

2009

# Enabling Semantic Interoperability in e-Government: A System-based Methodological Framework for XML Schema Management at National Level

Fenareti Lampathaki

*National Technical University of Athens, flamp@epu.ntua.gr*

George Gionis

*National Technical University of Athens, gionis@epu.ntua.gr*

Sotiris Koussouris

*National Technical University of Athens, skoussouris@epu.ntua.gr*

Dimitris Askounis

*National Technical University of Athens, askous@epu.ntua.gr*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2009>

## Recommended Citation

Lampathaki, Fenareti; Gionis, George; Koussouris, Sotiris; and Askounis, Dimitris, "Enabling Semantic Interoperability in e-Government: A System-based Methodological Framework for XML Schema Management at National Level" (2009). *AMCIS 2009 Proceedings*. 620.

<http://aisel.aisnet.org/amcis2009/620>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Enabling Semantic Interoperability in e-Government: A System-based Methodological Framework for XML Schema Management at National Level

**Fenareti Lampathaki**

National Technical University of Athens  
flamp@epu.ntua.gr

**George Gionis**

National Technical University of Athens  
gionis@epu.ntua.gr

**Sotiris Koussouris**

National Technical University of Athens  
skoussouris@epu.ntua.gr

**Dimitris Askounis**

National Technical University of Athens  
askous@epu.ntua.gr

## ABSTRACT

Articulating semantic interoperability in e-Government remains in question as long as the international standardization efforts do not reach a consensus on how to semantically annotate and exchange data, but merely focus on the syntactic aspects by publishing sets of XML Schemas. As one-stop governmental services at national and cross-county level become an imperative, the need for standardized data definitions, codification of existing unstructured information and a framework for managing governmental data in a unified way emerges. Effectively applied to the Greek e-Government National Interoperability Framework, this paper proposes a methodology for designing semantically enriched XML Schemas with which homogenized governmental information complies, based on the UN/CEFACT Core Components Technical Specification (CCTS). A discussion around a prospective architecture for managing large sets of XML Schemas is also motivated in order to recognize the necessary components and the key issues that need to be tackled when designing a Governmental Schema Registry.

## Keywords

e-Government, Semantic Interoperability, Data Standardization, UN/CEFACT CCTS, XML Schema, Core Components.

## INTRODUCTION

As Web evolves from global hypertext system to distributed platform for end-user interaction with the help of Web 2.0 tools (mash-ups, service front-ends, social software, etc.) (Schroth, 2007) and the core of Semantic Web matures (Hall et al., 2009), interoperability still remains a key enabler and a grand challenge to overcome (Charalabidis et al., 2008b). Current advancements in the e-Government domain (i.e. EU Services Directive 2006/123/EC and i2010) call, at the same time, for effective implementation of cross-border public sector information and interactive services, also known as pan-European public services (PEGS), which make the need for interoperability resolution more urgent and complex due to the additional language and legal barriers erected. As argued by (Ralyte et al., 2008), true interoperability is not yet here since organisations running different applications built with different designs and architectures still have difficulties talking to each other.

From the various types of interoperability, the semantic aspect is a particularly vexing problem that has arisen due to a disagreement about the meaning, interpretation, or intended use of the same or related data (Cardoso and Sheth, 2006). Following the trend that the cornerstone of government information exchanges is a common understanding of semantics at a syntax-independent level (Lampathaki et al., 2008), multiple interoperability frameworks have put semantic interoperability on the agenda (Bednar et al., 2008); for example, according to the European Interoperability Framework (EIF) (IDABC, 2004; 2005; 2008), semantic interoperability “enables systems to combine received information with other information resources and to process it in a meaningful manner”. Semantic interoperability is thus defined as a systematic way of interrelating information resources, so that we can allow for interpretation of pervasive and ever-growing information elements (Selvage et al., 2006), while it is also fraught with potential unexpected legal and social ramifications (Hovy, 2008).

