Controlled Vocabularies and Metadata Sets for Public Sector Information Management

Yannis Charalabidis
University of the Aegean, Greece
yannisx@aegean.gr

Abstract

Public Sector Information management frameworks, usually in the form of ontologies and taxonomies containing controlled vocabularies and relevant metadata sets, appear as a key enabler that assists the classification and sharing of resources related to the provision of open data and efficient digital services towards citizen and enterprises. As different authorities typically use different terms to describe their resources and publish them in various data and service registries that may enhance the access to and delivery of governmental knowledge, but also need to communicate seamlessly at a national and pan-European level, the need for unifying and inclusive digital public service metadata standards emerges. This paper presents the creation of an ontology-based extended metadata set that embraces public sectors services, documents, XML Schemas, codelists, public bodies and information systems. Furthermore, the paper presents experiences of application within the Greek Public Sector, as part of the National Interoperability Framework specification. Such a metadata framework is an attempt to formalize the automated exchange of information between various portals and registries and further assist the service transformation and simplification efforts, while it can be further taken into consideration when applying Web 2.0 techniques in governance.

1 Introduction

Over the years, several public and ICT initiatives across Europe have tried to describe services and other resources for use by systems and applications to serve citizens, business and administration agencies [5]. However, proposing a set of structural and syntactic metadata for digital public services and relevant resources is not adequate and effective to help services discovery and knowledge sharing [8], [13] leading to the conclusion that web-based resources and their mutual relationships can still be considered rather ungoverned [18].

The evolution of Internet, at the same time gives way towards building a Semantic Web that enables seamless communication between machines [3]. In this context, creating and populating rich semantic metadata on the Web has been commonly accepted as the route leading to the Government Semantic Web vision [8]. Metadata is a fundamental concept in building governmental digital collections or public information centres that describe and categorize e-government resources online [23]. According to [4], metadata can be defined as "data about data" and can exist in multiple levels for services and relevant resources.

In this direction, the present paper proposes a metadata set for describing e-Government resources gaining experience from relevant e-Government Metadata standardization efforts. Effectively applied in the context of the Greek e-Government Interoperability Framework [11], [15] and the Interoperability Registry Prototype implementation [21], the proposed metadata set is customized to the particular needs of the e-Government services, documents, XML Schemas, code lists, public bodies and information systems and formalizes their meaning. It further contributes to accelerate the exchange and retrieval of service-related information by governmental sites on the fly and to enhance the perspective over service provision guiding any transformation effort [8].

The structure of the present paper is as following: Section 2 describes the current state of the art in e-Government metadata schemas and standards, analyzing the main elements contained in most implementations worldwide. Section 3 presents an overview of the ontology that synthesizes the proposed metadata set creation, while the actual metadata sets for services, documents, XML schemas, code lists, public bodies and information systems are outlined in Section 4. Conclusions upon the merits and limitations of the approach, as well as next challenges to be tackled are provided in Section 5.

2 Relevant Work

Standardizing metadata sets for describing web resources has attracted great interest both from research and practical reality, as indicated in the following initiatives:

- *Dublin Core Metadata Initiative (DCMI)* [9] provides simple standards to facilitate the finding, sharing and management of information that extends over a broad range of purposes and business models.
- *United Kingdom's e-Government Metadata Standard (UK eGMS)* [16] lays down the elements, refinements and encoding schemes to be used by government officers when creating metadata for their information resources or designing search interfaces for information systems.
- *Australian Government Locator Service (AGLS)* Metadata Element Set [1] provides a set of metadata elements designed to improve the visibility, accessibility and interoperability of online information, organizations and services.
- *New Zealand Government Locator Service (NZGLS)* Metadata Element Set [22] originally designed for use by any governmental agency wishing to make information sources or services more readily discoverable is suitable for more general use.
- *Singapore Government Metadata Standard (SGMS)* [19] aims to help agencies achieve consistency when adhering to e-Government policies.
- *Canada Metadata Standards* [9] officially adopts Dublin Core as the core metadata standard for Web resource discovery since 2001.
- *IDABC Management Information Resources for e-Government (MIREG)* [12] came to supplement MOREQ (Model Requirements for the Management of Electronic Records) results and aimed to develop extensions to the Dublin Core for government information based primarily on the national metadata recommendations of the Member States' public administrations.

