 doi:10.20944/preprints202205.0344.v1

Article

Transforming Points of Single Contact Data into Linked Data

Pavlina Fragkou [1,†,‡], **Leandros Maglaras** [2,‡]

[1] General Secretariat of Information Systems, Ministry of Digital Governance, Athens Greece; e-mail: pavlinafragkou@gmail.com
[2] School of Computer Science and Informatics, De Montfort University, Leicester, UK; e-mail: leandros.maglaras@dmu.ac.uk
* Correspondence: leandros.maglaras@dmu.ac.uk; L.M.
† Current address: General Secretariat of Information Systems, Ministry of Digital Governance, Athens Greece
‡ These authors contributed equally to this work.

Abstract: In this paper, we present a method to map information regarding service activity provision residing in governmental portals across European Commission. In order to perform this, we used as a basis the enriched Greek e-GIF ontology, modeling concepts, and relations in one of the two data portals (i.e., Points of Single Contacts) examined, since relevant information on the second was not provided. Mapping consisted in transforming information appearing in governmental portals in RDF format (i.e., as Linked data), in order to be easily exchangeable. Mapping proved a tedious task, since description on how information is modeled in the second Point of Single Contact is not provided and must be extracted in a manual manner.

Keywords: Linked (open) Data; Semantic Interoperability; Data Mapping; Governmental Data; SPARQL; Ontologies.)

1. Introduction

Open government data is a valuable resource of information addressed to a significant number of recipients. However, this information is usually published in raw format i.e., without following specific guidelines. In addition to this, it does not contain some encoding that allows minimum data linkage such as json-ld thus, it remains unexploited. Moreover, information residing in governmental portals is modeled in a disparate way, even though it serves a common purpose. A characteristic example is the information residing in governmental portals for Directive 123/2006/EC (Directive 2006/123/EC) purposes. This hampers information exchange, which can be feasible using linked data technologies. Linked data technologies, aim at transforming data published in web sites into a machine-readable format (usually RDF using URIs) in order for them to be linked to other external datasets.

More specifically, Directive 123/2006/EC launched in 2006 focuses on simplifying the procedure of practicing a profession by a European citizen in another member state. Each national portal must contain information regarding required supporting documents for each service activity (for example Operation License to Tour Guide's) in two languages, the official of the member state and English. However, no linkage exists between relevant information inside each website, let alone linkage between websites supported by each member state. In other words, no linkage exists between "Cross border provision of services for tourism businesses" or " Tourist offices' notification of commencement of business" appearing in the corresponding Greek portal and "Tour operator and travel agency services" or "Tour guide services" appearing in the Hungarian one, in the case where a European citizen searches for related information.

The goal of our research is to examine how information residing in various Point of Single Contacts (PSCs) can be exchanged using RDF representation in order to be linked. This has as a prerequisite the transformation of this information into RDF triples. The

 doi:10.20944/preprints202205.0344.v1

transformation of this information is based on the data (or ontological) model of each Point of Single Contact (PSC), if and only if it exists. Information mapping thus results in mapping data models. In the case where, like the one examined in this paper, no data model exists for some Point of Single Contacts (PSCs), then existing model(s) are used to perform this information mapping. Current work uses an already existing tool presented in a previous work [1], which transforms data residing in the Greek PSC portal (http://www.eu-go.gr/sdportal/) into RDF format using as a basis the Enriched Greek e-GIF ontology [2]. The Enriched Greek E-GIF ontology is a two-layer ontology, which aims to capture and link all knowledge elements that are essential to describe service activities provided to citizens or businesses.

It should be stressed that, information residing in the Greek PSC portal is modeled based on the Enriched Greek E-GIF ontology. Since we made the assumption to use as a base model for our comparison the enriched Greek E-gif ontology, we have decided to expand the work performed in [1] in comparing information residing in Greek PSC with the one residing in other PSCs, where information on which data model is used is missing.

Current work aims to reveal the difficulty of exchanging information residing in portals related to actions originated by the European Commission, aiming thus at ameliorating citizen's everyday life. An organization can use the richness of the RDF model to capture the detailed relationships in their data and share that in multiple ways. This can only be accomplished if the information is semantically modeled in the same way (i.e., following a unique semantic data model which is designed in advance) and is transformed into rdf or json-ld format. The benefit to e-government will be the possibility to provide cross border electronic services to citizens and businesses and to exchange data between member states. The importance of linking governmental data becomes more and more crucial in view of the forthcoming Single Digital Gateway Regulation, which exploits and reuses - among others- information residing in Point of Single Contacts (PSCs). More specifically, Single Digital Regulation states that "A number of Union acts have aimed to provide solutions by creating sectorial one-stop shops, including points of single contact established by Directive 2006/123/EC of the European Parliament and of the Council, which offer online information, assistance services and access to procedures relevant for the provision of services."

More specifically, our work tries to bring out the steps required to transform information residing in two national portals containing information about Directive 123/2006/EC into linked data. Steps performed reveal that, not only such an approach requires a significant effort proving to be a tedious task, but it does not reach the goal due to a lack of uniform semantic representation (a well-adopted semantic model designed in advance) of the information residing in portals.

