

On Using the Core Public Sector Vocabulary (CPSV) to Publish a “Citizen’s Guide” as Linked Data

Alexandros Gerontas
University of Macedonia
Thessaloniki, Greece
agerontas@uom.gr

Efthimios Tambouris
University of Macedonia and
ITI-CERTH
Thessaloniki, Greece
tambouris@uom.gr

Konstantinos Tarabanis
University of Macedonia and
ITI-CERTH
Thessaloniki, Greece
kat@uom.gr

ABSTRACT

A large number of public authorities worldwide publish descriptions of their administrative procedures (public services) in public service catalogues. These public services descriptions are often based on a national or an ad hoc description template (also referred to as model). Moreover, usually they are not provided in a machine-readable format. At the end of 2016, the European Commission released CPSV-AP 2.0 (Core Public Sector Vocabulary-Application Profile) which is a Core Vocabulary for public service modeling and publishing, incorporating Linked Data technologies. In this paper, we present an approach for using CPSV-AP 2.0 to model and publish, as Linked Data, public services descriptions of a regional public service catalogue. Consequently, we discuss the challenges that we encountered during the modeling and publishing procedure.

CCS CONCEPTS

• Information systems → Information systems applications;

KEYWORDS

Public service, administrative procedures, public service modeling, public service catalogues, core vocabularies, CPSV, CPSV-AP 2.0, Linked Data

ACM Reference Format:

Alexandros Gerontas, Efthimios Tambouris, and Konstantinos Tarabanis. 2018. On Using the Core Public Sector Vocabulary (CPSV) to Publish a “Citizen’s Guide” as Linked Data. In *Proceedings of 19th Annual International Conference on Digital Government Research (dg.o’18)*, Anneke Zuiderwijk and Charles C. Hinnant (Eds.). ACM, New York, NY, USA, Article 4, 6 pages.

1 INTRODUCTION

An important field of eGovernment research and practice is the online provision of Public Services (PS). The European Union (EU) has defined a set of twenty public services that are considered as particularly important and measures regularly

the maturity of their online provision across the Member States [4]. It has been estimated that their online provision across the EU could contribute to high annual savings [5].

Public service (PS) modeling is a cornerstone of PS provision. Usually, PS modeling comprises the first step and the basis for the development of online Public Service Catalogues and/or eGovernment Information Systems.

In practice, the development of eGovernment Information Systems is usually based on ad hoc PS models or models included in national interoperability frameworks (e.g. the Greek eGIF [21]). In academic literature, a number of public service models have also been proposed, e.g. the Governmental Markup Language (GovML) [32], service model of the Governance Enterprise Architecture (GEA) [20], etc. Additionally, standardization bodies such as CEN and W3C promote the definition of commonly agreed standards for public service modeling towards facilitating eGovernment interoperability [12]. Furthermore, W3C has established the W3C EGOV Interest Group that is interested in advancing eGovernment through W3C technologies [13]. The development and use of a common public service model is among the issues of interest of this group [19].

Despite the significance of online PS provision and the plethora of public service models, both applied and theoretical, until recently a universally-accepted standard (model) did not exist. The absence of such standard results in wasting of recourses and hindering of interoperability [16, 28]. Additionally, it has been suggested that the introduction of unified public service models will bring a number of benefits, including reduced development costs, improved software quality, improved users experience and interoperability across different eGovernment Information Systems [31]. Furthermore, it has been articulated that sharing of common PS models would improve the analysis and development of eGovernment Information Systems, increase the efficiency of online PS provision, facilitate the adoption of public services by citizens and promote the reusability of public services [14, 15, 18]. Thus, the development and adoption of a relevant European standard is considered a prerequisite for eGovernment interoperability and cross-border pan-European PS provision, facilitating the vision of a Digital Single Market, which is a major priority of the European Union [6].

