Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

On using CPSV-AP to publish public service descriptions as linked open data

Alexandros Gerontas[1,*], Efthimios Tambouris[2], Konstantina Lazopoulou[3], Konstantinos Tarabanis[4]

Abstract: Public authorities worldwide provide a large number of Public Services (PS) to citizens, businesses and other authorities. In this context, they publish PS catalogues containing descriptions of these services, e.g., about cost, required documents, contact details, etc. Two main challenges in the design of PS catalogues are standardisation and interoperability. To address these challenges, the European Union (EU) has developed the Core Public Service Vocabulary (CPSV), as the proposed EU standard for PS modelling. CPSV-AP is an application profile of CPSV that uses linked data as an underpinning technology to exploit its benefits, e.g., interlinking PS descriptions with each other and with other web of data resources. Despite its potential however, there is only limited research on CPSV-AP use in practice. The aim of this paper is to devise a process for using CPSV-AP and conduct a pilot implementation, using this process, to investigate potential benefits or challenges from CPSV-AP practical usage. In the framework of this pilot implementation, we publish a set of PS descriptions, included in an official PS catalogue, as CPSV-AP compliant linked data. The contribution of this paper is a detailed process for CPSV-AP use in practice and relevant lessons learnt. We anticipate this research will be beneficial to both academics working on PS semantic interoperability and practitioners aiming to migrate existing PS catalogues exploiting CPSV-AP.

Keywords: *Core Public Service Vocabulary, CPSV, Core Public Service Vocabulary Application Profile, CPSV-AP, Public Services, Metadata, Linked Data, Core Vocabularies*

1. Introduction

Public service (PS) provisioning is at the heart of citizen-centric eGovernment. Modelling of PSs is a key activity in the development of eGovernment information systems, PS catalogues, portals and websites. A PS catalogue is usually published in the web at the local, regional, national or even international level. Such catalogues contain useful information about PSs helping citizens and businesses to understand their legal obligations and rights. This is an essential first step for administrative burden reduction. Additionally, publishing structured information about PSs or other government data as open data enhances transparency and trust between public authorities and citizens [1], [2], [3]. Moreover, citizens are empowered, as they become well-informed. Furthermore, they are facilitated to participate in PS co-creation [3] and to exploit the published open data, e.g., by reusing them [4], [5], [6]. Consequently, PS catalogues, including structured PS descriptions, ideally based on a standard PS model, are deemed very important.

For many years, public organizations and IT industry have modelled PSs either based on national interoperability standards, which usually include a PS model or in an *ad-hoc* way. Academics have also proposed various PS models [7], as for example the Governmental Markup Language (GovML) [8], the service execution object model of the Governance Enterprise Architecture (GEA) [9], etc. Nevertheless, until recently no PS model had been universally agreed. The absence of a standard PS model results in wasting of recourses and hindering of interoperability. Additionally, it has been suggested that the introduction of a standard public service model will improve the analysis and development of eGovernment Information Systems and will bring a number of benefits, including reduced development costs, improved software quality, improved users experience and facilitate the adoption of public services by citizens. Furthermore, it has been articulated that sharing of a common PS model, will facilitate the interoperability across different eGovernment Information Systems, increase the efficiency of online PS provision and promote the reusability of PSs [5], [10], [11]. Thus, the development and adoption of a relevant European standard is

[1] **corresponding author: Alexandros Gerontas, University of Macedonia, Thessaloniki, 54636, Greece. E-mail: agerontas@uom.edu.gr, Tel: 00306947563250, ORCID: 0000-0003-1786-9717*

[2] *Professor Efthimios Tambouris, University of Macedonia, Thessaloniki, 54636, Greece. E-mail: tambouris@uom.edu.gr, ORCID: 0000-0001-8036-9788*

[3] *Konstantina Lazopoulou, Hellenic Open University, Patras, Greece*

[4] *Professor Konstantinos Tarabanis, University of Macedonia, Thessaloniki, 54636, Greece. E-mail: kat@uom.edu.gr, ORCID: 0000-0002-4663-2113*

considered a prerequisite for eGovernment interoperability and cross-border pan-European PS provision, facilitating the vision of a Digital Single Market.

The European Commission (EC) identified the need for a standard model to unify PS modelling across EU Member States and potentially beyond the EU. For this purpose, a set of standards (namely Core Vocabularies) have been developed in the framework of the EC Interoperability Solutions for European Public Administrations (ISA) programme (2010 – 2015) and its follow-up programme EC ISA^2 programme (2016 - 2020). In this paper, we focus on Core Public Service Vocabulary (CPSV) [12], which is the Core Vocabulary for PS modelling, and more specifically on Core Public Service Vocabulary Application Profile (CPSV-AP), which is an application profile of CPSV incorporating linked data as an underpinning technology.

We note here that, in this paper, the term *PS model* is used to denote a data model that has been defined for describing and/or providing PSs. A PS model contains concepts such as title, description, required documents (input), cost, legal framework, output, contact details, communication channels, etc [13]. Additionally, the term *PS description* is used to denote an instance of a PS model, comprising the values for a set of concepts of that PS model.

In the rest of this article, Section 2 provides background information related to Services and PSs, Government Linked Data, CPSV and CPSV-AP. Additionally, the research objectives are articulated. The methodology followed is presented in Section 3. In Section 4, a literature review about processes for publishing linked data, focusing on Government Linked Data is conducted. Subsequently, a process for modelling and publishing PS descriptions as CPSV-AP compliant linked data is introduced in Section 5. This process is followed for implementing a pilot, which is described in Section 6. Based on this pilot, benefits and challenges of CPSV-AP practical usage are identified. In Section 7, the main lessons learnt are articulated, while in Section 8 the conclusions and proposals for future work are provided.

2. Background

This section outlines background work on Services and PSs, Government Linked Data, CPSV and its linked-data application profile, namely CPSV-AP. Finally, the research objectives of this study are presented.

2.1 Services and Public Services

Service is an important concept in a number of diverse domains, including marketing, public sector, information systems and computer science. Over the last 50 years, a number of definitions for service have been proposed. The majority tend to conceptualize service as "*an intangible value-creating activity or (complex) task provided by one entity, usually referred to as the service provider, to another entity, usually referred to as the customer*" or consumer [14]. Others define service as "*a simple or a complex task or activity, executed within an organisation on behalf of a customer or organisation*" [15]. In the Unified Services Theory [16], a service is perceived as a sequence of steps, constituting a process, that modifies inputs in a way that delivers benefits to the customers. This process-centric view highlights that the customer is the primary beneficiary of the service and receives the value created through the transformation and integration of the provided inputs.

Additionally, service is the fundamental concept in Service Oriented Architectures (SOA). In SOA, service has been defined as "*a discrete unit of business functionality that is made available through a service contract.*" A service contract "*specifies all interactions between the service consumer and service provider*" [17]. It may include business aspects as well as technical aspects [18], e.g., service interface, interface documents, service policies, quality of service (QoS), performance. QoS and performance are usually managed through a service level agreement (SLA). Moreover, a service should be explicitly managed throughout its entire life cycle [17].

A public service (PS) is a service that is provided by or on behalf of a public administration entity to a societal entity (i.e., a citizen or a business) or another public administration entity. PSs fulfil needs of the societal entities and may be mandatory or not [19]. PSs are governed by rules that are derived from relevant formal framework, i.e., some combination of legislation and policy which can be set at local, national or supranational level [12]. Thus, the service process, the input, the output, the entities involved, and numerous other parameters are defined by a formal framework. As a result, unlike most other service industries, the services offered by or on behalf of public sector have a lower degree of heterogeneity and are easier to standardise.

A service description encompasses the information needed in order to use a service [18]. A (Public) Service description is usually structured as a set of well-defined metadata. PS metadata are frequently embedded in a PS model. A PS description should be ideally provided in machine-readable format in order to facilitate semantic interoperability [20] and reusability of PS descriptions, e.g., for federation of PS catalogues. Additionally, PS descriptions included in a PS catalogue are usually grouped (bundled) around thematic categories in order to be more easily

findable by citizens. A thematic category could be an area of economic activity (e.g., agriculture), an element of citizens' profile (e.g, disabled, elderly), life-events (e.g., getting married) etc.