The structure of the paper is as follows. Section 2 provides an overview of related methods. Section 3 presents the steps performed for the creation of rdf triples in the pages originated from the Greek and the Hungarian Point of Single Contact portal. Section 4 provides a description of our experiments. Section 5 provides conclusions and future steps.

2. Related Work

Identifying the need for a standard public sector model, European Commission (EC) launched in 2012 the Core Public Service Vocabulary (CPSV) initiative in the framework of ISA and its successor ISA2 (Interoperability solutions for public administrations, businesses and citizens) programmes The Core Public Service Vocabulary (CSPV) is a simplified, reusable and extensible data model that captures the fundamental characteristics of a service activity offered by public administrations. By the term Public Service we mean service provision activities which are provided either in conventional or in electronic means by the public authorities to citizens and businesses. It allows public administrations to describe their service activities in a unified way and make these descriptions re-used on many governmental access portals. The vocabulary enables the seamless exchange of services and information across different e-Government systems. It should be noted

 doi:10.20944/preprints202205.0344.v1

that CPSV was published after Directive 123/2006/EC was launched. As a consequence information residing in every national portal is represented in a different way compared to the other ones. At the end of 2016, EC announced the final version of CPSV-AP 2.0, which is an application profile of CPSV incorporating Linked Data as underpinning technology. The latest specification of this vocabulary manages to find information regarding service provision activities more easily due to its enrichment with additional classes.

In [3] both the definition and verification of CPSV-AP v.1.1 were performed. Moreover, the paper describes the CPSV-AP's mapping to the data models used in the Point of Single Contacts (PSCs) of ten Member States (MS). The mapping of the PSC data models to the CPSV-AP was based on the review and analysis of the way information is provided on the PSC on the one hand, and the CPSV-AP on the other hand. The authors paid special attention to the Greek data model i.e., the Greek E-Gif ontology (http://www.e-gif.gov.gr/portal/page/portal/egif) taken as a basis for our work. The authors claim that mapping PSC data models with CPSV-AP revealed some interesting results. Firstly, the national data models describing Business Events and Public Services do not cover the whole set of classes or properties defined in the CPSV-AP. These electronic PSCs are currently facing several challenges, such as: (a) Lack of coordination between the electronic PSCs within the same country; (b) Heterogeneous, monolingual, descriptions of public services and business events; (c) National vs. cross-border public service provision. Nevertheless, all national data models had additional concepts (mostly properties but sometimes also classes for modeling, for example, required steps as well as prerequisite evidence documents) not defined in the CPSV-AP. This is attributed to the fact that PSCs want to provide detailed information to users, and it cannot be considered as a deficiency of the CPSV-AP. Secondly, this mapping exercise led to a few updates of CPSV-AP. One such important update is the connection of the Public Service and the Formal framework classes, the specification of the Channel class to Email, Homepage, Fax, Telephone, and Assistant as well as the Cost class.

The fact that national PSCs want to provide detailed information to users, which is not included in CPSV-AP was addressed in the Enriched Greek E-Gif ontology [2]. The later ontology contains entities that have more practical sense such as "Channel" i.e. different ways that the service activity as well as information regarding the person/organization to whom/which one can address to for further details is provided. The Enriched Greek E-Gif ontology also focuses on specific entities such as: (a) Document in the sense that it exposes prerequisite documents for service activity provision; (b) Service and more specifically its description; (c) Necessary Steps explaining in details the steps the applicant should perform as well as (d) Prerequisites conditions that the applicant must satisfy.

In [4], the authors present an approach for using CPSV-AP 2.0 released in 2016 to model and publish, as Linked Data, public services descriptions of a regional public service catalog. 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 Citizen's Guide of the Region of Epirus is a structured catalogue of descriptions of public services provided by the Region of Epirus, in Greece. The authors used CPSV-AP 2.0 tools and exported data residing in Citizen's guide of the Region of Epirus after performing a semi-automated process. The work performed there is extended in [5] where the authors try to publish RDF schema of CPSV-AP 2.0. The authors claim to publish 45 public services in an RDF store, without however providing further details. CPSV-AP is also used in [6] to denote the semantics to be used for the descriptions of public services by a chatbot application. In that work, public servants annotate public services using the semantics of CPSV-AP to create RDF triples, which are - at a second stage - transformed in JSON format, since the chatbot requires JSON files in order to analyze real-time data and make decisions.

Another effort to the direction of modeling concepts and relations used in a governmental portal was presented in [1]. Most specifically, the focus of that paper was on the exploitation of the Enriched Greek E-Gif ontology using a tool for semantic web applications which is able to transform semantic relations of an ontology into Resource Description Framework (RDF) relationships. The tool not only provides the capability to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

store, manage and manipulate RDF relations but also to pose queries using SPARQL, an RDF query language. The contribution of the work lies in the fact that, this tool can be incorporated into the Greek Point of Single Contact (PSC) portal, in order to transform static information appearing in it into linked open data that can be semantically queried. It is worth mentioning that, CPSV-AP 2.0 and Enriched Greek E-Gif ontology have many commonalities, however the Enriched Greek E-Gif ontology contains more attributes and manages to describe public services in a more detailed way.

