    - A Web Services Repository that incorporates a set of functionalities of a typical UDDI registry, for inserting and getting metadata and schemas.
    - A Public Bodies and Information Systems Repository containing details about the public authorities in Greece and the information systems they have deployed.
- The Enterprise Modelling Platform, based on ADONIS Modelling Tool, by BoC International.
- The XML Management and Authorware Platform, based on XMLSpy and SchemaAgent tools, by Altova Inc.
- A custom-made CCTS-Excel-to-XSD Tool incorporating the eGIF Naming and Design rules, developed to accept the metadata in tabular form as input and provide the XML Schemas as output.
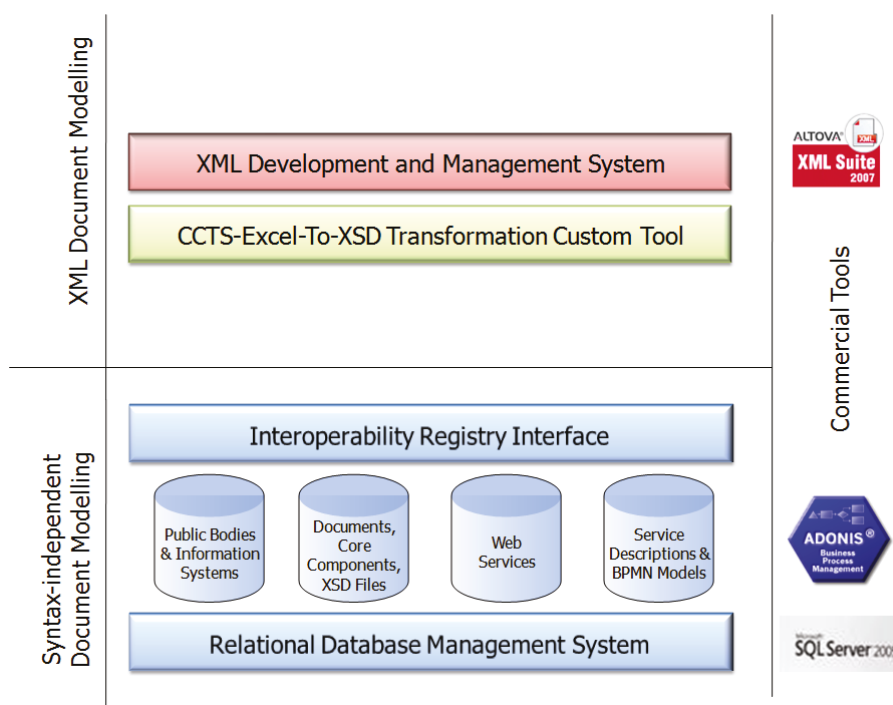

Figure 3: The Document Modelling Environment

The eGIF XML Schema Library built with the help of the aforementioned tools does not only meet the needs of the Greek public sector, but has the potential to be exploited in the context of pan-European Services (PEGS), since their definition has been expressed in the English language and only the content transferred within the actual XML documents is in the Greek language.

4.2   Modelling Principles

In order to enforce proper UN/CEFACT CCTS adoption in this new domain and uniform schema creation by all members of the team, but also involved Governmental Organizations in later cycles, a set of guidance rules were created. The most common such rules include:

i.   Metadata population in tabular form for Core Components, Business Information Entities and Qualified Data Types, as adopted   by the UN/CEFACT Naming and Design Rules [38]. A paradigm of a Business Information Entity that is also the root element for the document Passport can be found in Figure 4.