PS provision is usually based on the execution of organizational processes [21] and supported by information systems. Thus, (Public) Service Management can be examined from a business perspective [21] (e.g., reengineering of organizational processes), from a technical perspective [22] (e.g., user interfaces), or both. Additionally, in the public sector it is increasingly important to group together PSs in order to better satisfy citizens' needs. This leads to introducing integrated PSs as[5] *"the result of bringing together government services so that citizens can access them in a single seamless experience based on their wants and needs"*. Usually, an integrated PS encapsulates one or more PS chains, i.e., a PS group where the output of a PS is required as input for another PS.

2.2 Government Linked Data

From a technical point of view, 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.*" [23]. Linked Data technology exploits Uniform Resource Identifiers (URI) and the Resource Description Framework (RDF) to describe data and their interrelations with other data [24].

Recently, there has been an increasing trend for publishing data on the web as linked data [4], in order to be more easily sharable and reusable [25], [26]. When two datasets share common entities and predicates, they can be integrated more easily without downgrading the quality of information [27]. Linked data technology aspires to realize the web of data where datasets can be interlinked through typed links. Interlinking of datasets is achieved by using the same URI across multiple datasets [1], [4]. The web of data aims at interconnecting isolated data islands (data silos) to a giant distributed dataset, also called the Linked Open Data Cloud [4]. Additionally, liked data are self-describing as their schemas are part of the data and are described in exactly the same way as the data itself do [4]. Thus, linked data technology potentially transforms fragmented datasets (data silos) to a global data space [23], [4]. Moreover, according to the 5-star deployment and rating scheme for open data proposed by Sir Tim Berners-Lee[6,7], linked data that are published with an open licence and include links to other datasets of the web of data are of highest quality, i.e. they are 5-star open data.

Consequently, public sector linked data hubs development, maintenance and interlinking could contribute to the sharing, reusing, federation and integration of public sector information. Public sector information that is published as linked data becomes also part of the web of data[8] and thus can be interlinked and enriched with information included in other datasets, such as DBpedia. Additionally, public organisations are allowed to maintain the control of their data, while facilitating easily consumption and re-use of their published linked open data [28]. Furthermore, linked data technology enables public organizations to publish their data in a structured and modular format, facilitating the systematic data management and knowledge discovery. Thus, the use of linked data constitutes a paradigm for publishing public sector information, including information about PSs and public organizations, which could offer significant benefits to citizens and businesses, as well as public administrations, including civil servants and policy makers.

However, the benefits of publishing public sector information as linked data are counterbalanced by the difficulties of publishing and exploiting linked data [1], [25]. Although linked data technology and standards are mature and powerful there is still much work to be done on formalizing the process of publishing government linked data by developing straightforward patterns that could be easily adopted by public organizations or institutes [28]. The literature also highlights that there are no systematic guidelines for publishing government linked data, including technical details about all steps of the process [2].

2.3 CPSV and CPSV-AP

The Core Public Service Vocabulary (CPSV) is a European standard that has been designed to model PS descriptions and bring consensus on the basic PS metadata across EU Member States [12]. As every Core Vocabulary, CPSV has not been designed to model every characteristic or every property of a PS across all EU Member States and all domains. Rather, it can be used as the basis for new PS models or as the linking element between existing PS models, thus achieving a minimum level of semantic interoperability. More specifically, new PS models can be designed from scratch based on CPSV while existing PS models can be mapped to CPSV.

CPSV-AP is an Application Profile of CPSV which incorporates linked data as an underpinning technology. An application profile is "*a specification that reuses terms from one or more*

[5] https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/integrated-public-services

[6] https://5stardata.info/en/

[7] https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data

[8] https://www.w3.org/DesignIssues/LinkedData.html

base standards, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as

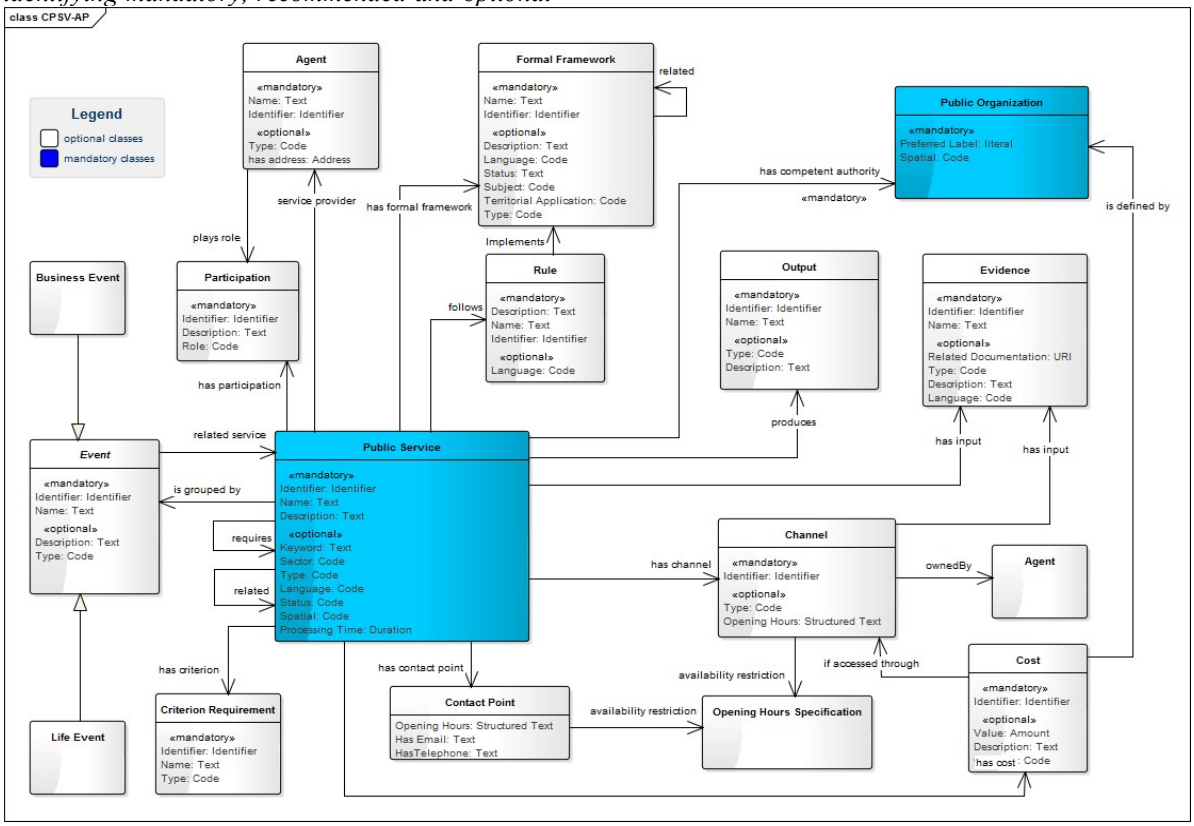

Fig. 1 UML diagram of the CPSV-AP 2.0 (Adapted from [29])

well as recommendations for controlled vocabularies to be used."[9]. In this paper, we use CPSV-AP 2.0, which was developed by the EC after consultation with EU Member States. CPSV-AP 2.0 was released in 2016 [29]. It reuses classes and properties of well-established semantic web vocabularies, including other EC ISA/ISA[2] Core Vocabularies, e.g., the Core Public Organisation Vocabulary (CPOV). The use of well-known vocabularies to represent data and their real-life meaning is the key factor that turns linked data to semantic data [30].

A PS model is compliant with CPSV-AP 2.0 if it includes, at least, information for the mandatory properties of the mandatory classes. In Figure 1, the mandatory classes are shown in blue. There are two mandatory classes, namely the *Public Service* class (which represents a public service itself) and the *Public Organization* class (which represents the public organisation that is the competent authority of a public service). In total, seven properties of these classes are mandatory. More specifically, the mandatory properties of *Public Service* class are *Identifier*, *Name*, *Description* and *Has Competent Authority*. The mandatory properties of *Public Organization* class are *Identifier*, *Preferred Label* and *Spatial* [29]. The property Identifier of *Public Organization* class share the same value (URI) with the property *Has Competent Authority* of *Public Service* class, which links *Public Service* class to *Public Organization* class. If an optional class of CPSV-AP is chosen to be included in a PS model, then its mandatory properties should be also included in that PS model [29].

2.4 Research Objectives

The objective of our research is two-fold: a) to device a process for using CPSV-AP to model and publish PS descriptions as linked data and b) to employ this process for conducting a pilot implementation in order to evaluate CPSV-AP practical usage. In the framework of this evaluation, potential benefits and challenges of CPSV-AP practical usage are identified and relevant lessons learnt are articulated.