The National Registry of Services (codename "Diavlos") has been developed for the Greek Ministry of Digital Governance on a national level and is already registering services using the CPSV-AP data mode. It publishes information about Public Services for citizens and businesses as well as internal procedures of public bodies. Currently, a JSON description of each process is provided as well as an API for accessing service descriptions (http://api.reg-diavlos.gov.gr:5000/v1/ui/) while the next step is to develop a module that converts it to RDF/XML. Information about services is provided in Greek, however the English version of Diavlos which is under development, will allow service descriptions to be interoperable in European and international level Adoption of other Core public service vocabularies (e.g. Core Public Organization Vocabulary for the development of an enhanced (in information) version of the Registry of Public Organizations within Diavlos). It is foreseen the newer Greek Point of Single Contact (PSC EUGO), will be based on the already available information in the National Register of Procedures 'Diavlos'.

It must be stressed that CPSV-AP consists a Core Vocabulary i.e., is generic enough by definition in order to be used as a starting point for specific domain cases. The most closely related approach from a semantically point of view is the Catalogue of Services Vocabulary (https://github.com/catalogue-of-services-isa/CPSV-AP) developed under the scope of Single Digital Gateway Regulation. It tries to capture all important information requirements from the Single Digital Gateway Regulation (SDGR) (https://github.com/catalogue-of-services-isa/SDG-services-model/wiki/SDG-services-model-Home-Page, github.com/catalogue-of-services-isa/SDG-services-model/tree/master/SDG%20concepts). More specifically, its goal is to make information easily findable through the YourEurope portal in order to help citizens and businesses to find and get the information they need at a glance, while minimizing the efforts from Member States to provide this information. In this regard, they have identified a first set of the most important information concepts from the Regulation, which have a lot of common attributes with the ones appearing in the enriched Greek e-Gif ontology, since the domain in question is very similar. This is due to the fact that, Single Digital Gateway Regulation is based – among others- on the 123/2006 Services Directive which forms the basis of the enriched Greek e-Gif ontology. However, due to the fact that the SDGR data model has not yet been finalized and it is still under creation couldn't be used for the problem in question. Moreover, no comparison can be made since no data portal exists based on the SDGR data model.

A similar approach from an architectural point of view but on a different problem was presented in [7] where an automatic semantic migration prototype based on Knowledge Discovery from Digital Archive Data for ontology population in the domain of Archives metadata, ISAD(G) was performed. Natural Language Processing (NLP) techniques for language processing and Semantic Web techniques for querying and updating the Ontology ArchOnto, a CIDOC-CRM (Conceptual Reference Model) extension were used. In the same vein and with many common architectural modules is the work presented in [8] where the aim is the creation of Linked Open Saudi Government Data Framework (LOSGDF). More specifically, the purpose is to collect, process, generate RDF datasets, interlink these datasets with other open datasets and store them into a triple store server taking as unlinked structured data. Linking open datasets involves investigating relations between Saudi Arabian government organizations (between Open Government datasets having different data sources such as CSV files, Excel sheets, online portals, and structured documents) supervised by certain ministries, establishing and then publishing them to their open governmental data portal.

 doi:10.20944/preprints202205.0344.v1

Lastly, it is worth noticing that, an ontology that captures impartially the notion of service provision is the European Skills, Competences, qualifications and Occupations (ESCO) ontology . This ontology consists of a central building block for an ecosystem of semantic assets on the labor market offering a multilingual classification of European Skills, Competences, Qualifications and Occupations to facilitate EU job market. ESCO is a hub thesaurus for mapping like classification system about either of - occupation - skills and competence - qualification (certification). It provides semantic relationships between concepts i.e. occupation - skills and competence - qualification. In version 1, occupations were tagged with one or more NACE codes It is worth mentioning that the Greek Point of Single Contact uses NACE codes to classify service provision information.

3. Methodology

As mentioned previously, current work extends the work performed in [1] i.e. where a limited number of pages residing in the Greek PSC portal were transformed into RDF format. Additionally, web page transformation was restricted to the ones residing in it, without taking into consideration how such a transformation can take place in pages residing in other PSC's having as a basis the Enriched Greek E-gif ontology. Current paper tries through data mapping between information residing in two PSCs, to prove what it was expected, i.e., that this mapping is a tedious task. Already existing work was focused on transforming information residing in the Greek PSC into rdf triples using the Enriched Greek E-gif ontology. Current work tries to map information residing in the Hungarian PSC to the Enriched Greek E-gif as well as to create rdf triples, aiming at being able to connect related information.

It must be stressed that CPSV-AP ontology does not include information such as documents, prerequisites as well required steps in order to exert a service provision activity i.e., profession, appearing in a number of Points of Single Contacts (PSCs).

Even though CPSV-AP ontology could have been designed in order to depict information found in various PSCs, making thus retrieval of information contained in them in a more effective way, this is not the case. The aforementioned situation is aggravated by the fact that information on how data is structured in various PSCs is agnostic due to the lack of a publicly available ontology describing each of them. This leads to take solely the Enriched Greek E-gif ontology as a base to compare how information is structured in various PSCs. Subsections that follow, explain in detail steps performed.