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

Unique ID	Name (eGIF XML Term)	Dictionary Entry Name (DEN)	ABIE/ BBIE/ ASBIE	Vers.	Definition	Object Class Term Qualifier	Object Class Term	Property Term Qualifier	Property Term
EL11030036	PassportDocument	Passport_ Document. Details	ABIE	1.0	Details about the travel document that uniquely identifies a person.	Passport	Document		
EL11030037	Name	Passport_ Document. Name. Name	BBIE	1.0	A name, expressed as text, for this specific passport.	Passport	Document		Name
EL11030038	BarcodeID	Passport_ Document. Barcode_ Identification. Identifier	BBIE	1.0	The barcode identifier for the passport.	Passport	Document	Barcode	Identification
EL11030039	IssueDateTime	Passport_ Document. Issue. Date Time	BBIE	1.0	The date time for the issuance of this passport.	Passport	Document		Issue
EL11030040	TypeCode	Passport_ Document. Type. Code	BBIE	1.0	A code specifying the type of the passport.	Passport	Document		Type
EL11030041	IssuerPublicOrganization	Passport_ Document. Issuer. Public_ Organization	ASBIE	1.0	The public organization that has issued this passport.	Passport	Document		Issuer
EL11030042	RelatedTravellerPerson	Passport_ Document. Related. Traveller_ Person	ASBIE	1.0	The person whom this passport identifies.	Passport	Document		Related
EL11030043	EffectiveSimplifiedPeriod	Passport_ Document. Effective. Simplified_ Period	ASBIE	1.0	The period during which the passport is effective.	Passport	Document		Effective

Associated Object Class Term Qualifier(s)	Associated Object Class Term	Repres. Term	Primitive Type	Qualified Data Type	Cardin. Min	Cardin. Max	Context: Business Process	Context: Organization	Context: Region	Related Term(s) in Greek
							In All Contexts	In All Contexts	Greece	Διαβατήριο
		Name	Name. Type	SimpleName100Type	0	1	In All Contexts	In All Contexts	Greece	Τίτλος Εγγράφου
		Identifier	Identifier. Type		1	1	In All Contexts	In All Contexts	Greece	Κωδικός / Αναγνωριστικό Εγγράφου
		Date Time	Date Time. Type	-	1	1	In All Contexts	In All Contexts	Greece	Ημερομηνία Έκδοσης
		Code	Code. Type		1	1	In All Contexts	In All Contexts	Greece	Τύπος Εγγράφου
Public	Organization				1	1	In All Contexts	In All Contexts	Greece	Εκδούσα Αρχή
Traveller	Person				1	1	In All Contexts	In All Contexts	Greece	Στοιχεία Πολίτη τον οποίο αφορά το Διαβατήριο
Simplified	Period				1	1	In All Contexts	In All Contexts	Greece	Περίοδος ισχύος διαβατηρίου

Figure 4: Example of the Document Passport in tabular form (CCTS template)

ii.   Consistency between the Document Views in tabular form and in formal XSD definition is ensured with the help of the CCTS-Excel-To-XSD Tool.

iii.   Compliance with the CCTS Modular Model that prescribes for the creation of the following separate XML schemas:

– A Root schema per document modelled (with prefix namespace rsm)

– A Reusable Aggregate Business Information Entities Module imported into every root schema (with prefix namespace ram) and containing all the definitions of the entities that are reused and shared among the documents

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

- A Qualified Data Type Module imported into the Reusable Aggregate Business Information Entities Module and every root schema (with prefix namespace qdt)

- An Unqualified Data Type Module imported into the Qualified Data Type Module, the Reusable Aggregate Business Information Entities Module and every root schema (with prefix namespace udt)

- A Schema per Codelist imported into the Qualified and Unqualified Data Type Modules (with prefix namespace clm<UniqueID of the CodeList>)

iv.   In order to create as more reusable and generic business information entities as possible, the required fields in a document have been eliminated to the minimum. To this end, every Core Component has been extended in one Business Information Entity per public organization.

v.   Data types have been detailed in qualified data types with particular facets, like length and pattern, whenever possible.