To the best of our knowledge, there are a few only implementations based on CPSV-AP. Furthermore, there is no research reporting detailed guidelines, benefits and challenges from CPSV-AP practical usage. This is surprising as CPSV-AP is a promising European standard with significant potential benefits for all EU Member States.

[9] https://joinup.ec.europa.eu/collection/catalogue-services/open-specifications-cpsv-ap-and-sdg-services-model-describing-procedures

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

It is noted that in this research we aim to devise a process specifically for exploiting CPSV-AP instead of constructing a more general process for publishing any public sector information as linked data. We believe that the use and support of standards, such as CPSV-AP, is of critical importance for their success. In addition, we believe that a more general process in the public sector would have limited added value, as it would lead to generalizations with limited applicability. Attempting to devise a new general process that fits everything would lead to "yet another RDFizing process", which is something we believe will not have significant impact in the public sector.

3. Methodology

The methodology followed reflects basic principles of Design (Science) Research methodology (DSR) [31]. DSR begins by identifying current problems that need to be solved or potential opportunities for improvements in an actual application environment consisting of people, organizational systems and technical systems. Furthermore, as pointed out in [32] design science research "*is motivated by the desire to improve the environment by the introduction of new and innovative artifacts and the processes for building these artifacts.*". These principles apply to our methodology, which aspires to introduce a process for modelling and publishing PS descriptions as CPSV-AP compliant linked data. The methodology followed consists of the following parts:

- Initially, a systematic literature review is conducted to identify existing processes for publishing linked data, focusing on Government Linked Data (Section 4).
- Exploiting the results of literature review that is presented in Section 4, a process for publishing PS descriptions as Government Linked Data based on CPSV-AP 2.0 is introduced (Section 5).
- Using the process that has been introduced in section 5, a pilot is implemented in order to record relevant benefits and challenges of CPSV-AP practical usage (Section 6)
- Finally, the main lessons learnt are articulated (Section 7).

4. Related work

An extensive search was carried out in Scopus and Google Scholar in order to identify processes for publishing linked data, focusing on those for publishing government linked data. For this purpose, the systematic literature review methodology, proposed by Webster and Watson, [33] was employed. The final number of selected processes was 12, which are included in 11 academic papers and one relevant web page of the EC. The identified processes are presented in Table 1.

It is noted that in the literature, some of the identified processes were originally proposed as *life-cycles* similarly to software development incremental and iterative life-cycles [2], while others as *sequential processes* [25], [26], [34]. However, most of the cases include the continuous improvement of published linked data by performing iterations for updating and extending them. Besides the academic literature, a process based on Asset Description Metadata Schema-Application Profile, ADMS-AP, was proposed in the framework of EC ISA programme [35]. ADMS-AP is an EC standard for modeling interoperability solution descriptions. Therefore, we also include this process in our analysis.

Subsequently, we recorded the steps of each process and analyzed their contents. |In addition, following the Webster and Watson methodology, we have grouped together steps with similar contents and name them accordingly. Table 1 presents the identified processes (rows) as well as the steps within each process (columns).

Most of the processes suggest the selection and reuse of appropriate vocabularies for data modelling [1], [2], [4]. It should be noted however that this step is redundant in our case, since CPSV-AP has been selected as the employed vocabulary.

Additionally, some processes consider the transformation of initial data to RDF as one big step [25]. Although these processes provide information about different tasks that are required for transforming data to RDF, they do not explicitly list those tasks as distinct steps [25].

Moreover, many processes highlight the need for a URI policy either by reusing an existing or by crafting a new (custom) one [1], [2], [4], [6], [36]. However, some processes do not consider that as a distinct step [1], [2].

The literature revealed that currently there is no concrete process for publishing PS descriptions as linked data. Additionally, the comparative analysis of the identified processes, as depicted in Table 1, resulted in 7 steps that are common in the majority of processes for publishing linked data.

5. The proposed process

Based on the literature review presented in Section 4, a 7-step process is proposed for publishing PS descriptions as CPSV-AP compliant linked open data. This process is depicted in Figure 2. We note that in this paper

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

the term dataset refers to a set of PS descriptions. Thus, the names of the steps in Table 1 are renamed accordingly. The steps of the proposed process are described below:

Table 1 Processes for publishing linked data, focusing on government data

Steps (S) / Reference	S1: Understanding the context, preparation and dataset selection	S2: Reuse or definition of a URI design policy	S3: Data modelling	S4: Linked data generation and interlinking	S5: Linked data validation	S6: Linked data publication (and interlinking)	S7: Linked data exploitation
[36]	Selecting Target Data	Assigning URIs		Publishing Data: 1) RDFizer		Publishing Data: 2) RDF2HTML, 3) void Generator Interlinking Data Resources: 1) Linking to other resources, 2) Interlinking between internal resources	
[2]	Specification (identification, analysis and definition of license of the government data sources).	Specification (adoption or implementation of a URI design methodology).	Modelling (Reusing as much as possible existing suitable vocabularies or ontologies. In case there are not any, create some).	Linked data, i.e., RDF data, generation (transformation of the selected data sources to RDF format, data cleansing and linking with other datasets).		Linked data, i.e., RDF data, publication (linked dataset publication, metadata publication, e.g., provenance information, and enabling effective discovery).	Linked Open Government Data Exploitation (Development of added-value applications on top of published data using user-friendly GUI).
[4]	Datasets identification: - Obtaining a copy of the database models and data extracts (or creation of data in a way that can be replicated). - Real world objects modeling	Naming things on the Web with URIs.	Modeling data by reusing vocabularies whenever possible.	Converting data to RDF.	Validating a sample of produced linked data.	- Announcing a new Linked Dataset. - Publishing human and machine-readable descriptions. - Publishing the schema as a VoID description. - Attaching a license statement to each dataset.	
[3]	Contextualization: - selection of dataset(s), - definition of access and request model of the data as well as data output formats - data consumers description	Definition of the URI patterns that will form the RDF Graph	- Ontology Design - RDF Graph modeling	- Performing an ETL process: extraction, transformation		- Performing an ETL: loading - RDF Graph implementation. - Updating Graph Service - Development of a web portal that allows publishing of documentation and support for data use	Visualizations or mashups implementations
[1]	- Acquiring machine readable dataset that are hosted in data catalogues - Data Clean-up	Reusing or definition of a URI design policy	Data modelling	- Data reconciliation - Converting raw data to RDF - Data Interlinking		Data sharing	
[5]	Revealing data		Refining data	Reusing data		Releasing data	Running data / data-driven apps
[6]	Step #1 Prepare stakeholders Step #2 Select a dataset Step #4 Specify an appropriate licence	Step #5 Definition of good URIs for linked data	Step #3 Model data Step #6 Use standard vocabularies	Step #7 Convert data		Step #8 Provide machine access to data Step #9 Announce new datasets	Step #10 Recognize the social contract
[39]	Business level (Stakeholder analysis, requirements analysis, use case development), Information system level (Data identification)		Data modelling	- Transformation Configuration		Data Processing	
[25]			Turn data into RDF (modeling)	Turn data into RDF		Link your data	- Query/link your data - Visualize/analyse your data
[34]	Stage 1: Cleaning,	Stage 2: Triplification (naming)	Stage 2: Triplification (extraction, enrichment/mapping)	Stage 2: Triplification (conversion)		Stage 3: Interlinking Stage 4: Storage	Stage 5: Visualization
[26]		Assigning IRIs to the entities of the dataset	Linked Data Modeling	Linked Data Generation	Linked Data Generation (Validation)	- Linked Data Publishing - Linked Data Licensing, Announcement and Maintenance)	
[35]			1. Fill in the ADMS-AP template spreadsheet.	2. Import the spreadsheet in OpenRefine. 3. Apply the ADMS-AP skeleton. 4. Export RDF.			