3.1. Dataset

During current work, the intention is to extend work performed in [1] to apply the toolchain to more web pages appearing in the Greek PSC, by also to expand RDF triples produced using additional Greek e-gif ontology's properties. This is the reason why, as a first step, all pages appearing in the Greek PSC were downloaded. This resulted in a 411 unique files containing descriptions of public service activities. Among them, 325 are written in Greek and 86 in English. In order to perform comparisons with information residing in other PSC's we focused on the 86 web pages written in English.

As a second step, we expanded the set of attributes extracted from each web page to form rdf triples and consequently those used for performing query search in the produced rdf triples. More specifically, during our current experiments for every service provision activity (i.e. profession) we focus on the following properties: title, provision method (i.e., establishment or cross border), NACE code classification, required time, cost, responsible public body, legal framework, required documents as well as comments. Each of the aforementioned properties is related to an equivalent relation i.e. predicate such as hasTitle, hasMethod etc, leading thus to a number of rdf triples for every web page.

A third step is our intention to expand our experiments to other Point of Single Contacts (PSCs). Since we made the assumption to have as a reference of comparison between PSCs belonging to different countries the Enriched Greek E-gif ontology, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

purpose was to try to extract as much as valuable information as possible i.e. as much as many attribute instances as those defined in it.

In order to examine whether the Enriched Greek E-gif ontology can capture information residing in other PSCs, we examine all PSCs. Our main prerequisite was the provision of information in another language apart from the country's native language, i.e., provision of information found in English.

A second criterion was whether each PSC under consideration contained similar professions with the ones described in the Greek PSC. A third criterion was how information was organized within a web page i.e., whether it contained information such as cost, time, legal framework, prerequisite documents, comments and public body. Another criterion taken under consideration was the actual html code of each web page, in the sense of performing an effective html parsing.

Among all PSCs, we restricted our search to the ones of Malta, Slovenia, Hungary and Cyprus. However, Hungarian PSC was the one to fulfill most of the criteria listed above. More specifically, in Hungarian PSC information is organized in a similar manner i.e., it provides information regarding the title, the administrative cost, the processing time limit, the list of required documents and the legal framework. On the other hand, it does not contain useful information such as NACE code. A significant difference between the Hungarian and the Greek PSC is the fact that, the first contains information regarding establishment as well as cross border provision into a single web page, while in the second the information resides into two distinct web pages since the legal framework (thus the prerequisite documents) for them might be different. Another point to note is that, in the Hungarian PSC there is not a unique web page layout, i.e., all pages do not present information in a unified way. This makes web page parsing more difficult. As a consequence, we extracted 15 out of 34 web pages containing information represented in a unified manner and describing professions that are similar to the ones found in the Greek PSC. Properties that appear to the fifteen pages of the Hungarian PSC that are in alignment with the Enriched Greek E-Gif ontology are: ServiceCost, LegalFramework, ServiceComment, Input (which represents prerequisite documents), Title and DeliveryTime.

3.2. Preprocessing Steps

As previously mentioned, we used the toolchain implemented in [1]. This toolchain involves the following steps

Step 1. HTML web page parsing - Web page retrieval: Even though information hosted in the Greek PSC portal is publicly available, no interventions, i.e. addition of RDF tags and annotations are permitted, leading us to download a copy for each page. For each web page of the dataset in question, a pre-processing step i.e., html parsing was performed using the JSOUP library (https://jsoup.org/) for extracting properties of interest, i.e., the content of specific fields and store the extracted values. JSOUP is an open source Java library used mainly for extracting data from HTML using DOM traversal or CSS selectors. It also allows to manipulate HTML elements, attributes, and text and out-put tidy HTML.

Step 2. Creation of the RDF model from data produced by the previous step: After downloading and parsing web pages belonging to the dataset in question, the creation of the RDF model was performed using the Enriched Greek E-Gif ontology. More specifically, taking as input the aforementioned ontology we created a model. As a result, class names were produced i.e., a namespace which consists of a prerequisite for searching every class. Predicates are produced by invoking appropriate methods which take as argument every property of interest. In the RDF model that was created, a node corresponds to a class appearing in the Enriched Greek E-Gif ontology. More specifically, each service provision activity i.e., the profession is represented as a node (i.e., a subject in the RDF model) and is connected either with other nodes or strings through appropriate predicates i.e., edges forming RDF triples. Each edge corresponds to an ontology property.

For the creation of the RDF model, Apache JENA Framework was used (https://jena.apache.org). Apache JENA is recognized by the W3C as a Semantic Web Standard, for creating semantic

 doi:10.20944/preprints202205.0344.v1

Table 1. Statistics on the number of pages, nodes and RDF triples extracted from the Greek and the Hungarian PSC.

Dataset	Number of Pages	Number of Nodes
Greek PSC	86	1683
Hungarian PSC	15	292

applications. Apache JENA provides an API to extract data from and write to RDF graphs. The graphs are represented as an abstract "model". A model can be sourced with data from files, databases, URLs or a combination of these. A model can also be queried through SPARQL. Apache JENA provides mapping correlation of names between those appearing in an already defined ontology (in our case the Enriched Greek E-GIF ontology) with those appearing in an RDF model generated by data samples.