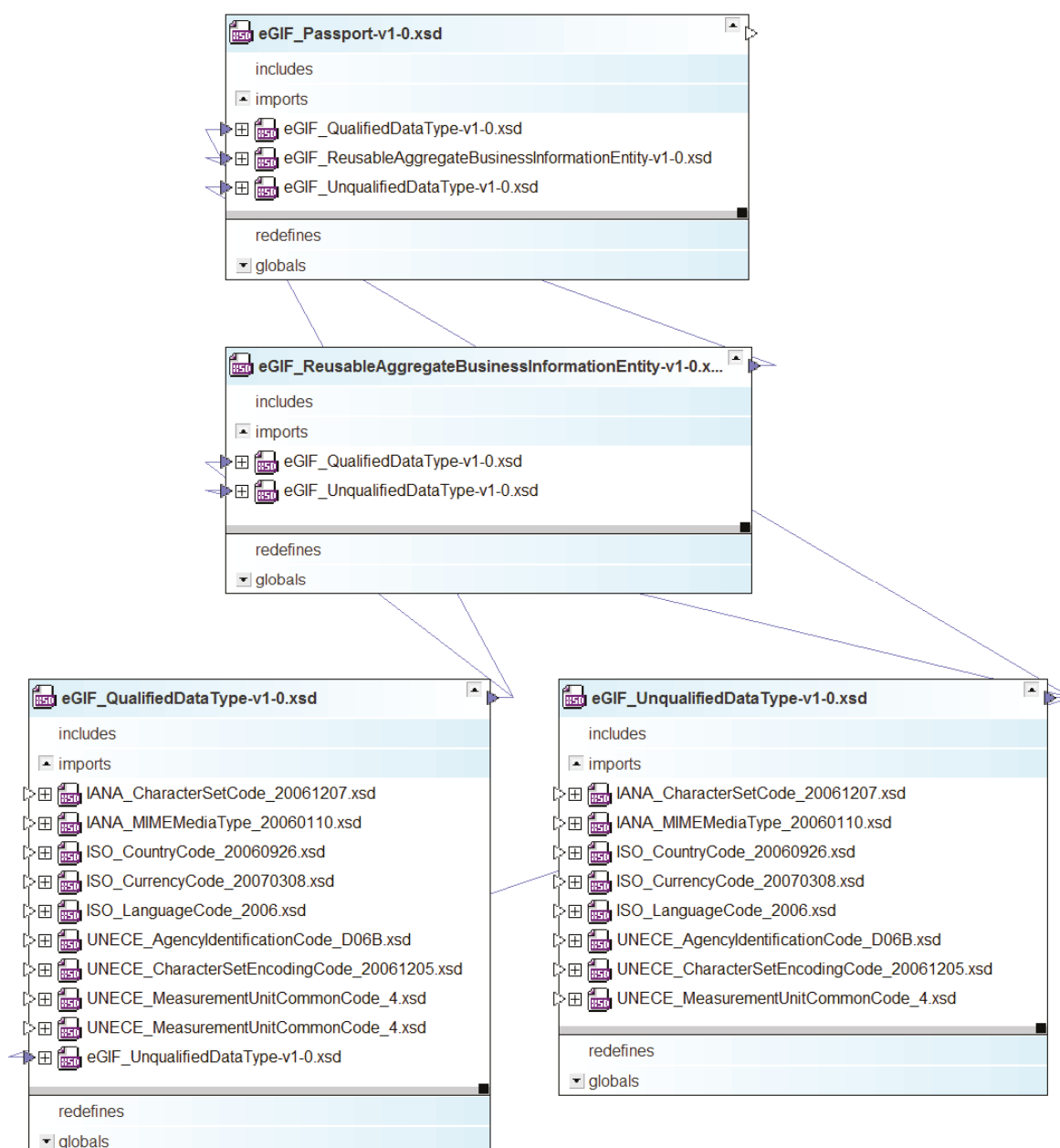Figure 5: Compliance of the XML Schemas with the CCTS Modular Model