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

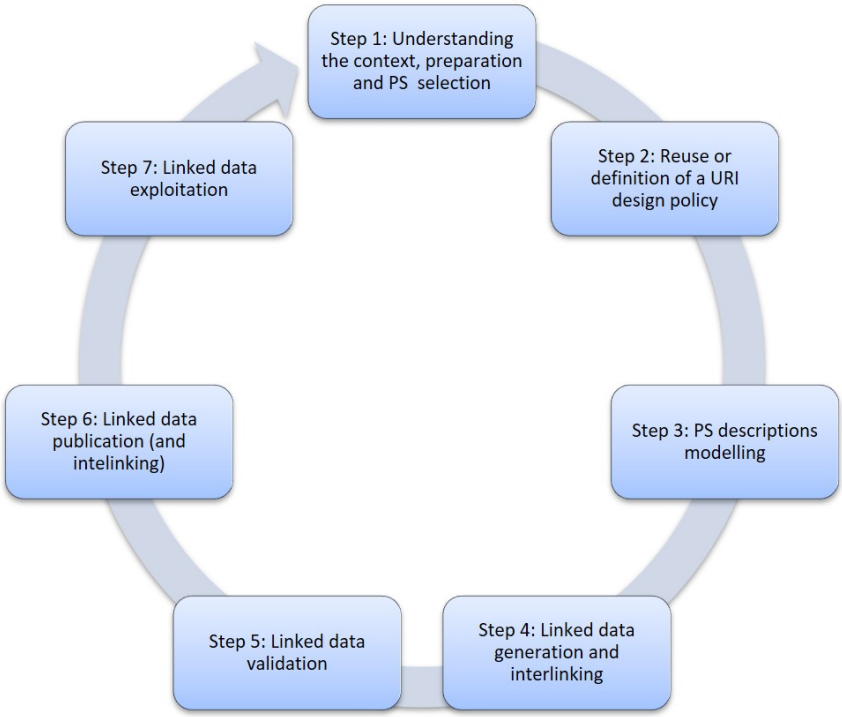

Fig. 2 The proposed process for publishing public services descriptions based on CPSV-AP

Step 1: Understanding the context, preparation, and PS selection

Initially, the details of the context should be well understood, as for example the geographical scope of the PSs, the PS model that is currently used, the competent public authority or authorities, etc.

The preparation includes the establishment of good cooperation of the project's team with the competent public authority or authorities. This requires top management consent and support as well as establishment of good working relations with relevant domain experts. Trust is a critical success factor of this cooperation.

PS selection could be based on various criteria, as for example: (a) PS usage statistics. High popularity of a PS probably means it is of high interest for citizens or businesses and thus of high value for the society [6]. (b) Availability of adequate, reliable, and structured PS descriptions.

PS descriptions are usually offered with an open licence even when this is not explicitly stated. It is a good practice to define an open licence for the linked dataset that will be produced to facilitate its reusability [6], [26].

Step 2: Reuse or definition of a URI design policy

CPSV-AP model uses URIs extensively. Although general guidelines for designing persistent URIs are available, see for example [4], [37] and [38], they are not directly applicable for PS modelling using CPSV-AP. Thus, a new URI design policy has to be defined [1]. The EU recommends the following URI structure: http://{domain}/{type}/{concept}/{reference} [37]. We suggest that *domain* starts with the address of the server that will host the PS descriptions as linked data. This enables adding related content that can be dereferenced [4].

For example, we suggest that the URI pattern of the identifier property of the *Public Service* class, could be:

http://<host>/PublicServices/id/ps/psxxxx, where *domain* equals *<host>/PublicServices*, *type* equals *id*, *concept* equals *ps* and *reference* equals to a code name of each PS, e.g., ps0001.

Step 3: PS descriptions modelling

In case PS descriptions exist in a structured form, i.e., modelled based on a PS model, then the concepts of the existing PS model should be mapped to the concepts of CPSV-AP. In case information about PSs exists in an unstructured form, e.g., plain text, then PS descriptions should be modelled from scratch using CPSV-AP. In both cases, we extract the information (metadata) of the selected PS descriptions and map it to CPSV-AP 2.0 properties.

The extraction of PS descriptions should be performed automatically or semi-automatically utilising a mapping software application (tool), if possible. However, even in cases where PS descriptions exist in structured form, it is not always easy to automatically transform the existing model to CPSV-AP, e.g., due to syntactic or semantic inconsistencies. For example, an *ad-hoc* PS model may suggest that

the names of all Evidences related to a PS are stored all together in one paragraph. However, CPSV-AP suggests that the name of each Evidence is stored separately. . Additionally, CPSV-AP extensively uses URIs which is not often the case in ad hoc PS models developed for publishing PS web catalogues. .

When a mapping tool is not available or appropriate the extraction of information, modelling of PS descriptions, and mapping to CPSV-AP should be performed manually. PS descriptions modelling can be facilitated by CPSV-AP 2.0 Template Spreadsheet[10], which is an Excel template for modelling PS descriptions using CPSV-AP 2.0.

Step 4: Linked data generation and interlinking
In this step, CPSV-AP compliant PS descriptions are converted to linked data. Additionally, the produced linked dataset can be interlinked with other linked datasets of the semantic web, as for example DBpedia. This interlinking would result in a better integration of the produced linked dataset into the Linked Open Data Cloud. Alternatively, the interlinking with other linked datasets can be done at Step 6. The interlinking of two linked datasets can be achieved using appropriate predicates, such as *owl:sameAs*.

There are various technical solutions and tools for transforming different types of structured or unstructured data to linked data. As the focus of this paper is on the process and not on the technical solutions (e.g., software tools) for supporting the transformation process, we will not further elaborate on the variety of technical solutions. Nevetheless, we provide some technical details for conducting this step utilising OpenRefine and CPSV-AP 2.0 Template Spreadsheet as a good practice example.

OpenRefine (formerly Google Refine) version 2.6-rc.2[11] along with the corresponding RDF extension, namely (alpha) Open Refine 2.6 compatible[12] developed by DERI[13], can be used to transform a set of PS descriptions to CPSV-AP compliant linked data. More specifically, the CPSV-AP 2.0 Template Spreadsheet file containing the PS descriptions in tabular format is first uploaded to OpenRefine. RDF extension provides a graphical user interface for aligning CPSV-AP 2.0 schema with sheets and columns of CPSV-AP 2.0 Template Spreadsheet. CPSV-AP 2.0 RDF Skeleton[14] automates this part of the process[15]. Finally, linked data can be extracted as an RDF/Turtle file, since this serialisation is compact and human understandable [1], [4]. However, more options regarding RDF serialisations are available.

Step 5: Linked data validation
In this step, produced linked data are validated for syntax violations and compliance with CPSV-AP 2.0 schema. There are different methods that can be employed in the validation process. A relevant good practice is to use the CPSV-AP 2.0 validator, a tool developed in the framework of EC ISA/ISA^2 programmes.

Step 6: Linked data publication (and interlinking)
After validation, the linked dataset (e.g., the RDF/Turtle file) is uploaded to an RDF store and becomes publicly available as an RDF graph through a SPARQL end-point. Additionally, triples using *owl:sameAs* predicate can be added, thus creating links to other linked datasets of the web of data. An example of such RDF store is Virtuoso Universal Server, which also provides a SPARQL end-point.

Step 7: Linked data exploitation
The published linked dataset can be exploited in a number of ways. Some examples are shown below:
(a) Valuable knowledge about important PS characteristics can be acquired more easily. For example, the most frequently required input documents (i.e Evidences using CPSV-AP 2.0 terminology) can be identified.
(b) The PSs that require as input the output of another PSs can be identified. In these cases, PS chains can be revealed, promoting integrated (one-stop) PS delivery.
(c) Federated queries with DBpedia or other linked datasets of the web of data, can result to generation of knowledge beyond that included into the produced linked dataset.

6. Pilot implementation

In this section, the proposed process is used to publish as CPSV-AP compliant linked data a set of PS descriptions included in an official PS catalogue. Relevant challenges and benefits are identified and presented.

10 https://joinup.ec.europa.eu/solution/core-public-service-vocabulary-application-profile/distribution/cpsv-ap-specification-v20-xlsx-template
11 https://github.com/OpenRefine/OpenRefine/releases/tag/2.6-rc.2
12 https://github.com/fadmaa/grefine-rdf-extension/releases/tag/v0.9.0
13 https://openrefine.org/download.html#
14 https://joinup.ec.europa.eu/solution/core-public-service-vocabulary-application-profile/distribution/cpsv-ap-specification-v20-rdf-skeleton
15 https://joinup.ec.europa.eu/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

6.1 Steps

Step 1: Understanding the context, preparation and PS selection
The context of the pilot implementation is the official PS catalogue of the Region of Epirus in Greece. This so-called "The Citizen's Guide of the Region of Epirus" has been awarded as "best practice" [40] at the European Public Sector Awards 2015, organised by the European Institute of Public Administration (EIPA) with the support of the European Commission[16].