- *CEN/ISSS Workshop on Discovery of and Access to e-Government Resources (CEN/ISSS WS/eGov-Share)* [5] presents the ontology for the description of e-Government resources (Services, Process descriptions, Standards and interoperability frameworks, (Requirements) documents) and the metadata schema that is used in its work.
- *ISA* Asset Description Metadata Schema (ADMS) [25] introduces several controlled vocabularies for interoperability assets, but mostly covering very basic terms and artefacts (e.g. interoperability levels, file formats, languages, themes, etc).
-

However, such metadata standards and schemes for network resources apply mainly to documents, electronic archives and public sites [1] or do not cover all the requirements for service-related modeling.

Research papers that have provided sets of metadata and ontologies for modeling services, such as [3], [6], [13], [17], [18], [24], as well as relevant initiatives for describing spatial information [13] and standards [14] have also been taken into account. However, despite the fact that a set of international standards and protocols, such as RDF (Resource Description Framework), OWL (Web Ontology Language), Atom Syndication Format, RSS (Really Simple Syndication), SKOS (Simple Knowledge Organization System) and XTM (XML Topic Maps), accompanies such metadata initiatives in order to formally depict e-Government metadata, a complete solution requires such a wide range of different technologies that to date have not rallied around a standard metadata representation [19].

The main emerging conclusions from studying the underlying state of the art thus include:
- Lack of a comprehensive, yet easy to use standardized metadata schema for e-Government resources, that adopts a service-oriented approach and captures the semantics of all the service-surrounding information, such as XML Schemas and code lists.
- Lack of orientation towards transforming services and real time service provision at web front-ends.
- Lack of easily accessible glossaries and predefined code lists for use in such metadata definitions, that resolve language issues as all the relevant metadata descriptions need to be in local language (for the government officials to understand, modify and approve) and at least in English (for ease of communication with other governments and practitioners).

3   The proposed e-Government Metadata Framework

The definition of the proposed extended e-Government Metadata Standard is driven by the e-Government ontology and emphasizes on the formalization and the representation of the following basic entities – classes:
- *Services* provided in conventional or electronic means by the public authorities to the citizens and businesses.
- *Documents*, in electronic or printed format, that constitute the inputs or outputs of a service or are involved during their execution.
- *Information Systems*, which support the service provision and encompass the web portals as well as the back-office and the legacy systems.
- *Public Bodies* embracing all the service points and the authorities of the public sector that provide services, issue documents, create XML Schemas and code lists and own supporting information systems.
- *Web Services* for the interconnection and the interoperability among information systems during a service execution.
- *Legal Framework* that regulates the service provision, the documents issuance and the overall operation of the public bodies.
- *XML Schemas* and *Code Lists* with which the electronically exchanged documents comply and which are exploited in web services.

Figure 1 presents an abstract overview of the e-Government Ontology which is described in detail in [6] and [21]. The basic clusters of attributes are provided within each class, as well as the main relationships between them giving way to further analysis in Section 4. It needs to be noted that as far as the class Web Service is concerned, the proposed approach adopts the metadata definition prescribed in the OWL-S standard.

Additional classes of the ontology, completing the representation but not presented in further details in the present paper, are the following:

- Classes representing service types, document types, information system types, the (functionally oriented) service category list, and relevant categorization elements.
- Classes representing activity steps (start, finish, decisions, etc), giving the ability for in-depth description of the service flows.
- Classes for representing user-oriented elements, such as life events and business episodes.
- Classes holding information on various characteristics of services and documents, such as authentication methods, ways of service provision, levels of service sophistication, etc.