46

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

vi.   Namespaces in the XML Schemas are declared as URNs (Unified Resource Names) following the pattern *<1st Level: Country>:<2nd Level: Responsible Public Body>:<3rd Level: Project>:<4th Level: Resource Type, i.e. Data, CodeList, etc.>:<5th Level: Resource Status, i.e. Draft, Standard>:<6th Level: Schema Module Type*, i.e. Root Schema Title, Reusable Aggregate Business Information Entity, Unqualified Data Type, Qualified Data Type, CodeList Title>:*<7th Level: Major Schema Version>*.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:ccts="urn:un:unece:uncefact:documentation:standard:CoreComponentsTechnicalSpecification:2"
xmlns:udt="gr:ypes:egif:data:draft:UnqualifiedDataType:1"
xmlns:ram="gr:ypes:egif:data:draft:ReusableAggregateBusinessInformationEntity:1"
xmlns:qdt="gr:ypes:egif:data:draft:QualifiedDataType:1" xmlns:rsm="gr:ypes:egif:data:draft:PassportDocument:1"
targetNamespace="gr:ypes:egif:data:draft:PassportDocument:1" elementFormDefault="qualified"
attributeFormDefault="unqualified" version="1-0" id="EL30000010">
        <xs:import namespace="gr:ypes:egif:data:draft:ReusableAggregateBusinessInformationEntity:1"
schemaLocation="..\Παράρτημα Δ - BIE\eGIF_ReusableAggregateBusinessInformationEntity-v1-0.xsd"/>
        <xs:import namespace="gr:ypes:egif:data:draft:UnqualifiedDataType:1" schemaLocation="..\Παράρτημα Ε -
DT\eGIF_UnqualifiedDataType-v1-0.xsd"/>
        <xs:import namespace="gr:ypes:egif:data:draft:QualifiedDataType:1" schemaLocation="..\Παράρτημα Ε -
DT\eGIF_QualifiedDataType-v1-0.xsd"/>
        <xs:element name="PassportDocument" type="rsm:PassportDocumentType" id="EL11030036">
                <xs:annotation>
                        <xs:documentation xml:lang="en">
                                <ccts:UniqueID>EL11030036</ccts:UniqueID>
                                <ccts:Acronym>ABIE</ccts:Acronym>
                                <ccts:DictionaryEntryName>Passport_ Document.
Details</ccts:DictionaryEntryName>
                                <ccts:Version>1.0</ccts:Version>
                                <ccts:Definition>The travel document that uniquely identifies a
citizen.</ccts:Definition>
                                <ccts:ObjectClassTermQualifier>Passport</ccts:ObjectClassTermQualifier>
                                <ccts:ObjectClassTerm>Document</ccts:ObjectClassTerm>
                                <ccts:BusinessProcessContextValue>In All
Contexts</ccts:BusinessProcessContextValue>
                                <ccts:OrganizationContextValue>In All Contexts</ccts:OrganizationContextValue>
                                <ccts:RegionContextValue>Greece</ccts:RegionContextValue>
                        </xs:documentation>
                </xs:annotation>
        </xs:element>
        <xs:complexType name="PassportDocumentType">
                <xs:sequence>
                        <xs:element name="Name" type="qdt:SimpleName100Type" id="EL11030037"
minOccurs="0">
                                <xs:annotation>
                                        <xs:documentation xml:lang="en">
                                                <ccts:UniqueID>EL11030037</ccts:UniqueID>
                                                <ccts:Acronym>BBIE</ccts:Acronym>
                                                <ccts:DictionaryEntryName>Passport_ Document. Name.
Name</ccts:DictionaryEntryName>
                                                <ccts:Version>1.0</ccts:Version>
                                                <ccts:Definition>A name, expressed as text, for this specific
passport</ccts:Definition>
                <ccts:ObjectClassTermQualifier>Passport</ccts:ObjectClassTermQualifier>
                                                <ccts:ObjectClassTerm>Document</ccts:ObjectClassTerm>
                                                <ccts:PropertyTerm>Name</ccts:PropertyTerm>
                                                <ccts:RepresentationTerm>Name</ccts:RepresentationTerm>
                                                <ccts:Cardinality>0..1</ccts:Cardinality>
                                                <ccts:BusinessProcessContextValue>In All
Contexts</ccts:BusinessProcessContextValue>