The categorization of PSs is based on both the administrative structure of the Region of Epirus and citizens' professional activities, thus facilitating PS discovery by citizens. More specifically, at the higher level of categorization, all public services are grouped by Direction Generals (DG). Every DG is related to a wide area of economic activity, for example DG Agricultural Economy and Veterinary. At the second and third level (when applicable) of categorization PSs are grouped by a thematic (usually professional) category. The PS descriptions included in this PS catalogue are structured based on an *ad hoc* PS model (henceforth referred to as the *Epirus PS model)*.

As the implementation described in this article is a pilot one, the preparation was restricted to communications and cooperation with domain experts of the Region of Epirus. The selection of PSs was made based on the following criteria: (a) the popularity of a PS, (b) the availability of adequate and reliable information about a PS, (c) the inclusion of PS chains. As a result, 45 PSs were selected.

Step 2: Definition of a URI design policy
The proposed URI patterns are shown in Table 2. In this table, xxxx in psxxxx and costxxxx refers to an increasing number, i.e., 0001, 0002, etc. Following linked data practice, the *Public Service* class of CPSV-AP 2.0 is interlinked with other classes by sharing the same URI. Specifically, a property of the PS class (1st column in Table 2) has as value a URI, which is also the value of the identifier property of another class of CPSV-AP 2.0 (3rd column in Table 2). In Table 2, the notation *A.B* is used to denote *Property B* of *Class A*.

Step 3: PS descriptions modelling
In this step, we extracted manually the information of the selected PS descriptions and mapped it to CPSV-AP 2.0 properties, using the CPSV-AP 2.0 Template Spreadsheet, as shown in Table 3. A row of the *Public Service* class sheet of the template spreadsheet contains a PS description. If a property (e.g., *PublicService.HasInput*) has multiple values (e.g., multiple required input documents) then these values are added to the same column in subsequent rows. Additionally, in the first column the URI of the *identifier* of that PS is added.

The *Category* property of *Epirus PS model* is mapped to *<PublicService:sector>* property of CPSV-AP 2.0, which represents the area of economic activities or sector that a PS is related to, e.g. environment, safety, housing, etc. In order for our PS descriptions to be compliant with CPSV-AP 2.0, we replaced the initial categories of PSs with categories included in the *List of NACE codes*, as recommended by CPSV-AP 2.0 [29].

Step 4: Linked data generation
In this step, we transformed the PS descriptions filled in the CPSV-AP 2.0 template spreadsheet from tabular format to an RDF/Turtle file using OpenRefine as described in Section 5.

The installation and use of Open Refine and the corresponding RDF extension is technically simple. However, it must be ensured that the version of Open Refine is compatible with the version of RDF extension.

Step 5: Linked data validation
The produced RDF/Turtle file was uploaded to the CPSV-AP 2.0 validator. The aim of the tool is to validate the file in terms of syntax violations and other inconsistencies with the CSV-AP 2.0 schema.

At this step, we encountered some minor technical problems that were fixed very quickly in cooperation with the developers of the tool. These problems concerned mainly some false warnings that the validator initially produced.

Step 6: Linked data publication (and interlinking)
After validation, we loaded the RDF/Turtle file to a Virtuoso Universal Server. As a result, the produced linked dataset became publicly available through a SPARQL end-point. The IRI of the graph of the published linked dataset is: http://data.dai.uom.gr:8890/CPSV-AP and the URL of the SPARQL endpoint is: http://data.dai.uom.gr:8890/sparql.

Additionally, in order to make a link with DBpedia, we added the following triple to the graph of the published linked dataset: *<http://data.dai.uom.gr:8890/PublicOrganization/id/nuts3/GRC_PRF_IPEIR> <http://www.w3.org/2002/07/owl#sameAs> <http://dbpedia.org/resource/Epirus_(region)>*.

[16] http://www.epsa2015.eu/

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

Table 2 URI design policy

Property of the Public Service class	URI pattern	Property of the interlinked class
PublicService.identifier	http://data.dai.uom.gr:8890/PublicServices/id/ps/psxxxx	
PublicService.hasInput	http://data.dai.uom.gr:8890/PublicServices/id/doc/<artifact>	Evidence.identifier
PublicService.produces	http://data.dai.uom.gr:8890/PublicServices/id/doc/<artifact>	Output.identifier
PublicService.hasCost	http://data.dai.uom.gr:8890/PublicServices/id/cost/costxxxx	Cost.identifier
PublicService.hasChannel	http://data.dai.uom.gr:8890/PublicServices/id/channel/<channel>	Channel.identifier
PublicService.hasCompetentAuthority	http://data.dai.uom.gr:8890/PublicOrganization/id/nuts3/GRC_PRF_IPEIR	PublicOrganisation.identifier
PublicService.hasFormalFramework	http://data.dai.uom.gr:8890/LegalFramework/id/law/<number_year>	FormalFramework.identifier

Table 3 Mapping of "the Citizen's Guide of the Region of Epirus" model to CPSV-AP 2.0

Property of the Epirus PS model	CPSV-AP 2.0 property (Class.Property)
Title of the PS	Public Service.Name
Category	Public Service.Sector
Description	Public Service.Description
Required input documents	Public Service.Has Input
Legal Framework	Public Service.Has Formal Framework
Cost	Public Service.Has Cost
Processing time needed for the completion of the execution of PS	Public Service.Processing Time
Produced output document(s)	Public Service.Produces
Validity time of the output document(s) (if exists)	-
Renewal procedure (if exists)	-
Competent organizational unit	Public Service.Has Contact Point
Additional Sources of Information	-
Notes	-
Relevant Files	-

Table 4 Mapping of the categories of the selected PSs of Epirus PS model to CPSV-AP 2.0

Directorate General (DG) – Thematic Category of Epirus PS model	Number of PSs	Corresponding property of the CPSV-AP 2.0, namely PublicService.sector
DG Agricultural Economy and Veterinary - Plant production	5	A1.6.1.-Support activities for crop production (2 PSs), G46.1.8. Agents specialiased in the sale of other particular products (1 PS), G46.3.1. Wholesale of fruit and vegetables (1 PS), G47.2.1. Retail sale of fruit and vegetables in specialised stores (1 PS)
DG Development - Technical professions	13	M71.1.2 - Engineering activities and related technical consultancy
DG Development Planning, Environment and Infrastructure – Environment	5	E38.1.1 - Collection of non-hazardous waste (1 PS), F43.9 - Other specialised construction activities (4 PS)
DG Public Health and Social Care - Creative centers	4	Q88.9.1 - Child day-care activities
DG Transport and Communication - Driving licenses	18	H52.2.1 - Service activities incidental to land transportation

Step 7: Linked data exploitation

Finally, we queried the produced linked dataset through the SPARQL endpoint in order to acquire new knowledge. We carried out the following scenarios: (a) extract information about cost and processing time of the PSs of a thematic category, (b) reduce administrative burden through the elimination of required input documents of PSs, (c) acquire information about a geographical area, for example a Region, that delimits the scope of the PS catalogue.

In the framework of *scenario (a)* we execute the following query: *How much is the cost and what is the expected processing time of each PS of the thematic category "DG Development-Technical professions"?* The relevant SPARQL query is shown in Listing 1. The result of this query is depicted in Table 5 and suggests that the cost and the expected processing time of a PS of the selected thematic category range from 15 € up to 400 € and from 15 days up to one month accordingly.