Step 3. Creation of RDF Triples: A subsequent step deals with the creation of the required triples using the information we derived through the application of the parser as well as class and property names derived from the ontology (steps 1 and 2). The Apache RDF API uses the notion of the namespace, which ensures the uniqueness of each entity of the model as well as the model itself in the semantic web. Apache JENA requires the use of namespace prefix in the model's name, as well as in each entity contained in the model (i.e., subject, predicate, and object). Once the RDF model (as well as the equivalent namespace NS) was created, RFD triples corresponding to web pages of the dataset in question can be created. Apache JENA RDF API was used in order to produce the triples based on information exported by the parser when it was applied to the dataset. Nodes are created resulting in the construction of triples through the interconnection of nodes. All triples are created and added to the model in a similar manner.

Step 4. Storage of the model into a Triple Database Store: A subsequent step was model storage, accomplished using the Triple Database Store API. An RDF model is stored via TDB into a recognizable format by Apache JENA API but also by dedicated servers such as FUSEKI. The Triple Database API provides a logical unit storage of the model through which the information is retrieved using SPARQL. It uses a simple and quick way of encoding nodes of RDF relations.

Step 5. Retrieval of semantic information: As long as the model is stored in an RDF format, retrieval of desired information can take place via SPARQL queries using the Apache JENA API. Queries can be expressed using the java language by invoking specific methods provided by the JENA API. Apache JENA supports its own API that supports the SPARQL RDF Query language (SPARQL Query and RDF Query Language). SPARQL is an information retrieval language from a database structured on the RDF model. More specifically, SPARQL seeks information that meets specific criteria, but giving a description of the type of ternary relations to which desired data be-longs to. For SPARQL, namespace definition is necessary. The FUSEKI server serves requests in the form of SPARQL queries. FUSEKI is a SPARQL server that provides HTTP endpoints to RDF data. Table 1 provides some statistics regarding the number of pages, the number of nodes is well as the number of triples resulted after applying the toolchain in pages (written in English) belonging to the Greek as well as the Hungarian PSC.

A graphical representation of the methodology followed appears in Figure 1.

4. Experiments

The purpose of our experiments is to examine the adequacy of Enriched Greek e-Gif ontology in representing knowledge residing to other PSCs. This can be explored by performing queries in both datasets, i.e., in the rdf triples produced after applying the toolchain presented in the previous section in web pages belonging to Greek (86 web pages) and Hungarian PSC (15 web pages).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

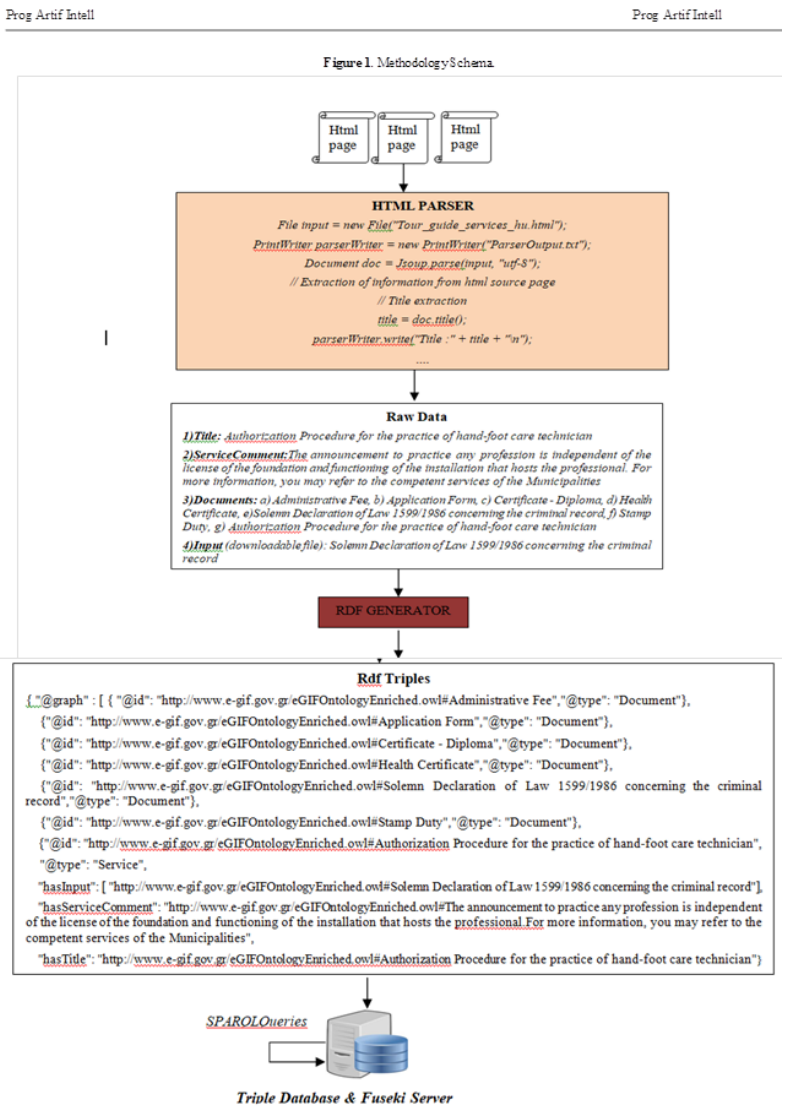

Figure 1. Methodology Schema.