                                                <ccts:OrganizationContextValue>In All
Contexts</ccts:OrganizationContextValue>
                <ccts:RegionContextValue>Greece</ccts:RegionContextValue>
                                        </xs:documentation>
                                </xs:annotation>
                        </xs:element>
<!-- == Other BBIE elements declaration============================================================ -->
<!-- == Other ASBIE elements declaration============================================================ -->
                </xs:sequence>
        </xs:complexType>
</xs:schema>
```

Namespace Declaration

Metadata Documentation

Figure 6: Extract from the Passport XSD file

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

vii.   Three Context categories, i.e. Business Process, Organization and Region, have been extracted out of the 8 context categories that have been adopted by the CCTS and UBL. The default values for the context Region is Greece, whereas for the contexts Business Process and Organization is In All Contexts.

viii.   Upon design, XML Schemas for the Greek public sector are placed in the following maturity levels:

- Level 0 – Draft when a public organization creates a new XML Schema but the approval by the eGIF is pending

- Level 1 – Recommended for reuse when an XML Schema has been incorporated into the eGIF XML Schema Library

- Level 2 – Standard when an XML Schema has not changed version for more than 6 months and at least 3 public organizations have reused this Schema

ix.   In accordance with UBL, strict customization rules have been defined and customization procedures can redefine every BIE only once per context except for the cases when the context is restricted.

x.   Three levels of compliance have been recognized:

- Customization Fully Compliant with the eGIF, when all changes in the XML Schemas apply the eGIF guidelines by the book

- Customization Partly Compliant with the eGIF, when the changes in the XML Schemas have resulted in a non compatible Schema to the initial, but at least reuse some of its Business Information Entities

- Customization Compliant with the eGIF Core Components, when there is no published XML Schema to be based on and the new XML Schema creates Business Information Entities on the basis of the Core Components from scratch.

4.3   Population of Schema Library

The first cycle of application of the presented methodology in the context of the Greek eGovernment Interoperability Framework yielded components for the Core-100 services of the Greek Public Sector. Those services include all the i2010 Core Services and cater for 85% of the service requests by citizens and businesses over the last years. Initial population of the Schema Library is presented in Table 1.

Table 1: Elements of the Greek eGIF Schema Library

Entity	Population
XML Schemas for Documents	93
Core Components	36
Reusable Business Information Entities	86
Qualified Data Types	46
Unqualified Data Types	20
Context Categories	3
Total Documents in the Registry	1,106

At this stage, the XML Schemas and the various Core Components developed are being adopted by public sector organizations (Ministries, Prefectures, Municipalities, Social Security Organizations, Universities, etc) for the development of web services requesting or providing the necessary documents in electronic form – thus gradually ceasing to request such intermediate documents from citizens or businesses during service provision.

5   Conclusions

The presented methodological framework focused on the key challenges of unified data modelling of Governmental Documents, for the provision of interoperable, one-stop shop services to citizens and businesses. Utilising practices and standardization from the eBusiness domain, the presented approach is claiming novelty in conceptualisation and overall coherence in the eGovernment domain, being one of the very first approaches internationally to bring the power of UN/CEFACT CCTS structure into a real-life application. Applied at large scales within the Greek eGovernment Interoperability Framework, the approach presents a sound collection of reusable principles and tools for other Governments and practitioners of the field, as following:

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

- An overall approach compatible with 2nd generation eGIFs, going beyond paper-based standardization to live systems and service / document registries.

- A 7-Step concise methodology to tackle the problem of creating unified, structured XML schemas for Governmental documents.

- Standardized Core Components, reusable BIEs and almost 100 XSD files (including Passports, Citizen ID, Driving License, Citizen Certificates (birth, marriage, civil status, death), VAT, Income and Intrastat Declarations, Criminal Record, Social Security Contributions and Benefits, Payments, etc) to be freely available within the Greek eGIF Web Portal.

- A set of naming and coding standards and principles for the uniform creation and maintenance of XML descriptions.

- More than 10 adopted or developed standard codelists for the most common values appearing in public forms and documents.

- A set of prototype tools that can be adapted to other cases and Governments' needs.

Problems faced during implementation and application were not trivial and have to be to be taken in mind during relevant attempts by government officials and practitioners. Tools development or integration is usually such an issue in relevant attempts, the present one not being excluded from this de-facto rule: significant effort has to be devoted to the development of the registry application that forms the central repository for governmental services and documents – as no commercial, ready to apply tools are generally available. Furthermore, integration of enterprise modelling tools (in our case BoC ADONIS platform) and XML authoring tools (Altova XMLSpy Suite) with the core registry should be performed with caution and supported by high-level technical support from the vendors. On the metadata creation field, language issues are extremely important as all the relevant descriptions should be in local language – for the government officials to understand, modify and approve) and at least in English (for easiness of communication with other governments and practitioners). One of the best ways to tackle this requirement is the creation of a glossary early in the process. Finally, adequate time and effort should be spent for communicating and working together with government officials at various levels, for the actual agreement on the schemas and for the final adoption. The use of eParticipation and eCollaboration tools, on top of the internet-based registry system has proven to be a worthwhile track in this direction.

Future steps of the team in the direction of the presented approach and also within the adoption phase of the Greek eGIF are the following:

- Tackling of the legal issues for the formal adoption of the new electronic documents in everyday practice, since most of the governmental services are ruled by specific laws and decrees – a case which is common in many European countries. Towards this goal, the legal elements ontology (already defined in the Interoperability Registry [34]) has to be fully populated with all the legal elements affecting and ruling services and then processed by mixed teams of public administration legal counsels and modelling engineers.

- Further applications of the "7 Steps Cycle" in order to cover more services and documents, as still a fraction of the existing manual services and their documents have been modelled. This task is specifically going to interfere with the so-called "vertical standards", that is document standards for Health, Banking, Telecommunications or Defence – to name a few of the vertical industries found within the public sector.

- Extension of the collaborative and participative functionalities of the Interoperability Registry, in an effort to accommodate the work of several committees and sub-committees in an electronic, ubiquitous platform.

Finally, application of the methodology has already been agreed with the Lithuanian eGovernment Interoperability Framework and discussions are being held with a number of European Union and Associated countries governmental organizations – in an effort to further adapt and test the principles, the tools and the assumptions.