Listing 1 SPARQL query regarding cost and expected processing time of each PS of the thematic category "DG Development-Technical professions"

```
PREFIX cpsv: <http://purl.org/vocab/cpsv#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX cv: <http://data.europa.eu/m8g/>
SELECT ?PSname ?cost_value ?Processing_Time
FROM http://data.dai.uom.gr:8890/CPSV-AP
WHERE{
  ?x a cpsv:PublicService.
  ?x dct:title ?PSname.
  ?x cv:hasCost ?cost.
  ?cost cv:value ?cost_value.
  ?x cv:processingTime ?Processing_Time.
  ?x cv:sector "M71.1.2 - Engineering activities and related technical consultancy".
}
ORDER BY(?PSname)
```

Table 5 Result of SPARQL query regarding cost and expected processing time of each PS of the thematic category "DG Development-Technical professions"

PSname	cost_value	processing_time
Announcement of the commencement of a construction machinery assistant operator professional activities	15.0	P1M
Announcement of the commencement of a mechanical engineer professional activities	15.0	P1M
Announcement of the commencement of a plumber professional activities	15.0	P1M
Announcement of the commencement of a refrigeration technician professional activities	15.0	P1M
Announcement of the commencement of a technician of mechanical equipment professional activities	15.0	P1M
Announcement of the commencement of an electrical engineer professional activities	17.0	P1M
Announcement of the commencement of an electrician professional activities	17.0	P1M
Approval of application for the participation in exams for obtaining a construction machinery operator license	31.0	P15D
Matching the old license of a construction machinery operator with a new type of license	10.0	P15D
Matching the old license of a plumber with a new type of license	400.0	P15D
Matching the old license of a refrigeration technician with a new type of license	400.0	P15D
Matching the old license of a technician of mechanical equipment with a new type of license	400.0	P15D
Matching the old license of an electrician with a new type of license	400.0	P15D

In the framework of *scenario (b)* we execute the following queries:

1) How many PSs require each input document? The SPARQL query is shown in Listing 2. The result of this query is depicted in Table 6 and indicates the most frequently required input documents. This information can be used in many different policy making scenarios. For example, we now have a clear idea on the number of PSs that will be affected if a certain required input document is eliminated or provided electronically.

Listing 2 SPARQL query regarding the number of PSs that require a specific input document

```
PREFIX dct:<http://purl.org/dc/terms/>
PREFIX cpsv: <http://purl.org/vocab/cpsv#>
SELECT DISTINCT ?Evidence (count(?x) as ?Number_of_PS)
FROM http://data.dai.uom.gr:8890/CPSV-AP
WHERE {
  ?x a cpsv:PublicService.
  ?x cpsv:hasInput ?i.
  ?i dct:title ?Evidence
}
ORDER BY DESC(?Number_of_PS)
```

Table 6 Part of the result of SPARQL query regarding the number of PSs that require a specific input document

Evidence	Number_of_PS
Certified copy of identity card or passport	24
Filled and signed application form	19
Receipt of fee payment	19
Driving license of B category	16
Residence permit issued in accordance with the provisions in force (only for citizens outside European Union)	16
Driving license of AM category	13
Driving license of A1 category	13
Driving license of A2 category	13
Driving license of A category	13
Driving license of C category	12
Driving license of C1 category	12
Driving license of D category	12
Driving license of D1 category	12

2) Which PS has as input the output of another PS thus constituting a PS chain? The relevant SPARQL query is shown in Listing 3. The result of this query is depicted in Table 7.

We could execute more queries, as for example: How many input documents are required for the execution of each PS? Which PS require a specific input document, for example the ID card? However, the relevant SPARQL queries are omitted, due to space limitation.

Listing 3 SPARQL query identifying PS chains

```
PREFIX dct:<http://purl.org/dc/terms/>
PREFIX cpsv: <http://purl.org/vocab/cpsv#>
SELECT DISTINCT ?PS1 ?PS2
FROM http://data.dai.uom.gr:8890/CPSV-AP
WHERE {
  ?ps1 a cpsv:PublicService.
  ?ps2 a cpsv:PublicService.
  ?ps1 cpsv:produces ?o.
  ?ps2 cpsv:hasInput ?o.
  ?ps1 dct:title ?PS1.
  ?ps2 dct:title ?PS2.
}
ORDER BY(?PS1)
```

Table 7 Part of the result of SPARQL query identifying PS chains

PS1	PS2
Adding of code 96 in valid driving license of B category	Issuing of Certificate of Professional Competence (C.P.C.) of Initial Training
Adding of code 96 in valid driving license of B category	Reissuing of driving license of categories restricted for health or retirement reasons
Adding of code 96 in valid driving license of B category	Conversion of American, Canadian, Australian, Japanese, South Korean and South African driving license into Greek driving license
Adding of code 96 in valid driving license of B category	Issuing of certified copy of driving license due to loss, theft, damage or alteration
Adding of code 96 in valid driving license of B category	Exchanging of driving license issued by a Member State of E.E. with a greek driving license of equivalent category
Adding of code 96 in valid driving license of B category	Replacement of driving license
Adding of code 96 in valid driving license of B category	Limitation of categories of valid driving license for health or retirement reasons

In the framework of *scenario (c)* we executed the following query: *How much population and what geographical area are covered by a PS of this PS catalogue?* The relevant federated SPARQL query obtains information from DBpedia and is shown in Listing 4. The result of this query is depicted in Table 8 and provides information about the geographical area and the population that is covered by a PS. These data may be useful for policy making. Additionally, they demonstrate the integration of the selected PS descriptions published as government linked data into the Linked Open Data Cloud.

Listing 4 SPARQL query identifying population and geographical area covered by a PS

```
PREFIX cv: <http://data.europa.eu/m8g/>
PREFIX cpsv: <http://purl.org/vocab/cpsv#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?ps ?population ?area
WHERE{
graph <http://data.dai.uom.gr:8890/CPSV-AP>{
  ?ps a cpsv:PublicService.
  ?ps cv:hasCompetentAuthority ?auth.
  ?auth owl:sameAs ?dbauth.}
service <https://dbpedia.org/sparql>{
  ?dbauth <http://dbpedia.org/ontology/populationTotal> ?population.
  ?dbauth <http://dbpedia.org/ontology/PopulatedPlace/areaTotal> ?area.}
}
ORDER BY(?ps)
```

Table 8 Part of the result of SPARQL query identifying population and geographical area covered by each PS

ps	population	area
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0001	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0002	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0003	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0004	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0005	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0006	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>
http://data.dai.uom.gr:8890/PublicServices/id/ps/ps0007	336856	"9203.22"^^<http://dbpedia.org/datatype/squareKilometre>

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

6.2 Benefits

We suggest that potential benefits of using CPSV-AP to publish PS descriptions as linked data are the following:

a) Citizens and businesses can acquire information about PSs, which has a clear and standardised structure. This could improve PS consumers experience and promote PS catalogues utilisation. This in turn will save resources (e.g., time and cost) for fulfilling PSs.

b) Policy makers can study *what-if* scenarios for PS reengineering or improvement. For example, they can answer questions such as "*How many PSs require each input document?*", as shown in Subsection 6.1. Thus, they can more easily estimate the effect of the abolishment of a required input document. Furthermore, they can review a chain of PSs concerning a specific life or business event [29], [41].

c) Public organizations can increase transparency and democratic accountability as data about PSs becomes publicly available, open, sharable and reusable [1]. Additionally, they can achieve more effective and more efficient PS portfolio management. Moreover, they can participate more easily in federation portals of PS descriptions, resulting in increasing their interoperability and their collaboration with other public organisations.

d) IT industry can develop new added-value data-driven services and applications, based on published linked open datasets containing PS descriptions. This would result among others in new job positions and thus, economic growth.

6.3 Challenges

Based on the pilot implementation presented in Section 6 we describe below the main challenges of CPSV-AP practical usage for every step of the proposed process presented in Section 5.

Step 1: PS selection

It is not always easy to find statistics about PS popularity.

Step 2: Reuse or definition of a URI design policy

The definition of a URI design policy constitutes a major challenge. To the best of our knowledge, no such PS URI design policy exists at European level. Such policy should provide specific URI sets for essential concepts of PS modelling such as required (input) documents (or other Evidences), produced (output) documents (or other Output), etc.

Step 3: PS descriptions modelling

The main challenges of PS modelling, using CPSV-AP 2.0, are shown in Table 9. These challenges concern mainly the mapping of Epirus PS model concepts to CPSV-AP 2.0 concepts.