Manual comparison of the fifteen service provision activities selected from the Hungarian PSC to the 86 corresponding ones appearing into the Greek PSC based on their title, resulted to the fact that only 7 of them are common in both PSCs.

Practically, two distinct datasets were created one for pages residing in every PSC. The aim was to run queries on both datasets based on keywords on the triples created (based on predicates). For comparison reasons, we performed the same queries - as the one executed in both datasets - in the actual PSCs web pages using the search option provided by each of them. Most specifically, in those queries we used as keywords words appearing in each service provision activity's title, comment and document (i.e., input). However, we are not aware of how the search engine in each PSC works, i.e., in which part(s) of the information of every profession is based on. The aforementioned search was selected since no other comparison can be performed.

The rationale behind this was to try to simulate how an end-user will behave. This is the reason why text search was introduced. The result of those queries is strongly related to the produced rdf triples, which in its turn is related to the level of data mapping accomplished. This, by no means, leverages the typical strengths of having the information as Linked data, which is exactly what current research tries to highlight. For our queries, we used as basis (starting point) information residing in the fifteen pages of the Hungarian PSC site. More specifically, our queries were mainly based on the predicates

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

Table 2. Examples of queries and matches returned by the Greek and the Hungarian PSC.

Query	Matches in the Greek PSC	Matches in the Hungarian PSC
PREFIX OntologyPrefix: <http://www.e-gif.gov.gr /eGIFOntologyEnriched.owl#> SELECT * WHERE { ?Service OntologyPrefix:hasTitle ?Title; filter contains (?Title, "veteri") }	Establishment - Authorized to perform artificial insemination and eustrus synchronization program in breeding animals	Practicing as a private veterinarian
	Establishment - Licensed establishment and operation of veterinary office for livestock	
	Establishment - Authorization for establishment and operation of ungulates and poultry slaughterhouses	
PREFIX OntologyPrefix: <http://www.e-gif.gov.gr /eGIFOntologyEnriched.owl#> SELECT * WHERE { ?Service OntologyPrefix:hasServiceComment ?Comment; filter contains (?Comment, "construct") }	Establishment - Granting of license to practice the profession of Degree holder EngineerJun	Statement of eligibility to practice as architectural design engineer
	Cross Border - Declaration of temporary and occasional provision of Services of Architect Engineer	

hasServiceComment, hasTitle and hasInput. We restricted our search on those predicates since we believe that, they contain the most valuable information compared to hasServiceCost, hasLegalFramework or hasDeliveryTime. The search was performed by extracting potential keywords from each of the aforementioned predicates of each of the 15 service activities of the Hungarian PSC and used them to query the 86 web pages of the Greek PSC (written in English). Performing our queries we observed that, case sensitivity produces different results. Case sensitivity has a significant impact on the number of obtained results. Queries performed using keywords appearing in the hasTitle, hasServiceComment as well as hasDocument resulted in the outcomes described in the following paragraph.

More specifically, we managed to find matches in all fifteen service activities examined from the Hungarian PSC when we used keywords taken from service provision activity (profession)'s title. Matching also varies according to the service provision activity examined. More specifically, we found matches when we expanded our keywords taking information from hasServiceComment and hasDocument predicates for the following service provision activities: 1) activity directed at the organization of vocational examination, 2) the attestation activity of contributing entity involved in the preliminary vehicle identity check, 3) practicing as a private veterinarian, 4) the registration of economic organizations and their shops engaged in the trade of precious metal, jewelry, articles and ornaments, 5) statement of eligibility to practice architectural design engineer, 6) statement of eligibility to practice as construction engineer inspector. Those service provision activities coincident with the seven found as common in both PSCs during our manual search.

Table 2 provides examples of the matches returned by querying both PSCs showing how queries are expressed in SPARQL, while Table 3 provides all queries conducted with all variations of keywords as well as matching results.

From the obtained results we reach the following conclusions: while hasDocument turns out to be the least useful predicate, the most valuable ones turns out to be the hasTitle and hasComment predicates. On the other hand, hasService predicate is not as reliable as hasTitle, since it returns non-relevant documents as matches. It must be stressed that, results obtained after performing the sparql queries using specific keywords, as well as performing on site queries in both PSCs are very similar. This means that, the selection of the keywords for search was close to the one used by the search engine implemented in

 doi:10.20944/preprints202205.0344.v1

both sites, even though, in both PSCs and for every service provision activity (profession)'s description, keyword information is missing.

Obtained results are in alignment with results obtained when performing the same queries on the search field in each portal. It must be stressed that, search functionality in Greek PSC is case sensitive. Moreover, since we are not aware of which keywords are used to describe each page in the search functionality, each query performed in the Greek PSC portal, resulted in more than one page, some of them being incorrect. However, the correct one(s) always appeared in the query's outcome.

5. Conclusion

In this paper, we have presented work concerning the transformation of publicly available information regarding service activity provision enforced by Directive 123/2006/EC to Linked Open Data. The purpose of performing those experiments was triggered by two factors: the first one was the fact that data published as open by public administrations remains poorly exploited (which was not the intention of service directive i.e., Directive 123/2006). The second factor is that the forthcoming Single Digital Gateway Regulation i.e., Regulation (EU) 2018/1724 is going to be based – among others – to the outcome of Service Directive. At the current moment, Single Digital Gateway Regulation focuses on 21 procedures (http://data.europa.eu/eli/reg/2018/1724/oj). However, the number of procedures might rise after 2023, when the Regulation will put into action. Thus, the focus was not on jobs' details across PSCs but on common information residing in PSCs.