Websites List

Site 1: HKSARG Common Schemas
http://www.xml.gov.hkhttp://www.jtaer.com/

Site 2: NIEM
http://www.niem.gov/index.php

Site 3: SEMIC.EU
http://www.semic.eu

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

References

[1] Cabinet Office – e-Government Unit, UK Government. (2005) e-Government Interoperability Framework, Version 6.1, [Online], Available: http://www.govtalk.gov.uk/documents/eGIF%20v6_1(1).pdf.

[2] Cabinet Office – e-Government Unit, UK Government. (2004) e-Government Schema Guidelines for XML, Version 3.1, [Online], Available: http://www.govtalk.gov.uk/documents/schema-guidelines-3_1(1).pdf.

[3] Cabinet Office – Office of the e-Envoy, UK Government. (2006) e-Government Metadata Standard, Version 3.1, [Online], Available: http://www.govtalk.gov.uk/documents/eGMS%20version%203_1.pdf.

[4] Cabinet Office, UK Government. (2008) UK GovTalk Schema Library, [Online], Available: http://www.govtalk.gov.uk/schemasstandards/schemalibrary.asp.

[5] Y. Charalabidis, F. Lampathaki, D. Askounis, A. Stassis, Shifting to Second Generation E-Government Interoperability Frameworks, in Proceedings of the International EGOV 2007 Conference, Regensburg (Germany), September 3-7, 2007.

[6] Y. Charalabidis, M. Tschichholz, A. Hopkirk, Advancing the eGovernment Interoperability Framework in European Countries: Architectures, Challenges and Perspectives from the new Greek eGIF, in Proceedings of the eChallenges 2007 Conference, The Hague, The Netherlands, October 24 – 26, 2007.

[7] Y. Charalabidis, Workshop on Interoperable Services for Businesses and Governments, iESA Conference on Interoperability of Enterprise Systems and Applications (IEEE, IFIP, ACM), Madeira, Portugal, March 26-30, 2007.

[8] C. Codagnone, M. Wimmer (Eds.), Roadmapping eGovernment Research. Visions and Measures towards Innovative Governments in 2020, eGovRtd2020 Project, [Online], Available: http://www.egovrtd2020.org/navigation/results.

[9] Dublin Core Metadata Initiative. (2008) [Online], Available: http://dublincore.org.

[10] European Commission, DG Information Society and Media, Enterprise Interoperability Research Roadmap, Version 5.0, March 2008

[11] European Commission, Public Sector Information: A Key Resource For Europe, Green Paper on Public Sector Information in the Information Society, COM(1998)585

[12] Gartner Group. (2007) Preparation for Update European Interoperability Framework 2.0 - Final Report, [Online], Available: http://ec.europa.eu/idabc/servlets/Doc?id=29101.

[13] Greek Ministry of Interior. (2008) Greek eGovernment Interoperability Framework, [Online], Available: http://www.e-gif.gov.gr.

[14] L. Guijarro, Interoperability frameworks and enterprise architectures in e-government initiatives in Europe and the United States, Government Information Quarterly, vol. 24, no. 1, pp. 89-101, January 2007.

[15] HKSARG (Hong Kong Special Administrative Region). (2007, July) Interoperability Framework, Version 5.1, [Online], Available: http://www.ogcio.gov.hk/eng/infra/download/s18.pdf.

[16] IANA Character Set Code, [Online], Available: http://www.unece.org/uncefact/codelist/standard/IANA_CharacterSetCode_20050128.xsd.

[17] IANA MIME Media Type, [Online], Available: http://www.unece.org/uncefact/codelist/standard/IANA_MIMEMediaType_20060110.xsd.

[18] IDA. (2004) IDA Architecture Guidelines For Trans-European Telematics Networks for Administrations, Version 7.1, [Online], Available: http://ec.europa.eu/idabc/en/document/3485/5585.

[19] IDABC. (2005) Content Interoperability Strategy, Working Paper, [Online], Available: http://ec.europa.eu/idabc/servlets/Doc?id=24405.

[20] IDABC. (2004) European Interoperability Framework for pan-European e-Government Services, Version 1.0, [Online], Available: http://europa.eu.int/idabc/en/document/3761.

[21] ISO 639-2 (language code list), [Online], Available: http://www.loc.gov/standards/iso639-2/php/code_list.php.

[22] ISO-3166-1 (country code list), [Online], Available: http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html.

[23] ISO-4217 (currency code list), [Online], Available: http://www.iso.org/iso/en/prods-services/popstds/currencycodeslist.html.