Table 9 Identified modelling challenges, solutions followed and suggestions

Modelling Challenge	Solutions followed or relevant suggestions
There is not always an exact match between the format of the properties of Epirus PS model and the properties of CPSV-AP 2.0. For example, the list of required (input) documents is a text field in the Epirus PS model while the value of the corresponding property of CPSV-AP 2.0, namely *<PublicService.hasInput>,* is a list of URIs, where a URI identifies only one required (input) document.	The information for every required (input) document of a PS, was manually extracted from the Epirus PS catalogue and filled in the corresponding property of the CPSV-AP 2.0 template spreadsheet. Additionally, a URI for every required document was created. The URIs of the required documents were filled in the *<PublicService.hasInput>* property of each PS, while additional information for very required document was filled once in the *Evidence* class.
The text of the Epirus PS catalogue is in Greek. An automated translation tool is provided by its website and thus the translation might not be accurate.	The text of the selected PS descriptions was translated manually in English by the authors.
The categorization of PSs cannot be directly mapped to CPSV-AP 2.0 categorization.	We mapped, as closely as possible, the categories of the Epirus PS catalogue to the proposed list of values for *<Public Service.sector>* property of CPSV-AP 2.0.
The cardinality of some properties of the Epirus PS model does not match with the proposed cardinality of the corresponding CPSV-AP 2.0 properties	We suggest that the cardinalities of the properties of CPSV-AP 2.0 should be updated in a future version. Actually, in CPSV-AP 2.1, some cardinalities have been changed. For example, the cardinality of the property *<PublicService.hasContactPoint>* has been changed to [0..n]. In CPSV-AP 2.0, it was [0..1], which was not suitable as the contact points for a PS are sometimes more than one. In the case of the Epirus PS catalogue, they are usually four for every PS, i.e., one for every regional unit of the Region of Epirus.
Taxonomies or code lists, e.g., for required (input) documents, are not available.	Taxonomies or code lists should be defined for as many PS modelling concepts as possible in national level or ideally in European level to facilitate semantic interoperability and cross border PS provision.

Modelling Challenge	Solutions followed or relevant suggestions
Concepts related to Provenance or Ownership of PS descriptions are missing from CPSV-AP 2.0 specification.	Although, the property <*PublicService:isDescribedAt*> has been added to CPSV-AP 2.1 [42] in order to link a PS description to its authoritative source, e.g., a regional or a national PS catalogue, we suggest that more effort is needed for the integration to CPSV-AP concepts related to PS descriptions provenance, as well as to PS descriptions owner(s) [26] or steward(s). These properties are particularly important due to the authoritative nature of PS descriptions [1], [25]. Provenance could include information about how an RDF graph was created (e.g., the steps of the transformation process), its authoritative source [1], etc. This is very important for the assessment of the quality of the published linked data [2], [25], [36], as well as the enhancement of accountability and trust [1] between public authorities and PS consumers.
A concept related to the licensing scheme of PS descriptions is missing from CPSV-AP 2.0 model.	It is a good practice to publish information about the licensing scheme [4] in human readable and machine readable format (e.g., as linked data [26]) as it increases the potential of data (here PS descriptions) reuse [6].

Step 4: Linked data generation, Step 5: Linked data validation and Step 6: Linked data publication

The conversion of CPSV-AP compliant PS descriptions to linked data (Step 4 and Step 5) and publishing (Step 6) is not fully automated or supported by an integrated software application (tool). Thus, it is a technical task that requires knowledge and experience on semantic web technology [25], [1]. Usually, public servants and even software engineers are not familiar with this technology [1].

Step 7: Linked data exploitation

The identified challenges for the exploitation of a linked dataset containing CPSV-AP 2.0 compliant PS descriptions are the following:

- It requires in-depth knowledge of CPSV-AP 2.0 schema as well as technical background on semantic web technology.
- There is lack of guidance for the presentation of PS descriptions to humans, e.g. in html web pages.

7. Lessons learnt

The main lessons learnt, through the pilot implementation presented in Section 6 are:

a) A concrete *URI design policy* as well as *taxonomies* or *code lists* for the major concepts of PS modelling domain, such as required (input) documents, should be defined ideally at the European level.

b) The *categorization* of PSs should be also defined ideally at the European level. Such categorization may include elements of PS consumers' profiles, life or business events, thematic categories (e.g., professional or economic activities), etc.

c) An *integrated software tool* as well as detailed *guidelines* supporting and automating as much as possible the entire process of PS descriptions modelling and publishing using CPSV-AP, might facilitate the adoption of CPSV-AP. The proposed process (Section 5) constitutes a potential basis for developing such a tool or issuing relevant guidelines.

A relevant *integrated software tool* should (a) Provide a user-friendly graphical user interface [1], [2], [26] and a set of wizards, hiding the underlying complexity from users that are not familiar with linked data and semantic web technology [25]. (b) Incorporate the ability to publish, manage and update PS descriptions in machine readable (e.g., as linked-data) as well as in human readable (html) format [1], [25]. (c) Incorporate natural language processing technology or a predefined set of questions for querying the linked data repository without requiring knowledge of SPARQL or CPSV-AP schema. This tool could be useful for citizens and public servants without relevant technical background [25]. Such functionality is provided, for example, by chatbots [43], [44]. (d) Include visualization and analytics modules [25].

8. Conclusions and future work

CPSV-AP is a promising PS model proposed by the EU as the European standard for PS descriptions modelling. Despite its potential, there is limited research on CPSV-AP practical usage. In order to facilitate and systematise PS descriptions modelling and publishing as linked data (which is the highest quality, 5-star open data according to Sir Tim Berners-Lee) using CPSV-AP we propose a relevant process, including technical details and guidelines. Additionally, benefits and challenges of CPSV-AP practical usage are identified through a pilot implementation and relevant lessons learnt are provided.

We argue that the exploitation of CPSV-AP could potentially provide significant benefits for all stakeholders of PS provision process, e.g.

citizens, businesses, policy makers, public organisations (including public servants), and IT industry. This includes accompanying the proposed process with relevant tools and guidelines.

We suggest that the research presented in this article contributes also to the linked data corpus of knowledge. In future work, the proposed process could be exploited for systematizing the modelling and publishing process of various government datasets as linked data using EC ISA/ISA2 Core Vocabularies, e.g. the Core Public Organisation Vocabulary (CPOV) for datasets including information about public organisations or using other ontologies. Moreover, tools and solutions that are used in the framework of the proposed process based on CPSV-AP can be reused in the framework of similar processes that are based on other Core Vocabularies or ontologies.

Additionally, future work includes combining, or integrating, CPSV-AP with other EC ISA/ISA^2 Core Vocabularies. Additionally, it would be interesting to interlink or federate PS descriptions of different public organisations to create a government linked data space where knowledge management and visualisation methods could be applied. That would enable PS portfolio management, facilitating also comparative review, analysis and collaborative continuous improvement of PSs.

References

[1] Maali, F., Cyganiak, R., Peristeras, V. (2012). A Publishing Pipeline for Linked Government Data. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds) The Semantic Web: Research and Applications. ESWC 2012. Lecture Notes in Computer Science, vol 7295, pp 778-792. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30284-8_59

[2] Villazón-Terrazas, B., Vilches-Blázquez, L.M., Corcho, O., Gómez-Pérez, A. (2011). Chapter2: Methodological Guidelines for Publishing Government Linked Data. In: Wood, D. (eds) Linking Government Data. Springer, New York, NY, pp 27-49. https://doi.org/10.1007/978-1-4614-1767-5_2

[3] Cifuentes-Silva F, Sifaqui C, Labra-Gayo JE (2011) Towards an architecture and adoption process for Linked Data technologies in Open Government contexts: A case study for the Library of Congress of Chile. I-Semantics '11: Proceedings of the 7th Int. Conf. on Semantic Systems, Sept. 7-9, Graz, Austria, pp 79–86. https://doi.org/10.1145/2063518.2063529

[4] Hyland, B., Wood, D. (2011). Chapter 1: The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web. In: Wood, D. (eds) Linking Government Data. Springer, New York, NY, pp 3-26. https://doi.org/10.1007/978-1-4614-1767-5_1

[5] Kaschesky M, Selmi L (2013) Fusepool R5 Linked Data Framework: Concepts, Methodologies, and Tools for Linked Data. dg.o '13: Proceedings of the 14th Annual International Conference on Digital Government Research, June 17-20, Quebec City, Canada, pp 156–165. https://doi.org/10.1145/2479724.2479748

[6] Hyland B, Atemezing G, Villazón-Terrazas B (2014) Best Practices for Publishing Linked Data. W3C Working Group Note. https://www.w3.org/TR/2014/NOTE-ld-bp-20140109/. Accessed 20 February 2021

[7] Tambouris E, Manouselis N, Costopoulou, C (2007) Metadata for digital collections of e-government resources. The Electronic Library. 25(2):176–192.