In the area of e-Government, 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 (Charalabidis et al., 2008a).

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.

In this direction, the present paper proposes a framework and a prospective architecture 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. The presented modelling methodology has been tested within the environment of the National Interoperability Framework (NIF) in Greece (Greek Ministry of Interior, 2009), a project that aims at effectively supporting the provision of one-stop shop, interoperable electronic services from National and Municipal organizations in Greece.

## BACKGROUND

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 (UN/CEFACT, 2003). 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) (UN/CEFACT, 2006b) 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) (UN/CEFACT, 2006a) represents the repository for generic business data components, the so called Core Components.

Current National Interoperability Frameworks (NIF's) that have issued semantic interoperability assets (i.e. XML Schemas, Ontologies, Code Lists) or guidelines for their design as a first step towards semantic interoperability are:

- In Australia, the Australian Government Interoperability Framework (AGIF) issued and maintained by the Australian Government Information Management Office (AGIMO, 2008) is supplemented by GovDex, a secure internet based space for government agencies to deliver information in one central location, and Government Information Exchange Methodology (GIEM) as a Whole of Government desktop modelling tool.
- In Denmark, the Interoperability Framework is released by KIU. InfoStructureBase, 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 and an UDDI (Universal Description, Discovery and Integration) repository.
- At a pan-European level, the European Interoperability Framework which is currently being revised by IDABC is met. As far as the Semantic Interoperability aspect is concerned, an EU-Project SEMIC.EU (Semantic Interoperability Centre Europe) has also been established in order to support the data exchange for pan-European e-Government services, and has published to date interoperability assets coming from Denmark, Germany and France.
- In Germany, the Standards and Architectures for e-Government Applications (SAGA) by the KBSt are accompanied by the XRepository, that contains XML Schemas, Core Components and codelists, and the XGenerator tool that automatically transforms UML class diagrams into XML Schemas.
- In Hong Kong, the Information Technology Services Department (ITSD) has established the HKSARG Interoperability Framework (Version 5.1) and an XML Schema Design and Management Guide while Common Schemas are published on the XML registry. Such schemas are in line with the Core Components Technical Specification (CCTS) but incline more towards being characterized as codelists and core components and do not present any actual governmental documents.
- In the United Kingdom, the e-Government Interoperability Framework and its relevant specifications (e.g. the e-Government Metadata Standard and the Schema Guidelines) are issued by the e-Government Unit. The GovTalk XML

Schema Library contains a set of XML Schemas most of which are peripheral, do not refer to “core” governmental documents or just map to simple components like address and person.

- In the United States of America, the National Information Exchange Model (NIEM) is the outcome of the partnership between the U.S. Department of Justice and the Department of Homeland Security (2009) in order to develop, disseminate and support enterprise-wide information exchange standards and processes. NIEM has adopted a component-based approach quite similar to the UN/CEFACT CCTS and has published today the broadest set of XML Schemas.

The gaps in the current state of the art that this paper is attempting to fill, lie in the following areas:

- A methodology for data modelling that on the one hand promotes homogenization and proposes a canonical semantically-enriched XML Schema format for exchanging data instead of complex 1-to-1 matchings; and on the other hand identifies, reuses and extends existing components in the eBusiness-oriented UN/CEFACT CCL.
- 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.
- An architecture for a Governmental Schema Registry that takes into account current advances in XML modelling, storage and matching and can indeed promote semantic interoperability in the public sector, especially when large numbers of documents are to be managed.