Identifying the need for a standard PS model, the European Commission (EC) launched in 2012 the Core Public Service Vocabulary (CPSV) initiative [11], in the framework of ISA (Interoperability Solutions for Public Administrations)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

dg.o’18, May 30-June 1, 2018, Delft, Netherlands

© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6526-0/18/05...\$15.00

<https://doi.org/10.1145/3209281.3209362>

and its successor ISA² (Interoperability solutions for public administrations, businesses and citizens) programmes [29], aiming at developing a simplified, reusable and extensible model that captures the fundamental characteristics of a service offered by public administrations. CPSV was released in 2013 aiming to become the de facto European Standard for public service modeling. Subsequently, at the end of 2016, EC announced the final version of CPSV-AP 2.0 [30], which is an application profile of CPSV. CPSV-AP 2.0 was developed by a working group that included representatives from all European Union (EU) Member States. CPSV-AP 2.0 incorporates Linked Data as underpinning technology.

The “Citizen’s Guide of the Region of Epirus”, which was initiated in July 2011, is an example of a regional PS catalogue. This catalogue has been awarded as “best practice” in the European Public Sector Award 2015 competition, which was organised by the European Institute of Public Administration (EIPA) with the support of the European Commission. Despite this fact, this catalogue is based on an ad hoc PS model.

The objective of this paper is to present the use of CPSV-AP 2.0 in order to migrate the “Citizen’s Guide” to linked data technology. The motivation is based on the assumption that exploiting linked data technology using a European Standard, namely the CPSV-AP 2.0, will provide value to all stakeholders, including citizens, policy makers, public organizations, etc. For example, PS modeling using Linked Data would enable PS visualizations and application of knowledge management and knowledge representation methodologies. Additionally, the use of a European standard would facilitate semantic interoperability across European public organizations for the provision of pan-European public services.

In the rest of this paper, we present the steps towards using CPSV-AP 2.0 to publish the PS descriptions included in “the Citizen’s Guide of the Region of Epirus” as Linked Data. In section 2, we provide some background information, including a short description of CPSV, reflections of the literature on Linked Government Data and a short presentation of CPSV-AP 2.0. Following, in section 3 “the Citizen’s Guide Project” is briefly presented. Subsequently, in section 4, the approach that was adopted for using CPSV is described. In section 5, our effort for using CPSV-AP 2.0 to model and publish the public services of the Region of Epirus is summarized. In section 6, we discuss the main challenges and in section 7 we draw our conclusions and articulate plans for future work.

2 BACKGROUND

2.1 Core Public Service Vocabulary

The Core Public Service Vocabulary (CPSV) is a European Standard designed to model PS descriptions and bring consensus on the basic PS metadata across EU Member States. CPSV has not been designed to model every characteristic or every property of PS across all European Countries as well as all domains and all Public Organizations. In the position paper of European Commission [10] it is suggested that “A Core Concept is a simplified data model that captures the

minimal, global characteristics/attributes of an entity in a generic, country and domain neutral fashion. It can be represented as Core Vocabulary using different formalisms (e.g. XML, RDF, JSON)”. Rather, CPSV is proposed as basis for new PS models and as the linking element between existing PS models for achieving a minimum level of semantic interoperability. Thus, new PS models can be designed based on CPSV while existing PS models can be mapped to CPSV.

2.2 Linked Government Data

Linked data refers to “*data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets*” [2]. Linked Open Data (LOD) technology aspires to realize the Web of Data where data exist as self-descriptive entities. Here, data are not integrated with applications but rather exist autonomously and can be interlinked through typed links. Linked data technology exploits Uniform Resource Identifiers (URI) and the Resource Description Framework (RDF) to describe data and their interrelations with other data [9]. During the previous years, there has been an increasing trend to publish data as linked data to be more easily sharable and reusable [17]. The Web of Data aims at interconnecting isolated data islands (data silos) to a giant distributed dataset, also called the linked open data (LOD) cloud. Thus, LOD technology transforms fragmented datasets, which exist at data silos, to a global data space, that can be visualized as a graph. Public sector linked data hubs could contribute to sharing, reusing, federation and integration of Public Sector Information. Furthermore, LOD technology enables public organizations to publish their data in a structured and modular format, facilitating the systematic data and knowledge management of the produced semantic database. Thus, the use of linked data could potentially be an integration paradigm for public services and organizations offering significant benefits to citizens (individuals and enterprises) as well as public organizations (including civil servants and policy makers).