The majority of the above additional classes constitute an important addition to existing ontologies, such as the eGMS or the CEN/ISSS, providing for automated reconciliation and semantic matching of relevant annotations among systems of different organizations, directly contributing to semantic interoperability achievement. They have been modeled as Controlled Lists in the metadata sets that follow in the next section.
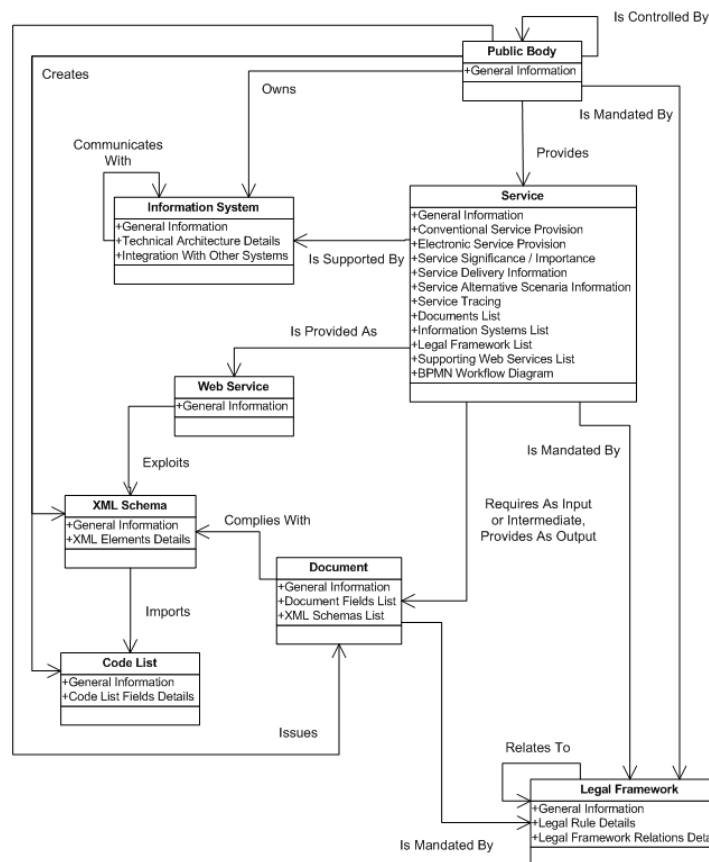

Fig. 1. The main entities of the framework

4 The Detailed Service and Resources Metadata

4.1 Services

The metadata set customized to the conventional and / or electronic services' requirements consists of 9 groups and is provided in Table 1. It needs to be noted that the metadata definition of a service needs to be accompanied by the BPMN (Business Process Modelling Notation) diagram of its workflow.

Table 1. Metadata Set for Services[1]

General Information		
Identifier	Title	Responsible Public Body
Final Service (*)	Addressee (*)	Type (*)
Aggregation (3 level-GCL) (*)	Life Event (*)	Business Episode (*)
Service In Abstract Level (*)	Parent Service	Service Delivery Channels (*)
Conventional Service Provision		
Demand On Physical Presence In Submission (*)	Demand On Physical Presence In Receipt (*)	Conventional Authentication Method (*)
Electronic Service Provision		
Website	Electronic Service Delivery Method (*)	Current Online Sophistication Level (*)
Target Online Sophistication Level (*)	Multilingual Content (*)	Offline Operation (*)
Progress Monitoring Support (*)	Personal Data Level (*)	Trust Level (*)
Required Authentication Level (*)	Current Authentication Mechanism (*)	Registration Process (*)
Service Significance / Importance		
Transactions Volume (per year)	Frequency of Service Requests	Based On European Policies (*)
Service Delivery Information		
Delivery Cost	Delivery Time	Responsible Department
Responsible Public Servant		Service Preconditions
Related Announcements		Related Attachments
Service Alternative Scenaria Information[2]		
Identifier	Title	Conditions
Resources for Public Administration	Resources for Addressee	Total Resources
Cost for Public Administration	Cost for Addressee	Total Cost
Time for Public Administration	Time for Addressee	Total Time
Service Tracing		
Source	Date. Published	Date. Modified
Date. Valid (From-To)	State (*)	Language (*)
Documents List		
Identifier	Title	Position In Service (*)	Mandatory (*)
Replaces Document		Self-appointed Call (*)	
Information Systems List		
Legal Framework List		
Supporting Web Services List		
BPMN Workflow Diagram		

---

[1] The fields marked with (*) take values from appropriate predefined, controlled lists.

[2] It includes computed fields based on the step-by-step calculation of cost, time and resources, taking into account the possibility of faults.

4.2 Documents

The metadata set that accompanies the documents claims novelty in incorporating the documents fields' definition that guides the XML Schema design and the code lists creation at later stages.