### 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) (Simson and Witt, 2005). 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 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 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) and the 460 public services in Greece, modelled today in the Interoperability Registry (Sourouni, et al., 2008).
- 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:

- *Step 1:* Selection of the documents to be modelled and documentation of their context, i.e. 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 documents and service forms at field level based on the metadata of the Documentation Model for Public Administration Processes and Data (Greek Ministry of Interior, 2009), so that common structures can be identified.
- *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 and the Greek CCL and were identified upon the initial documents screening are:

- A set of Core Components, like Document, Query and Response, has been reinforced with additional fields, distinguishing, for example, between the Person, Party and Organization as Issuer, Addressee and Applicant and including references to the majority of the remaining Core Components.
  - Specific Core Components (i.e. Person, Organization, Contact, Event, Status, etc) have obtained a broader scope in order to cover any possible governmental information exchange.
  - New Core Components added in the Greek CCL include the Digital Signature of a physical Person or a Contact and the Attachment,
  - Core 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. Indicatively, the following Codelists appear in the list of the approved international Codelists: ISO-3166-1 for country codes, ISO-4217 as the currency code list and ISO 639-2 as the language code list, while additional adopted (custom) codelists indicatively refer to Gender, Marital Condition, Public Organization Type and Vehicle Type.
  - *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.
  - *Step 6:* Semiautomatic XML Schema Definition (XSD) files creation, using sets of naming, structuring and verification rules prescribed in the Documentation Model for Public Administration Processes and Data of the Greek NIF.
  - *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 current cycle of application of the presented methodology in the context of the Greek e-Government Interoperability Framework yielded components for the Core (1<sup>st</sup> Cycle)-100 and 2<sup>nd</sup> Cycle-360 services of the Greek Public Sector. Current population of the Greek NIF Schema Library includes: 404 XML Schemas for Documents, 36 Aggregate Core Components, 154 Reusable Aggregate Business Information Entities, 46 Qualified Data Types, 20 Unqualified Data Types, 16 Code Lists and 3 Context Categories.

Such a Schema Library 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.

## GOVERNMENTAL XML SCHEMA LIBRARY AND ARCHITECTURE

### 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 in the Documentation Model for Public Administration Processes and Data in order to summarize the gained experience and provide best practices to anyone seeking for semantic interoperability guidelines. The most common such rules include:

- 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. An example of a Business Information Entity that is also the root element for the document Driving Licence can be found in Figures 1 and 2.
- Consistency between the Document Views in tabular form and in formal XSD definition.
- Compliance with the CCTS Modular Model that prescribes for the creation of the following separate XML schemas:
  - A Root schema per document modelled

- A Reusable Aggregate Business Information Entities Module imported into every root schema and containing all the definitions of the entities that are reused and shared among the documents
- A Qualified Data Type Module imported into the Reusable Aggregate Business Information Entities Module and every root schema
- An Unqualified Data Type Module imported into the Qualified Data Type Module, the Reusable Aggregate Business Information Entities Module and every root schema
- A Schema per Codelist imported into the Qualified and Unqualified Data Type Modules

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	Associated Object Class Term Qualifier(s)	Associated Object Class Term
EL11030028	DrivingLicenceDocument	Driving Licence_ Document. Details	ABIE	1.1	The driving licence document represents the certificate that authorities handle in a driver in order for him to drive a vehicle.	Driving Licence	Document				
EL11030029	EULicenceID	Driving Licence_ Document. EU Licence_ Identification. Identifier	BBIE	1.1	A unique identification for this licence at EU level.	Driving Licence	Document	EU Licence	Identification		
EL11030030	IssueDateTime	Driving Licence_ Document. Issue. Date Time	BBIE	1.1	The date, time, date time or other date time value for the issuance of this document.	Driving Licence	Document		Issue		
EL11030031	ReceiptDateTime	Driving Licence_ Document. Receipt. Date Time	BBIE	1.1	The date, time, date time or other date time value for the formal receipt of this document.	Driving Licence	Document		Receipt		
EL11030032	Remarks	Driving Licence_ Document. Remarks. Text	BBIE	1.1	A remark, expressed as text, regarding this document.	Driving Licence	Document		Remarks		
EL11030033	IssuerPublicOrganization	Driving Licence_ Document. Issuer. Public Organization	ASBIE	1.1	The issuer organization of the driving licence.	Driving Licence	Document		Issuer	Public	Organization
EL11030034	RelatedDriverPerson	Driving Licence_ Document. Related. Driver Person	ASBIE	1.1	The driver is the person to whom this licence refers.	Driving Licence	Document		Related	Driver	Person
EL11030035	ReferenceVehicleCategoryStatus	Driving Licence_ Document. Reference. Vehicle Category_ Status	ASBIE	1.1	The various vehicle categories associated with this driving licence.	Driving Licence	Document		Reference	Vehicle Category	Status