However, the benefits of publishing public services descriptions as linked data are counterbalanced by the difficulties of the publishing procedure. The literature suggests that although linked data technology and standards are mature and powerful there is still much work to be done on formalizing the process of publishing linked government data by developing straightforward patterns that could be easily adopted by public organizations or institutes [27]. Additionally, it is pointed out that there are no systematic guidelines for publishing Government Linked Data, including technical details about all the steps of the publishing process [1].

2.3 Core Public Service Vocabulary Application Profile 2.0

CPSV-AP 2.0 is the Application Profile, version 2.0, of CPSV. An Application Profile is “*a specification that re-uses terms from one or more base standards, adding more specificity by identifying mandatory, recommended and optional elements*

The benefits for the public administration (Region of Epirus) include:

- Increased decongestion of the authority by the large volume of people waiting to be served. As a result, better utilization of the time of civil servants is achieved and thus their productivity is increased.
- Increased cooperation between the organizational units of public authorities for improving public services.
- Development of a solid basis to further model and reengineer public services.
- The public services descriptions could be used as core material for the developing of relevant transactional applications for these public services.
- Public services are becoming a part of Regions Total Quality Management (TQM) methodology that would potentially be implemented in the future.

In addition, the Citizen's Guide can be used as a platform for promoting the cooperation between all Regions of the country as many of the administrative processes are common to all regions.

Finally, it is worth noting that the implementation team of Citizen's Guide was composed exclusively of civil servants whereas Free/Open Source software was used. As a result, the implementation, maintenance and support costs were kept very low. The implementation period of the Citizen's Guide was about two and a half years, specifically from the summer of 2011 until the beginning of 2014.

4 MOVING INTO CPSV AND LINKED DATA: THE APPROACH FOLLOWED

In the framework of ISA programme and its successor ISA² programme, a procedure for public service modeling based on CPSV-AP 2.0 and the transformation to linked data has been proposed [23]. To support this procedure, a number of software tools has been developed. From this toolbox, in our work, we have used the following tools:

- CPSV-AP 2.0 Template Spreadsheet [25], which is an Excel template for publishing Linked Data according to CPSV model.
- CPSV-AP 2.0 RDF skeleton for mapping the sheets and columns of the Template Spreadsheet to the Classes and Properties of the CPSV-AP 2.0. This also includes the RDF predicates of the CPSV-AP 2.0 model [24].
- CPSV-AP 2.0 Validator [26], which can be used to validate the compliance of produced linked data with the CPSV model.

In addition, we have also used the following tools for transforming PS descriptions from a tabular format to a machine-readable (RDF/Turtle) format:

- OpenRefine (formerly Google Refine) [7] and
- RDF Extension, developed by DERI [8].

The adopted procedure [23] has been initially devised for the modeling of interoperability solutions descriptions, based on another EC ISA model, namely the ADMS-AP. It comprises the following steps:

- Step 1: Fill in the spreadsheet. CPSV-AP 2.0 Template Spreadsheet consists of one sheet per CPSV-AP 2.0 class, which in turn contains a column per property. Cells contain the values of the properties, i.e. the actual information about a Public Service. Two classes can be interlinked by having properties sharing the same URI. To provide a second value for a property, two rows need to be used. All rows with the same URI in the first column will be merged by OpenRefine.
- Step 2: Import the spreadsheet in OpenRefine. In this step, a new project is created in OpenRefine and the spreadsheet from the previous step is imported to this project.
- Step 3: Apply the CPSV-AP 2.0 skeleton. The tabular data are transformed to Linked Data by using OpenRefine along with the corresponding RDF extension. These tools provide a graphical environment for the alignment of classes, properties and their values to sheets, columns and cells of the Template Spreadsheet. CPSV-AP 2.0 Skeleton automates this procedure.
- Step 4: Export RDF. After applying the RDF skeleton, the transformed content of the template spreadsheet is exported to an RDF file.
- Step 5: RDF file validation. The CPSV-AP 2.0 Validator is used to validate the compliance of machine-readable public services descriptions, included in the produced RDF file, with the CPSV-AP 2.0 specification. It should be noted that, to the best of our knowledge, no further instructions are provided for publishing public services descriptions as linked data or validating the produced linked data.