Table 2. Metadata Set for Documents

General Information		
Identifier	Title	Creator
Publisher	Type (*)	Subject
Coverage	Format (*)	Language (*)
Addressee	Audience	Mandate. Authorizing Statute
Source	Date. Published	Date. Modified
Date. Valid (From-To)	State (*)	
Document Fields List		
Name	Description	Aggregation
Filled In By	Mandatory (*)	Complex Type (*)
Multiple Values (*)	Predefined Values from Code List	
Type (*)	Length	
XML Schemas List		

The metadata set built around XML Schemas is further customized according to the type of the XML Schema [7] and has been based on the UN/CEFACT Core Components Technical Specification.

Table 3. Metadata Set for XML Schemas

General Information		
Identifier	Title	Creator
Version	Type (*)	Format (*)
Date. Published	Date. Valid (From-To)	State (*)
XML Elements Details		
Unique Identifier	Name	Dictionary Entry Name
Type (*)	Version	Definition
Object Class Term Qualifier	Object Class Term	Property Term Qualifier
Property Term	Associated Object Class Term Qualifier	Representation Term (*)
Date Type Qualifier	Primitive Type (*)	Qualified Data Type
Cardinality Min	Cardinality Max	Facets (i.e. Pattern, Length, Min/Max Length, Enumeration, Total Digits, Fraction Digits, Min/Max Inclusive/Exclusive)
Context: Business Process	Context: Organization	Context: Region (*)
Business Term	Example	Remarks

4.3 Controlled Lists

In order to avoid populating the metadata set with unstructured information, a set of controlled lists has been created and an indicative extract is provided in Table 4.

Table 4. Indicative Controlled Lists – Service Categorisation

Government Category List (1st Level out of 3)[3]		
City Planning and Land Registry	Civilization and Free Time	Education and Research
Environment and Natural Resources	Finance and Economy	Health and Social Care
Information and Communication	International Affairs and European Union	Justice, State and Public Administration
People, Communities and Way of Living	Public Order and Defence	Services for Companies
Transportation Means, Trips and Tourism		Work, Insurance and Pension
Service Type		
Request	Information Discovery	Declaration
Return	Registration	Participative Actions
Permit / Licence	Certificate Issuance	Payment
Life Event[4]		
Birth	Marriage	Divorce
Death in Family	Sickness	Employment
Dismissal	Retirement	Property Acquisition
Property Loss	Residence Change	Education
Travel Abroad	Military Service	Other …
Not Applicable		Unknown

5   Conclusions

The proposed service description framework brings the power of annotating services with commonly agreed metadata into the exchange and the retrieval of service-related information stored in Interoperability Registries by governmental sites on the fly. As the need for transforming services to obtain a more citizen-centric orientation based on their real needs and life events is more and more stressed, the proposed approach has already included metadata around service delivery scenaria that can guide any business process re-engineering effort in the public sector.

Problems faced during the adoption and application of the proposed metadata set were not trivial and have to be to be taken in mind during relevant attempts by government officials and practitioners. The adoption of a common "governance policy" over metadata appears as a worthwhile track towards this direction. Language issues also need to be taken care of early in the process, as the provision of pan-European e-Government Services is already on the way. Finally, adequate time and effort should be spent for educating and working together with government officials at various levels, for obtaining a common perspective over the metadata set.

As the proposed metadata set is incorporated into the Greek e-Government Interoperability Framework [11] and the Interoperability Registry Prototype implementation [21], future steps along our work mainly include exploration of how such a metadata set as the proposed one can: (a) be exploited in intelligent governmental service front-ends that enhance end users experience and have recently started to gain momentum at the international research scene [4], mainly when it comes to provided public services cataloguing and user groups profiling information, and (b) be further elicited in order to take into account service addressees' feedback when creating the service alternative scenaria.

---

[3] Indicative list: more than 300 service categories included in the 3-level Government Category List.

[4] Indicative list: more than 60 life events identified in total.