Figure 1: Document Driving Licence in tabular form (1<sup>st</sup> part of fields)

Unique ID	Name (eGIF XML Term)	Representation Term	Primitive Type	Qualified Data Type	Cardin. Min	Cardin. Max	Context: Business Process	Context: Organization	Context: Region	Related Term(s) in Greek
EL11030028	DrivingLicenceDocument						In All Contexts	Ministry of Transport and Communications	Greece	Άδεια Οδήγησης
EL11030029	EULicenceID	Identifier	Identifier. Type		1	1	In All Contexts	Ministry of Transport and Communications	Greece	Αριθμός Άδειας κατά την ΕΕ
EL11030030	IssueDateTime	Date Time	Date Time. Type	-	1	1	In All Contexts	Ministry of Transport and Communications	Greece	Ημερομηνία Έκδοσης
EL11030031	ReceiptDateTime	Date Time	Date Time. Type	-	1	1	In All Contexts	Ministry of Transport and Communications	Greece	Ημερομηνία Παραλαβής
EL11030032	Remarks	Text	Text. Type	SimpleText2000Type	0	unbounded	In All Contexts	Ministry of Transport and Communications	Greece	Συμπληρωματικά Στοιχεία
EL11030033	IssuerPublicOrganization				1	1	In All Contexts	Ministry of Transport and Communications	Greece	Εκδούσα Αρχή
EL11030034	RelatedDriverPerson				1	1	In All Contexts	Ministry of Transport and Communications	Greece	Πολίτης με Άδεια Οδήγησης
EL11030035	ReferenceVehicleCategoryStatus				1	unbounded	In All Contexts	Ministry of Transport and Communications	Greece	Κατηγορία Οχήματος για το οποίο ισχύει η Άδεια Οδήγησης

Figure 2: Document Driving Licence in tabular form (2<sup>nd</sup> part of fields)

- 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.
- Data types have been detailed in qualified data types with particular facets, like length and pattern, whenever possible.
- Namespaces in the XML Schemas are declared as URNs (Unified Resource Names) following the pattern <1<sup>st</sup> Level: Country>:<2<sup>nd</sup> Level: Responsible Public Body>:<3<sup>rd</sup> Level: Project>:<4<sup>th</sup> Level: Resource Type, i.e. Data, CodeList,>:<5<sup>th</sup> Level: Resource Status, i.e. Draft, Standard>:<6<sup>th</sup> Level: Schema Module Type, i.e. Root Schema Title, Reusable Aggregate Business Information Entity, Unqualified Data Type, Qualified Data Type, CodeList Title>:<7<sup>th</sup> Level: Major Schema Version>.
- 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. The default values for the context Region is Greece, whereas for the contexts Business Process and Organization is In All Contexts.
- 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 NIF is pending
  - Level 1 – Recommended for reuse when an XML Schema has been incorporated into the NIF 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
- In accordance with UBL (Universal Business Language) (OASIS, 2004), 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.
- Three levels of compliance have been recognized:
  - Customization Fully Compliant with the NIF, when all changes in the XML Schemas apply the NIF guidelines by the book
  - Customization Partly Compliant with the NIF, 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 NIF 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 Core Components from scratch.