5 PUBLISHING "CITIZEN'S GUIDE" AS LINKED DATA

The publishing procedure started with the selection of a set of public services. The selection was made by the authors in cooperation with domain experts of the Region of Epirus. A basic criterion was PS popularity. As a result, five public services were selected for inclusion in this first pilot phase. For modeling the selected public services according to CPSV-AP 2.0 and consequently the transformation of the tabular data to linked data format we followed the procedure described in the previous section.

More specifically, first, we extracted the selected public services descriptions from the Citizen's Guide and copied it to the CPSV-AP 2.0 Template Spreadsheet (step 1). The mapping between the properties of the Citizen's Guide of the Region of Epirus and the properties of the CPSV-AP 2.0 is depicted in Table 2.

The population of CPSV-AP 2.0 concepts was performed using the CPSV-AP 2.0 Template Spreadsheet (step 1). This is a simple to use Excel file with a clear structure that corresponds to the CPSV-AP 2.0 model. Here, we had to map the concepts of the existing template to the Excel template and to migrate the content. In the process, we faced syntactic

Table 2: Mapping of “the Citizen’s Guide of the Region of Epirus” to the CPSV-AP 2.0

Property of the “the Citizen’s Guide of the Region of Epirus”	Mapping to CPSV-AP 2.0 (Class: Property)
Title of the PS	PublicService:Name
Category	PublicService:Sector
Description	PublicService:Description
Required input documents	PublicService:HasInput
Legal Framework	PublicService:HasFormalFramework
Cost	PublicService:HasCost
Processing time needed for the completion of the execution of PS	PublicService:ProcessingTime
Produced output document(s)	PublicService:Produces
Validity time of the output document(s) (if exists)	-
Renewal procedure (if exists)	-
Competent organizational unit	PublicService:HasContactPoint
Additional Sources of Information	-
Notes	-
Relevant Files	-

and semantic interoperability obstacles. For example, all documents, required as an input to a public service, were located in a single text field in the original PS Catalogue. However, these had to be split manually, as in the Excel template each input document should be inserted in a separate row (this corresponds to the value of the property `PublicService:HasInput` which contains the URI of a document).

Subsequently, we transformed the public service descriptions from a tabular format to a machine-readable format (namely RDF/Turtle) using OpenRefine and the corresponding RDF extension (steps 2-4). In this process, we utilised the RDF skeleton (see Fig.2) to automate the alignment of the tabular data with the CPSV-AP 2.0 schema.

The installation and use of OpenRefine and the corresponding RDF extension was straightforward. Attention should be only paid to ensure that the version of RDF extension is compatible with the version of OpenRefine. Moreover, the RDF skeleton was very helpful, facilitating the transformation of tabular data to linked data, in the form of an RDF file.

Following, we imported the RDF file to the CPSV-AP 2.0 Validator (step 5), which provides a user-friendly interface for validating the produced RDF file for syntax violations or other inconsistencies with the CSV-AP 2.0 schema.

CPSV-AP 2.0 Validator is a user-friendly tool. Here, we encountered some minor technical issues, which were fixed. More specifically, some false warnings were produced by the tool due to a minor syntactic error in the RDF skeleton.

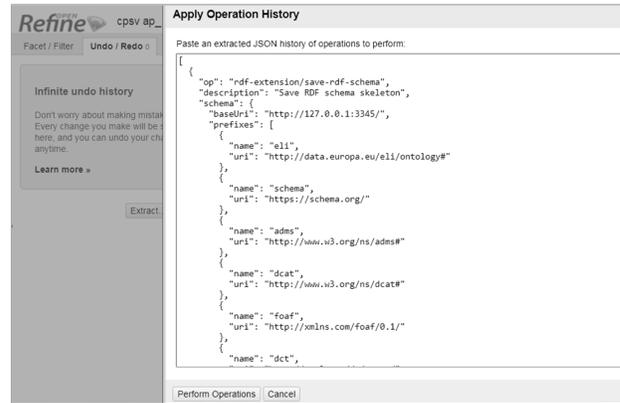


Figure 2: The CPSV-AP 2.0 RDF skeleton

6 DISCUSSION

In the first step of the followed approach, namely the modeling of public services based on CPSV-AP 2.0, we encountered a number of challenges. Here, we discuss the most important.

Firstly, the format of the properties of the “Citizen’s Guide” and the CPSV-AP 2.0 were not identical. For example, in the original catalogue, the titles of all required (input) documents for a public service were included in one text cell while in CPSV-AP 2.0 model each supporting document must be placed (as a URI) in a separate cell.