From the aforementioned experiments, the main observation is the lack of uniform semantic representation of information residing in national PSC's (i.e., lack of definition beforehand and/or lack of conformity to a unique semantic model), makes data mapping tedious to achieve. This must be taken under consideration in forthcoming actions such as the Single Digital Gateway directive, whose purpose is - among others- to provide information on online and offline procedures and links to online procedures, established at European Union or national level via a common user interface, which shall be accessible in all official languages of the European Union.

Information residing in PSCs can be addressed to a number of recipients, i.e., European citizens. Ideally, in order to be better exploitable, this information must follow a uniform semantic representation such as json-ld, containing annotations based on a common semantic representation i.e. ontology. ESCO ontology might be used as a starting point for this purpose, since it contains a related class i.e., occupation. A combination of more than one ontologies, vocabularies or models such as CPSV-AP 2.0, ESCO as well as the Core Criterion and Core Evidence vocabulary (https://joinup.ec.europa.eu/release/core-criterion-and-core-evidence-vocabulary-v100) or ISA2 Core Vocabularies (Core Person, Core Organization and Core Business, https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/core-vocabularies) can also prove to be beneficial.

Our vision for future work consists of three steps. The first step aims to create appropriate links i.e., URI's between pages belonging to our core dataset, taking under consideration that for a number of pages their translated version in English web page exists. The second step involves the interconnection of our dataset with the equivalent ones found in the dedicated sites of other European countries i.e. the creation of interconnection links between sites focusing on the English content of each site. In order to perform this, DBpedia ontology (http://www4.wiwiss.fu-berlin.de/dbpedia/dev/ontology.htm) can be considered as a common ontology to use. We are also considering the creation of a tool in order to transform information residing in Greek PSC into JSON-LD, as an upper layer, both using the Enriched Greek E-gif ontology as well as CSPV-AP in order to be used by other PSC's.

6. Conclusions

This section is not mandatory, but can be added to the manuscript if the discussion is unusually long or complex.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

Table 3. Queries and matches returned by the Greek and the Hungarian PSC, keywords used and attributes in which queries were applied.

Keyword used in the query	Attribute in which the query was applied	Matches in the Hungarian PSC Dataset	Matched profession(s)	Matches in the Greek PSC Dataset	Matched profession(s)
accounting or Accounting	hasTitle	1	Accounting services	0	
Vocational or vocational	hasTitle	1	Activity directed at the organisation of vocational examinations	1	Issuance of a permit to establish and operate a Vocational Training School for Hazardous Goods Drivers (SEKOOMEE)
vehicle	hasTitle	1	The attestation activity of the contributing entity involved in the preliminary vehicle identity check	2	1) Accreditation of Test Facilities for ADR testing for hazardous goods vehicles 2) Licensing for establishment and operation of Driving School (and its branch) of candidate motor-vehicles and motorcycles drivers
Vehicle	hasComment	0		3	1) Issuance of a permit to establish and operate a Private Vehicle Technical Control Centre (IKTEO) 2) Accreditation of Test Facilities for ADR testing for hazardous goods vehicles 3) Licensing for establishment and operation of Driving School (and its branch) of candidate motor-vehicles and motorcycles drivers
Vehicle	hasTitle	0		1	Issuance of a permit to establish and operate a Private Vehicle Technical Control Centre (IKTEO)
condomi	hasTitle	1	Commercial condominium management	0	
facility	hasComment	0		1 incorrect	Accreditation of Test Facilities for ADR testing for hazardous goods vehicles
facility	hasTitle	1	Commercial facility management activity	0	
Commercial	hasTitle	2	1) Commercial facility management activity 2) Commercial condominium management	0	
Funeral	hasTitle	1	Funeral services	0	
procure	hasTitle	1	Official public procurement consultation activities	0	
consult	hasTitle	1	Official public procurement consultation activities	0	
estate	hasTitle	1	Performance of real estate brokerage, real estate asset appraisal and intermediary activities	0	
veteri	hasComment	0		3	1) Authorized to perform artificial insemination and eustrus synchronization program in breeding animals 2) Licensed establishment and operation of veterinary office for livestock 3) Authorization for establishment and operation of ungulates and poultry slaughterhouses
veteri	hasTitle	1	Practicing as a private veterinarian	1	Licensed establishment and operation of veterinary office for livestock

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 May 2022 doi:10.20944/preprints202205.0344.v1

Table 4. Queries and matches returned by the Greek and the Hungarian PSC, keywords used and attributes in which queries were applied (Cont.).