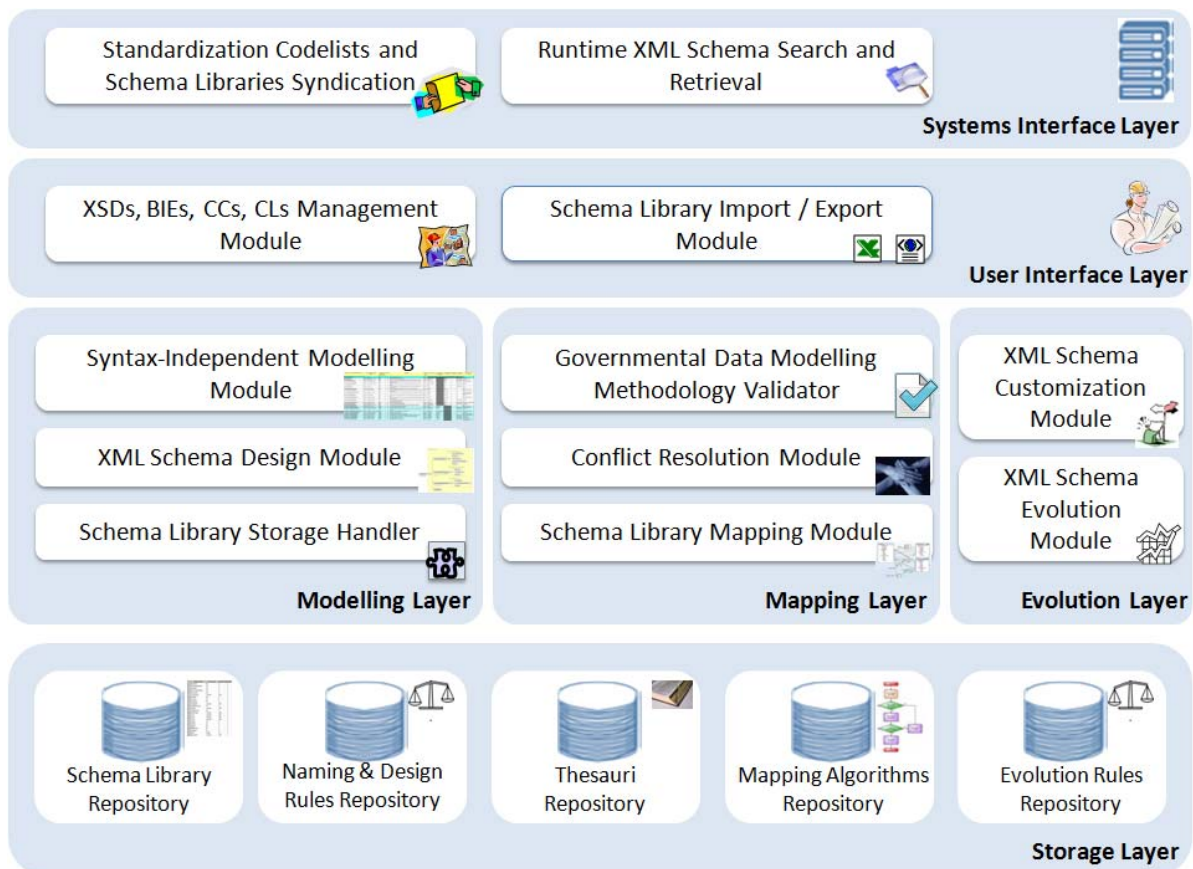
### Governmental Schema Registry Architecture

In the case of the Greek NIF, data modelling activities have been effectively supported with the help of a semantically-rich and context-aware Service and Documents Repository (the Interoperability Registry) (Sourouni et al., 2008) and the CCTS-Excel-To-XSD Tool that is developed to accept the metadata in tabular form as input and provide the XML Schemas as output in compliance with the Greek NIF Naming and Design rules (Charalabidis et al., 2008a).