References

1. Alasem, A.: An Overview of e-Government Metadata Standards and Initiatives based on Dublin Core. Electronic Journal of e-Government 7 (1), 1-10 (2009)
2. Australian Government, National Archives of Australia, AGLS Metadata Standard, Version 1.3, http://www.naa.gov.au/records-management/publications/AGLS-Element.aspx
3. Barone A., Di Pietro P.: Semantic of e-Government Processes: a Formal Approach to Service Definition (Arianna), in Proceedings of eGovINTEROP 2006, Bordeaux (France)
4. Cantera, J. M., Reyes, M., Hierro, J., Tsouroulas, N.: White Paper on Service Front-Ends in the Future Internet of Services and Next Generation SOA, http://www.nexof-ra.eu/sites/default/files/nxt_gen_soa_workingdraft_21july2008.pdf
5. CEN/ISSS Workshop on Discovery of and Access to e-Government Resources (CEN/ISSS WS/eGov-Share): Sharing e-Government resources: a practical approach for designers and developers, http://www.cen.eu/cenorm/businessdomains/businessdomains/isss/workshops/wsegovshare.asp
6. Charalabidis, Y., Askounis, D.: Interoperability Registries in e-Government: Developing a Semantically Rich Repository for Electronic Services and Documents of the new Public Administration, in Proceedings of the 41st Hawaiian International Conference of System Sciences, HICCS-08, Hawaii (2008)
7. Charalabidis, Y., Lampathaki, F. and Askounis, D.: 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 (2008)
8. Charalabidis, Y., Lampathaki, F., Psarras, J.: Combination of Interoperability Registries with Process and Data Management Tools for Governmental Services Transformation, in Proceedings of the 42nd Hawaiian International Conference of System Sciences, HICCS-09, Hawaii (2009)
9. Devey, M., Cote, M.-C., Bain, L., McAvoy, L.: Celebrating 10 Years of Government of Canada Metadata Standards. In Proceedings of International Conference on Dublin Core and Metadata Applications (2010)
10. Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/dces/
11. Greek Ministry of Interior, Greek e-Government Interoperability Framework, http://www.e-gif.gov.gr/
12. IDABC Management Information Resources for e-Government, http://ec.europa.eu/idabc/en/document/3615/5585
13. INSPIRE (Infrastructure for Spatial Information in Europe) Draft Implementing Rules for Metadata (Version 2): http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/draftINSPIREMetadataIRv2_20070202.pdf
14. ISO/IEC WD 24706 Information technology – Metadata for technical standards and specifications documents: http://jtc1sc32.org/doc/N1251-1300/32N1257-WD24706.pdf
15. Lampathaki, F., Charalabidis, Y., Sarantis, D., Koussouris, S., Askounis, D.: E-Government Services Composition Using Multi-faceted Metadata Classification Structures, In: Wimmer, M.A., Scholl, H.J. and Grönlund, A. (Eds.), Proceedings of 6th EGOV Conference 2007, LNCS vol. 4656, pp. 116-126 (2007)
16. Office of the e-Envoy, e-Government Metadata Standard, Version 3.1, http://www.govtalk.gov.uk/documents/eGMS%20version%203_1.pdf
17. Palmonari, M., Viscusi, G., Batini, C.: A semantic repository approach to improve the government to business relationship. Data & Knowledge Engineering. 65, 485–511 (2008)
18. Peristeras, V., Tarabanis, K.: Governance Enterprise Architecture (GEA): Domain Models for E-Governance, in Proceedings of the ICEC'04, Sixth International Conference on Electronic Commerce (2004)
19. Singapore Government Metadata Standard (SGMS), http://www.igov.gov.sg/NR/rdonlyres/B90E7579-E494-477A-86B7-E532F3FBC1B4/18178/2010ReportonSingaporeeGovernment.pdf
20. Smith, J.R.: The Search for Interoperability. Multimedia, IEEE. 15 (3), 84-87 (2008)
21. Sourouni, A.-M., Lampathaki, F., Mouzakitis, S., Charalabidis, Y., Askounis, D.: 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, vol. 5184, pp. 340-351 (2008)
22. State Services Commission, NZGLS Metadata Element Set, Version 2.1, http://www.e.govt.nz/standards/nzgls/standard/element-set-21/nzgls-element-set-2-1.pdf
23. Tambouris, E., Manouselis, N., Costopoulou, C.: Metadata for digital collections of e-government resources. The Electronic Library. 25 (2), 176-192 (2007)
24. Vassilakis C., Lepouras G.: An Ontology for e-Government Public Services. Encyclopedia of E-Commerce, E-Government and Mobile Commerce. 865-870 (2006)
25. ISA Asset Description Metadata Schema (ADMS) v1.0, European Commission, 2012, https://joinup.ec.europa.eu/catalogue/distribution/adms_specification_v100zip