Keyword used in the query	Attribute in which the query was applied	Matches in the Hungarian PSC Dataset	Matched profession(s)	Matches in the Greek PSC Dataset	Matched profession(s)
placement	hasTitle	1	Private employment placement activities	0	
property	hasTitle	1	Property management land surveyor activity	0	
jewel	hasTitle	1	The registration of economic organizations and their shops engaged in the trade of precious metal jewellery, articles and ornaments	1	Exercise of the profession of money-changers, second-hand dealers, pawnbrokers and of those involved in the purchase and sale or the casting of used gold jewellery and gold artistic objects and other valuables
metal	hasTitle	1	The registration of economic organizations and their shops engaged in the trade of precious metal jewellery, articles and ornaments	0	
architect	hasComment	0		2	1) Declaration of temporary and occasional provision of Services of Architect Engineer 2) Granting of license to practice the profession of Degree holder Engineer
architect	hasTitle	1	Statement of eligibility to practice as architectural design engineer	1	Declaration of temporary and occasional provision of Services of Architect Engineer
construct	hasTitle	1	Statement of eligibility to practice as construction engineering inspector	0	
inspect	hasComment	0		2 incorrect	1) Approval of the activity of Inspection and Certification Body for organic products 2) Recognition of Private Bodies certifying Agricultural Products and Systems
inspect	hasTitle	1	Statement of eligibility to practice as construction engineering inspector	1 incorrect	Approval of the activity of Inspection and Certification Body for organic products
Inspect	hasTitle	1	Statement of eligibility to practice as construction engineering inspector	1 incorrect	Approval of the activity of Inspection and Certification Body for organic products

 doi:10.20944/preprints202205.0344.v1

Table 5. Queries and matches returned by the Greek and the Hungarian PSC, keywords used and attributes in which queries were applied (Cont.).

Keyword used in the query	Attribute in which the query was applied	Matches in the Hungarian PSC Dataset	Matched profession(s)	Matches in the Greek PSC Dataset	Matched profession(s)
enginne	hasComment or has Title	0		20	Declaration of temporary and occasional provision of Services of: 1)Mechanical Engineer 2)Mechanical Engineer and Aircraft Builder Engineer 3)Architect Engineer 4)Agronomist and Topographer Engineer 5)Civil Engineer 6)Electrical Engineer and Computer Engineer 7)Electrical Engineer and Computer Technology Engineer 8)Shipbuilder Mechanical Engineer 9)Chemical Engineer 10)Mineralogist - Metallurgist Engineer 11)Environment Engineer 12)Production and Management Engineer 13)Mineral Resources Engineer 14)Information and Communication Systems Engineer 15)Zone & Urban Planning and District Development Engineer 16)Zone Planning and Development Engineer 17)Electronic Engineer and Computer Engineer 18)Computer and Information Science Engineer 19)Computer, Telecommunications and Networks Engineer and 20)Granting of license to practice the profession of Degree holder Engineer
Tour or tour or travel	hasComment	0		4	1) Cross border provision of services for tourism businesses 2) Authorizing the entry into service of Tourist Trains in special routes 3) Tourist offices' notification of commencement of business 4) Notification of commencement of Tourism Road Transport Companies
Tour or travel	hasTitle	1	Tour operator and travel agency services	4	1) Cross border provision of services for tourism businesses 2) Authorizing the entry into service of Tourist Trains in special routes 3) Tourist offices' notification of commencement of business 4) Notification of commencement of Tourism Road Transport Companies

Author Contributions: Conceptualization, P.F.; methodology, P.F.; validation, P.F. and L.M.; formal analysis, P.F.; investigation, P.F.; resources, P.F.; data curation, P.F.; writing—original draft preparation, P.F.; writing—review and editing, L.M.; visualization, P.F.; supervision, L.M.; project administration, P.F.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Fragkou, P.; Kritikos, N.; Galiotou, E. Querying greek governmental site using sparql. In Proceedings of the Proceedings of the 20th Pan-Hellenic Conference on Informatics, 2016, pp. 1–6.
2. Fragkou, P.; Galiotou, E.; Matsakas, M. Enriching the e-GIF ontology for an improved application of linking data technologies to greek open government data. *Procedia-Social and Behavioral Sciences* **2014**, *147*, 167–174.
3. Loutas, N.; De Keuzer, M.; Tarabanis, K.; Alvarez-Rodriguez, M.; Burian, P. Harmonising the public service models of the points of single contact using the core public service vocabulary application profile. *Innovation and the Public Sector* **2015**, p. 277.
4. Gerontas, A.; Tambouris, E.; Tarabanis, K. On using the core public sector vocabulary (CPSV) to publish a" citizen's guide" as linked data. In Proceedings of the Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018, pp. 1–6.
5. Gerontas, A.; Peristeras, V.; Tambouris, E.; Kaliva, E.; Magnisalis, I.; Tarabanis, K. Public Service Models: a systematic literature review and synthesis. *IEEE Transactions on Emerging Topics in Computing* **2019**, *9*, 637–648.
6. Karamitsos, I. Chatbots for Greek/EU Public Services **2019**.
7. Melo, D.; Rodrigues, I.P.; Koch, I. Knowledge Discovery from ISAD, Digital Archive Data, into ArchOnto, a CIDOC-CRM based Linked Model. In Proceedings of the KEOD, 2020, pp. 197–204.
8. AlSukhayri, A.M.; Aslam, M.A.; Arafat, S.; Aljohani, N.R. Leveraging the Saudi Linked Open Government Data: A Framework and Potential Benefits. *International Journal of Modern Education & Computer Science* **2019**, *11*.