However, as the Greek NIF Schema Library is populated with more and more XML Schemas, the need for designing a Governmental Schema Registry (GSR) that will incorporate advanced modelling, validating and customizing capabilities comes into the foreground. A prospective server-based architecture for a Registry that shall articulate semantic interoperability issues in the public sector comprises the following layers:

- *Storage Layer*, which stores the XML Schema Library in a relational database since research result have proved that XML native databases are not mature enough to handle component-based XML Schemas (Kappel et al., 2004; Pardede et al., 2008); the Naming and Design Rules that emerge from the NIF Data Modelling Methodology; the Thesauri, such as custom dictionaries and the lexical English database WordNet (Cognitive Science Laboratory, 2009) which provide a list of synonyms, homonyms, opposites and can be exploited when defining Definition, Object Class Term Qualifier, Property term Qualifier, Associated Object Class Term Qualifier, Context and Related Terms in Figure 1; the Mapping Algorithms utilized in order to create possible mappings between an approved and a candidate XML Schema or Core Component; and the Evolution Rules that provide guidance on how to transition between an older version of an XML Schema to the current one.

- *Modelling Layer* which provides the core functionality for modelling a governmental data exchange in a syntax-independent, tabular format and automatically designing its XML Schema in compliance with the Naming and Design Rules (Charalabidis et al., 2008a; Janner et al., 2008; Lampathaki et al., 2008).
- *Mapping Layer* that ensures that when a new component (Core Component, Business Information Entity, Document Schema, Qualified Data Type or Code List) is inserted or updated, it is valid and compliant with the NIF Data Modelling Methodology in terms that it doesn't conflict the already stored components or violates any naming and design rules. In case a conflict is detected, it needs to activate appropriate manual or automatic mechanisms for its resolution. In order to achieve the XML Schema and Components Mapping, a set of methodologies that indicatively include linguistic similarity, structural similarity, overlap in data types and pattern recognition based on neural networks and heuristics have been proposed in the bibliography (Do and Rahm, 2007; Jeong et al., 2008; Shin and Lee, 2006) without though presenting an effective automatic mapping that outperforms the others to date or having been tested in large sets of modular, complex XML schemas.
- *Evolution Layer* taking into account the customization needs that may emerge from organization to organization in order to adhere to new requirements as time goes by (Guerrini, et al., 2007). On the one hand, it ensures that customization of the existing NIF XML Schema Library goes hand in hand with the NIF Customization Rules and on the other hand, it can expand to include appropriate algorithms that make the older XML Schemas conform to the evolved XML Schemas, based on formally defined rules that guide the creation of appropriate XSLT scripts for the real XML documents exchanged.



**Figure 3: Architecture for a Governmental Schema Registry**

- *User Interface Layer* for the improved modellers' experience in managing, importing and exporting the governmental schemas and components.
- *Systems Interface Layer* for XML Schema search and retrieval at runtime by other systems (i.e. other Registries, Portals or Back-office Systems) and for syndicating Code Lists and Schema Libraries which are maintained by other standardization bodies.



## 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. Utilizing practices and standardization from the eBusiness domain, the presented approach is claiming novelty in conceptualisation and overall coherence in the e-Government 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 e-Government Interoperability Framework, the approach presents a sound collection of reusable principles for other Governments and practitioners of the field, as following:

- An overall approach for resolving semantic interoperability issues compatible with 2<sup>nd</sup> generation NIF's, 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 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 NIF Website.
- A set of naming and design rules and principles for the unified creation, maintenance and customization of XML Schemas.
- More than 15 adopted or developed standard codelists for the most common values appearing in public forms and documents.

Problems faced during implementation and application were not trivial and have to be taken in mind during relevant attempts by government officials and practitioners: Gathering the necessary information in order to design complete XML Schemas is time-consuming and needs to be done with caution in order to ensure reusability, while adequate 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. Language issues are also extremely important on the metadata modelling phase, 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), leading to the decision to adopt a glossary from the very beginning.

Future steps in the direction of achieving e-Government semantic interoperability 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.
- 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”, i.e. document standards for Health (such as HL7) or Banking (i.e. UNIFI).
- Gradual implementation of a prototype Governmental Schema Registry initially based on the Storage, Modelling and User Interface Layers with the perspective to eventually expand to the Mapping and Evolution Layers of the proposed architecture.
- Extension and scalability of the Greek NIF Schema Library in order to embrace cross-country transactions following on the methodology proposed by Janner et al. (2008) and Lampathaki et al. (2007).