Furthermore, all properties of the original catalogue were of string type. Thus, there were no URI values in any field of the model. More importantly, there is no URI policy at National or European level for any property (field) that requires a URI as a value. We believe this is a major drawback of the whole process.

Another difficulty is that the categorization of the public services of the Region of Epirus cannot be directly mapped to the proposed categorization of the CPSV-AP 2.0 (which is reported in the `PublicService:Sector` property). In addition, the cardinality of some properties of the models is not the same. Finally, some properties of CPSV-AP 2.0 are in TBC (to be confirmed) status, e.g. the property “type” of the “Evidence” class.

Beyond the above challenges, there are also additional challenges that should be addressed. For example, there are no guidelines on how to publish the RDF file or how to evaluate its potential added-value.

7 CONCLUSIONS FUTURE WORK

The evolution of the semantic web and linked open data technologies is promising for providing benefits in eGovernment and public service provisioning [9, 27]. In this respect, CPSV-AP 2.0 seems like a promising European standard for modeling and publishing public services, which also incorporates Linked Data technologies.

However, there are still challenges for exploiting CPSV-AP 2.0 at a large scale. One of the most important, is the absence

of a concrete and tested procedure, including detailed guidelines, that a public organisation could follow to easily model and publish its administrative procedures (public services) using CPSV-AP 2.0. These guidelines should include information about all needed tools, a URI design policy, detailed examples, etc. These would enable us to evaluate whether the potential benefits are worth the effort needed towards the exploitation of CPSV-AP 2.0 for public service modeling and publishing.

Our future work includes the introduction of a concrete procedure for using CPSV-AP 2.0. Consequently, we will apply it to a major part of “Citizen’s Guide” content and evaluate the potential added value produced by its adoption. The evaluation process will follow a number of scenarios for different stakeholders (e.g. citizens, policy makers, public servants etc).

ACKNOWLEDGMENTS

The work reported in this paper is funded by the European Commission within the H2020 Programme in the context of the project OpenGovIntelligence (<http://OpenGovIntelligence.eu>) under grand agreement No. 693849.