[24] J.A. Mykkanen, M.P. Tuomainen. (2008) An evaluation and selection framework for interoperability standards, Information and Software Technology, vol. 50, no. 3, pp. 176-197, February 2008.

[25] KBSt unit at the Federal Ministry of the Interior. (2006) SAGA Standards and Architectures for e-Government Applications, Version 3.0, [Online], Available: http://www.kbst.bund.de/cln_012/nn_836802/SharedDocs/Anlagen-kbst/Saga/saga__3__0,templateId=raw,property=publicationFile.pdf/saga_3_0.pdf.

[26] KIU. (2008) Danish e-Government Project, InfoStructureBase, [Online], Available: http://isb.oio.dk/info.

[27] KIU. (2006) Danish Interoperability Framework, Version 1.2.14, [Online], Available: http://standarder.oio.dk/English/Guidelines.

[28] MODINIS. (2007) Study on Interoperability at Local and Regional Level, Version 2.0, [Online], Available: http://www.epractice.eu/files/media/media1309.pdf.

[29] S. Mouzakitis, F. Lampathaki, C. Schroth, U. Scheper, T. Janner, Towards a common repository for governmental data: A modelling framework and real world application, in Enterprise Interoperability II: New Challenges and Approaches (Springer), in Proceedings of the 3rd International Conference Interoperability for Enterprise Software and Applications I-ESA 2007, Funchal (Madeira Island) Portugal, March 2007

[30] OASIS, Guidelines for the Customization of UBL v1.0 Schemas, [Online], Available: http://docs.oasis-open.org/ubl/cd-UBL-1.0/doc/cm/wd-ubl-cmsc-cmguidelines-1.0.html.

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis

[31] OASIS. (2006, December) Universal Business Language (UBL) Version 2.0, Standard, [Online], Available: http://docs.oasis-open.org/ubl/os-UBL-2.0.zip.

[32] M. Selvage, D. Wolfson, B. Zurek, E. Kahan. (2006, June) Achieve semantic interoperability in a SOA: Patterns and best practices, IBM, [Online], Available: http://www.ibm.com/developerworks/webservices/library/ws-soa-seminterop.html.

[33] G. Simson G., G.C. Witt, Data Modelling Essentials, Third Edition, Morgan Kaufmann Publications, Elsevier, 2005.

[34] A.-M. Sourouni, F. Lampathaki, S. Mouzakitis, Y. Charalabidis, D. Askounis, Paving the way to eGovernment Transformation Interoperability Registry Infrastructure Development, in Proceedings of 7th EGOV Conference 2008, LNCS vol. 5184, pp. 340-351, 2008.

[35] G. Stuhec, How to solve the Business Standards Dilemma – the Context Driven Business Exchange, SAP Developer Network, 2005.

[36] UN/CEFACT. (2006) Core Component Library (UN/CCL), version 1.0, [Online], Available: http://www.unece.org/cefact/codesfortrade/codes_index.htm.

[37] UN/CEFACT. (2003, November) Core Components Technical Specification, Part 8 of the ebXML Framework, Version 2.01, [Online], Available: http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf.

[38] UN/CEFACT. (2006, February) XML Naming and Design Rules, Version 2.0, [Online], Available: http://www.unece.org/cefact/xml/XML-Naming-and-Design-Rules-V2.0.pdf.

[39] UNECE Agency Identification Code, [Online], Available: http://www.unece.org/uncefact/codelist/standard/UNECE_AgencyIdentificationCode_D05A.xsd.

[40] UNECE Character Set Encoding Code, [Online], Available: http://www.unece.org/uncefact/codelist/standard/UNECE_CharacterSetEncodingCode_2005.xsd.

[41] UNECE Units of Measure used in International Trade, [Online], Available: http://www.unece.org/cefact/recommendations/rec20/rec20_rev3_Annex2e.pdf.

[42] United Nations Development Programme (UNDP). (2007) e-Government Interoperability: A Review of Government Interoperability Frameworks in Selected Countries, [Online], Available: http://www.apdip.net/projects/gif.

[43] F. Van Blommestein, B. Broeksema, UN/CEFACT Core Components as the basis for structured business communication by SMEs, employing auto-generated, user adjustable forms, Enterprise Interoperability II New Challenges and Approaches (Ricardo J. Gonçalves, Jörg P. Müller, Kai Mertins and Martin Zelm eds.), 2007.

[44] Y. Yarimagan, A. Dogac, Semantics based customization of UBL document schemas, Distributed and Parallel Databases, vol. 22, no. 2-3, December 2007.

Unified Data Modeling and Document Standardization Using Core Components
Technical Specification for Electronic Government Applications

Yannis Charalabidis
Fenareti Lampathaki
Dimitris Askounis