At this stage, the XML Schemas and the various Core Components developed begin to be 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.

## REFERENCES

1. Australian Government Information Management Office (AGIMO) (2008) Australian Government Interoperability Framework, <http://www.finance.gov.au/agimo/>
2. Bednar, P., Furdik, K., Kleimann, M., Klischewski, R., Skokan, M. and Ukena, S. (2008) Semantic Integration of e-Government Services in Schleswig-Holstein, in Wimmer, M.A., Scholl, H.J. and Ferro, E. (Eds.), *Proceedings of 7th EGOV Conference 2008*, LNCS 5184, 315-327.

3. Cabinet Office (2007) UK e-Government Interoperability Framework, <http://www.govtalk.gov.uk>
4. Cardoso, J. and Sheth, A. (2006) The Semantic Web and its Applications, in Cardoso, J. and Sheth, A. (eds.): *Semantic Web and Beyond: Computing for Human Experience, Semantic Web Services, Processes and Applications*, Springer.
5. Charalabidis, Y., Lampathaki, F. and Askounis, D. (2008a) Unified Data Modeling and Document Standardization Using Core Components Technical Specification for Electronic Government Applications, *Journal of Theoretical and Applied Electronic Commerce Research*, 3, 3, 38-51.
6. Charalabidis, Y., Gionis, G., Moritz Hermann, K. and Martinez, C. (2008b) Enterprise Interoperability Research Roadmap, Draft Version 5.0, [ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/enet/ei-roadmap-5-0-draft\\_en.pdf](ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/enet/ei-roadmap-5-0-draft_en.pdf)
7. Cognitive Science Laboratory, Princeton University (2009) WordNet: A lexical database for the English language, <http://wordnet.princeton.edu/>
8. Do, H.-H. and Rahm, E. (2007) Matching large schemas: Approaches and evaluation, *Information Systems*, 32, 857–885.
9. Greek Ministry of Interior (2008) Greek e-Government Interoperability Framework, <http://www.e-gif.gov.gr>.
10. Guerrini, G., Mesiti, M. and Sorrenti, M. (2007) XML Schema Evolution: Incremental Validation and Efficient Document Adaptation, in D. Barbosa et al. (Eds.): *XSym 2007*, LNCS 4704, 92–106.
11. Hall, W., De Roure, D. and Shadbolt, N. (2009) The evolution of the Web and implications for eResearch, *Philosophical Transactions of the Royal Society A*, 367, 991–1001.
12. Hovy E. (2008) Data and Knowledge Integration for e-Government, in Chen H. et Al. (Eds.), *Digital Government: e-Government Research, Case Studies and Implementation*, Springer, 219-232.
13. HKSARG (Hong Kong Special Administrative Region). (2007, July) Interoperability Framework, Version 5.1, <http://www.ogcio.gov.hk/eng/infra/download/s18.pdf>.
14. IDABC (2004) European Interoperability Framework for pan-European e-Government Services, Version 1.0, <http://europa.eu.int/idabc/en/document/3761>
15. IDABC (2005) Content Interoperability Strategy, Working Paper, <http://ec.europa.eu/idabc/servlets/Doc?id=24405>
16. IDABC (2008) European Interoperability Framework draft version 2.0, <http://ec.europa.eu/idabc/servlets/Doc?id=31508>
17. IDABC (2008) Semantic Interoperability Centre, <http://www.semic.eu>
18. Janner, T., Lampathaki, F., Hoyer, V., Mouzakitis, S., Charalabidis, Y. and Schroth, C. (2008) A Core Component-based Modelling Approach for Achieving e-Business Semantics Interoperability, *Journal of Theoretical and Applied Electronic Commerce Research*, 3, 3, 1-16.
19. Jeong, B., Lee, D., Cho, H. and Lee, J. (2008) A novel method for measuring semantic similarity for XML schema matching, *Expert Systems with Applications*, 34, 1651–1658.
20. Kappel, G., Kapsammer, E. and Retschitzegger, W. (2004) Integrating XML and Relational Database Systems, *World Wide Web: Internet and Web Information Systems*, 7, 343–384.
21. KBSt unit at the Federal Ministry of the Interior (2006) SAGA Standards and Architectures for e-Government Applications, [http://www.cio.bund.de/cln\\_102/DE/Standards/SAGA/saga\\_node.html](http://www.cio.bund.de/cln_102/DE/Standards/SAGA/saga_node.html)
22. KIU (2006) Danish Interoperability Framework, <http://standarder.oio.dk/English/Guidelines>
23. Lampathaki, F., Mouzakitis, S., Janner, T., Schroth, C., Askounis, D. and Hoyer, V. (2008) Achieving Cross-Country Electronic Documents Interoperability with the help of a CCTS-based Modelling Framework, in Charalabidis, Y., Panetto, H., Loukis, E. and Mertins, K. (eds.), *eJETA Special Issue on "Interoperability for Enterprises and Administrations Worldwide"*.
24. OASIS (2004) Guidelines for the Customization of UBL v1.0 Schemas, <http://docs.oasis-open.org/ubl/cd-UBL-1.0/doc/cm/wd-ubl-cm-sc-cmguidelines-1.0.html>.
25. Pardede, E., Rahayu J.-W. and Taniar, W. (2008) XML data update management in XML-enabled database, *Journal of Computer and System Sciences*, 74, 170–195.
26. Ralyte, J., Jeusfeldb, M., Backlundc, P., Kuhn, H. and Arni-Blocha, N. (2008) A knowledge-based approach to manage information systems interoperability, *Information Systems*, 33, 754–784.
27. Schroth, C. (2007) The Internet of Services: Global Industrialization of Information Intensive Services, *Proceedings of IEEE International Conference on Digital Information Management (ICDIM'07), Web X.0 and Web Mining Workshop*, Lyon, France.