REFERENCES

- [1] Villazn-Terrazas B., L.M. Vilches-Blzquez, O. Corcho, and A. Asuncin Gmez-Prez. 2011. *Methodological Guidelines for Publishing Government Linked Data in Linking Government Data*. Springer, Chapter 2.
- [2] C. Bizer, T. Heath, and T. Berners-Lee. 2009. Linked data: the story so far. *International journal on Semantic Web and Information Systems* 5, 3 (2009), 122.
- [3] J. Bosse, M. Burnett, S.M. Nielsen, C. Rongione, and H. Scholtens. 2015. *European Public Sector Award 2015: The Public Sector as Partner for a Better Society*. Technical Report. European Institute of Public Administration. http://www.epsa2015.eu/files/site/EP2015_Publication_updated.pdf
- [4] Capgemini, IDC, Rand Europe, Sogeti, and DTi. 2010. *Digitizing public services in Europe: putting ambition into action, 9th benchmark measurement*. Technical Report. European Commission, Directorate General for Information Society and Media.
- [5] European Commission. 2014. *A Digital Agenda for Europe*. Technical Report.
- [6] European Commission. 2015. *Why we need a Digital Single Market*. Technical Report.
- [7] OpenRefine Community. 2016. OpenRefine. (2016). <http://openrefine.org/index.html>
- [8] DERI. 2014. RDF Refine. (2014). <http://refine.deri.ie/>
- [9] L. Ding, V. Peristeras, and M. Hausenblas. 2010. Linked open government data. *IEEE Intelligent Systems* 27 (1 2010), 1115.
- [10] Interoperability Solutions for Public Administrations (ISA) programme. 2011. *eGovernment Core Vocabularies: The SEMIC.EU approach*. Technical Report. European Commission.
- [11] Interoperability Solutions for Public Administrations (ISA) programme. 2013. *Core Public Service Vocabulary specification*. Technical Report. European Commission.
- [12] European Committee for Standardization (CEN). 2008. *Final Report of the CEN/ISSS eGovernment Focus Group on the eGovernment Standards Roadmap*. Technical Report.
- [13] T. Janowski and J. Holm. 2012. *Draft Roadmap*. Technical Report. W3C, EGOV Interest Group.
- [14] M. Janssen and R. Wagenaar. 2004. Developing generic shared services for e-government. *Electronic Journal of e-Government* 2, 1 (2004), 31–38.
- [15] A. Kayed, M. Nizar, and M. Alfayoumi. 2010. Ontology concepts for requirements engineering process in e-government applications. In *Fifth International Conference on Internet and Web Applications and Services*. 396.
- [16] N. Loutas, S. Goedertier, C. Colas, M. De Keyzer, D. Di Giacomo, J.R. Frade, M. Dekkers, and V. Peristeras. 2013. Realising a Federation of Repositories of Reusable Metadata EC. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*. 156.
- [17] S. Mouzakitis, D. Papaspyros, M. Petychakis, S. Koussouris, A. Zafeiropoulos, E. Fotopoulou, L. Farid, F. Orlandi, J. Attard, and J. Psarras. 2017. Challenges and opportunities in renovating public sector information by enabling linked data and analytics. *Information Systems Frontiers* 19, 2 (2017), 321336.
- [18] E. Ostasius, Z. Petraviciute, and G. Kulvietis. 2010. Constructing a generic e-service model in public sector. In *16th International Conference on Information and Software Technologies (IT-2010)*. 33.
- [19] V. Peristeras. 2008. *Use Case 9 - Common Service Model*. Technical Report. W3C, EGOV Interest Group. http://www.w3.org/egov/wiki/Use_Case_9_-_Common_Service_Model
- [20] V. Peristeras and K. Tarabanis. 2004. Governance enterprise architecture (GEA): domain models for e-governance. In *ACM International Conference Proceeding Series*. International Conference on Electronic Commerce, ICEC04, Delft, Netherlands.
- [21] Information Society S.A. 2008. *The Greek e-Government Interoperability Framework*. Technical Report. Hellenic Ministry of Interior, General Secretariat of Public Administration and e-Government.
- [22] M. Schmachtenberg, C. Bizer, and H. Paulheim. 2014. Adoption of the linked data best practices in different topical domains. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 245.
- [23] Semantic Interoperability Community (SEMIC). 2012. Generate ADMS asset descriptions from a spreadsheet with Refine RDF. (2012). <https://joinup.ec.europa.eu/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf>
- [24] Semantic Interoperability Community (SEMIC). 2016. CPSV-AP 2.0 - RDF Skeleton. (2016). https://joinup.ec.europa.eu/rdf_entity/http_e_f_data_ceuropa_ceu_fw21_f359c90ab_bee2d_b400b_babdb_bdb299cff7430
- [25] Semantic Interoperability Community (SEMIC). 2016. CPSV-AP 2.0 - Spreadsheet template. (2016). https://joinup.ec.europa.eu/rdf_entity/http_e_f_data_ceuropa_ceu_fw21_f48579d79_bc9c0_b40d5_b8c0f_b8d2f1ae3182b
- [26] Semantic Interoperability Community (SEMIC). 2016. CPSV-AP 2.0 - Validator. (2016). <http://cpsv-ap.semic.eu/cpsv-ap-validator/>
- [27] J. Sheridan and J. Tennison. 2011. Linking UK Government Data. In *Proceedings of the Linked Data on the Web Workshop (LDOW2010)*. 1.
- [28] G. Shukair, N. Loutas, V. Peristeras, and S. Sklar. 2013. Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. *Computers in Industry* 64, 1 (2013), 1018.
- [29] Interoperability solutions for public administrations businesses & citizens (ISA2 Program). 2016. . Technical Report. European Commission.
- [30] Interoperability solutions for public administrations businesses & citizens (ISA2 Program). 2016. *D04.01 Core Public Service Vocabulary Application Profile 2.0*. Technical Report. European Commission.
- [31] E. Tambouris. 2008. Introducing the need for a domain model in public service provision (psp) e-government systems. In *3rd International Conference on Data Information Management*. 794.
- [32] E. Tambouris, G. Kavadias, and E. Spanos. 2005. The Government Markup Language (GovML). *Journal of E-Government* 1, 2 (2005), 5970.