[8] Tambouris E, Kavadias G, Spanos E (2005) The Governmental Markup Language (GovML). Journal of E-Government 1(2):59–70. https://doi.org/10.1300/J399v01n02_05 https://doi.org/10.1108/02640470710741313

[9] Peristeras, V., Tarabanis, K. (2004). Advancing the Government Enterprise Architecture – GEA: The Service Execution Object Model. In: Traunmüller, R. (eds) Electronic Government. EGOV 2004. Lecture Notes in Computer Science, vol 3183. Springer, Berlin, Heidelberg, pp 476-482. https://doi.org/10.1007/978-3-540-30078-6_83

[10] Tambouris E. (2008) Introducing the need for a domain model in Public Service Provision (PSP) egovernment systems. 3rd International Conference on Digital Information Management, ICDIM 2008, pp. 794-799. https://doi.org/10.1109/ICDIM.2008.4746837

[11] Shukair G, Loutas N, Peristeras V, Sklarß S (2013) Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. Computers in Industry 64(1):10-18. https://doi.org/10.1016/j.compind.2012.09.003

[12] European Commission (2013) Core Public Service Vocabulary specification. ISA Programme. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary/distribution/cpsv-documentation-pdf. Accessed 23 May 2022

[13] Gerontas A, Peristeras V, Tambouris E, Kaliva E, Magnisalis I, Tarabanis K, (2021) Public Service Models: a systematic literature review and synthesis. IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 2, pp 637-648. https://doi.org/10.1109/TETC.2019.2939485

[14] Loutas N (2015) On improving service provision through the use of customer-centric semantic service models. PhD Thesis. https://doi.org/10.12681/eadd/39330

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

[15] Dumas, M., O'Sullivan, J., Heravizadeh, M., Edmond, D., ter Hofstede, A. (2003). Towards a Semantic Framework for Service Description. In: Meersman, R., Aberer, K., Dillon, T. (eds) Semantic Issues in E-Commerce Systems. IFIP - The International Federation for Information Processing, vol 111. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-35658-7_17

[16] Sampson SE, Froehle CM (2006) Foundations and implications of a proposed Unified Services Theory. Production and Operations Management 15(2):329-343. https://doi.org/10.1111/j.1937-5956.2006.tb00248.x

[17] Rosen M, Lublinsky B, Smith KT, Balcer MJ (2008) Applied SOA: Service-Oriented Architecture and Design Strategies. Wiley Publishing, Inc.

[18] Norton B, Kerrigan M, Mocan A, Carenini A, Cimpian E, Haines M, Scicluna J, Zaremba M (2008) Reference Ontology for Semantic Service Oriented Architectures. Version 1.0. OASIS. http://docs.oasis-open.org/semantic-ex/ro-soa/v1.0/pr01/see-rosoa-v1.0-pr01.html#_Toc214854775. Accessed 20 February 2021

[19] Peristeras V, Tarabanis K (2008) The GEA: Governance enterprise architecture-framework and models. Advances in Government Enterprise Architecture, pp. 229-262. https://doi.org/10.4018/978-1-60566-068-4.ch011

[20] Directorate-General for Informatics (European Commission) (2015) e-Government Core Vocabularies handbook: Using horizontal data standards for promoting interoperability. https://doi.org/10.2799/97439

[21] Osborne SP, Radnor Z, Nasi G (2013) A New Theory for Public Service Management? Toward a (Public) Service-Dominant Approach. American Review of Public Administration, 43 (2):135-158. https://doi.org/10.1177/0275074012466935

[22] Galup SD, Dattero R, Quan JJ, Conger S (2009) An overview of IT service management. Communications of the ACM, 52(5):124-127. https://doi.org/10.1145/1506409.1506439

[23] Bizer C, Heath T, Berners-Lee T (2009) Linked data-the story so far. International journal on Semantic Web and Information Systems 5(3):1-22. https://doi.org/10.4018/jswis.2009081901

[24] Ding L, Peristeras V, Hausenblas M (2012) Linked open government data. IEEE Intelligent Systems. 27(3):11–15. https://doi.org/10.1109/MIS.2012.56

[25] Mouzakitis, S., Papaspyros, D., Petychakis, M. et al. (2017) Challenges and opportunities in renovating public sector information by enabling linked data and analytics. Inf Syst Front 19, 321–336. https://doi.org/10.1007/s10796-016-9687-1

[26] Dimou A, Heyvaert P, Taelman R, Verborgh R (2017) Modeling, Generating, and Publishing Knowledge as Linked Data. P. In: Ciancarini et al. (ed) EKAW Satellite Events, LNAI vol 10180:3-14. Springer, Cham. https://doi.org/10.1007/978-3-319-58694-6 1

[27] Marshall MS et al (2012) Emerging practices for mapping and linking life sciences data using RDF - A case series. Journal of Web Semantics 14:2-13. https://doi.org/10.1016/j.websem.2012.02.003

[28] Sheridan J, Tennison J (2011) Linking UK Government Data. Proceedings of the Linked Data on the Web Workshop, pp.1-4. http://ceur-ws.org/Vol-628/ldow2010_paper14.pdf. Accessed 23 May 2022

[29] European Commission (2016) D04.01 – Core Public Service Vocabulary Application Profile 2.0. ISA Programme. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary-application-profile/distribution/cpsv-ap-specification-v20-pdf. Accessed 23 May 2022

[30] Schmachtenberg M, Bizer C, Paulheim H (2014) Adoption of the linked data best practices in different topical domains. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol 8796. Springer, Cham. pp 245-260. https://doi.org/10.1007/978-3-319-11964-9_16

[31] Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Quarterly. 28(1):75-105. https://doi.org/10.2307/25148625

[32] Hevner AR (2007) A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems 19(2):87-92. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1017&context=sjis. Accessed 23 May 2022

[33] Webster J, Watson R (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly 26(2):13-23. https://www.jstor.org/stable/4132319. Accessed 23 May 2022

[34] Rao SS, Nayak A (2017) LinkED: A Novel Methodology for Publishing Linked Enterprise Data. CIT. Journal of Computing and Information Technology 25(3):191-209 https://doi.org/10.20532/cit.2017.1003477

[35] SEMIC (2017) Generate ADMS asset descriptions from a spreadsheet with Refine RDF. European Commission. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf. Accessed 20 February 2021

Gerontas, A., Tambouris, E., Lazopoulou, K. *et al.* On using CPSV-AP to publish public service descriptions as linked open data. *SOCA* (2022). https://doi.org/10.1007/s11761-022-00344-6

[36] Behkamal B, Kahani M, Paydar S, Dadkhah M, Sekhavaty E (2010) Publishing Persian linked data; challenges and lessons learned. 5th International Symposium on Telecommunications, art. no. 5734119, pp. 732-737. https://doi.org/10.1109/ISTEL.2010.5734119

[37] European Commission (2012) D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. ISA Programme. https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf. Accessed 23 May 2022

[38] European Commission (2014) D04.02.01.B: Common approach for the management of persistent URIs by EU institutions. Towards a common policy for the governance and management of persistent URIs by EU institutions. ISA Programme. https://joinup.ec.europa.eu/sites/default/files/inline-files/Towards-a-common-approach-for-the-management-of-persistent-URIs-by-EU-institutions_v.0.53.pdf. Accessed 23 May 2022

[39] Klein E, Gschwend A, Neuroni AC (2016) Towards a Linked Data Publishing Methodology. International Conference for E-Democracy and Open Government. https://doi.org/10.1109/CeDEM.2016.12

[40] Bosse J., Burnett M., Nielsen S.M., Rongione C., and Scholtens H. (2015) European Public Sector Award 2015: The Public Sector as Partner for a Better Society. European Institute of Public Administration. http://www.epsa2015.eu/files/site/EPSA2015_Publication_updated.pdf. Accessed 25 May 2022

[41] Tambouris E, Tarabanis K (2008) Understanding and scoping life events. International Journal of Electronic Governance 1(2):139-154. https://doi.org/10.1504/IJEG.2008.017901

[42] European Commission (2019) D02.02 – Core Public Service Vocabulary Application Profile 2.2.1. ISA^2 Programme. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary-application-profile/distribution/cpsv-ap-specification-v221-pdf. Accessed 23 May 2022

[43] Stamatis A, Gerontas A, Dasyras A, Tambouris E (2020) Using chatbots and life events to provide public service information. International Conference on Theory and Practice of Electronic Governance (ICEGOV), Athens, Greece. https://doi.org/10.1145/3428502.3428509

[44] Stamatis A, Gerontas A, Tambouris E (2019) On using Chatbots and CPSV-AP for Public Service Provision. Joint conference EGOV-CeDEM-EPART, San Benedetto Del Tronto, Italy. https://biblio.ugent.be/publication/8626904/file/8626906.pdf#page=147. Accessed 23 May 2022