28. Selvage, M., Wolfson, D., Zurek, B. and Kahan, E. (2006) Achieve semantic interoperability in a SOA: Patterns and best practices, IBM, Available: <http://www.ibm.com/developerworks/webservices/library/ws-soa-seminterop.html>.
29. Shin, D.-H. and Lee, K.-H. (2006) Towards the faster transformation of XML documents, *Journal of Information Science*, 32, 261-276.
30. Simson, G. and Witt, G.C. (2005) *Data Modelling Essentials*, Third Edition, Morgan Kaufmann Publications, Elsevier.
31. Sourouni, A.-M., Lampathaki, F., Mouzakitis, S., Charalabidis, Y. and Askounis, D. (2008) Paving the way to e-Government Transformation Interoperability Registry Infrastructure Development, in Wimmer, M.A., Scholl, H.J. and Ferro, E. (Eds.), *Proceedings of 7th EGOV Conference 2008*, LNCS 5184, 340-351.
32. UN/CEFACT (2003) Core Components Technical Specification, Part 8 of the ebXML Framework, Version 2.01, [http://www.unece.org/cefact/ebxml/CCTS\\_V2-01\\_Final.pdf](http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf).
33. UN/CEFACT (2006a) Core Component Library (UN/CCL), version 1.0, [http://www.unece.org/cefact/codesfortrade/codes\\_index.htm](http://www.unece.org/cefact/codesfortrade/codes_index.htm).
34. UN/CEFACT (2006b) XML Naming and Design Rules, Version 2.0, <http://www.unece.org/cefact/xml/XML-Naming-and-Design-Rules-V2.0.pdf>.
35. U.S. Department of Justice and Department of Homeland Security (2009) National Information Exchange Model (NIEM), <http://www.niem.gov/>