d2kg-OWL: An Integrated Ontology for Knowledge Graph-based Representation of Government Decisions and Acts

The Greek Programme Diavgeia case

Konstantinos Serderidis [a,*], Ioannis Konstantinidis [b], George Meditskos [a], Vassilios Peristeras [b] and Nick Bassiliades [a]

[a] *School of Informatics, Aristotle University of Thessaloniki, Greece*
E-mails: kserderc@csd.auth.gr, nbassili@csd.auth.gr, gmeditsk@csd.auth.gr
[b] *International Hellenic University, Greece*
E-mails: i.konstantinidis@ihu.edu.gr, v.peristeras@ihu.edu.gr

Abstract. Public Administration is a rich source of data and potentially new knowledge. It is a data intensive sector producing vast amounts of information encoded in government decisions and acts, published nowadays on the World Wide Web. The knowledge shared on the Web is mostly made available via semi-structured documents written in natural language. To exploit this knowledge, technologies such as Natural Language Processing, Information Extraction, Data mining and the Semantic Web could be used, embedding onto documents explicit semantics based on formal knowledge representations such as ontologies. Knowledge representation can be made possible by the deployment of Knowledge Graphs, collections of interlinked representations of entities, events or concepts, based on underlying ontologies.

This paper presents a new ontology, d2kg [d(iavgeia) 2(to) k(nowledge) g(raph)], integrating in a unique way standard EU ontologies, core and controlled vocabularies to enable exploitation of publicly available data from government decisions and acts published on the Greek platform Diavgeia with the aim to facilitate data sharing, re-usability and interoperability. It demonstrates a characteristic example of a Knowledge Graph based representation of government decisions and acts, highlighting its added value to respond to real practical use cases for the promotion of transparency, accountability and public awareness. The proposed d2kg ontology is accessible on "http://lpis.csd.auth.gr/ontologies/2022/d2kg/d2kg.owl"

Keywords: Semantic Web, Linked Open Data, Ontologies, Knowledge Graphs, Government decisions and acts, Diavgeia Programme

1. Introduction

During the last decades there has been a constant effort to bring citizens closer to public policies and to raise their awareness of government programmes and policies so that the civil society becomes more actively engaged, better informed and adequately capable to assess the decision-making bodies and processes.

This effort has been driven by introducing concepts such as "Open Government" aiming at establishing cooperation among the main actors in the public sphere, that is politicians, public administrators, entrepreneurs and

*Corresponding author. E-mail: kserderc@csd.auth.gr.

citizens through enhanced transparency, accountability and participation. To put "Open Government" in effect free access, use and re-use of data and information in general, are essential prerequisites, which makes "Open Government Data" (OGD) a pillar for establishing "Open Government" [1].

Towards the direction of implementing effective (Open) Governance models, a crucial element is the efficient use of the big amounts of data produced in the public domain in order to promote transparency and accountability amongst public actors, as well as to raise awareness amongst citizens. It is evident that public domain data offers a rich source of valuable data with high potential for consumption, sharing and exploitation. In public administration, though, data is made available via inter-linked documents written traditionally in natural language. To this end, the emergence of the World Wide Web has contributed to the production and sharing of vast amounts of data that could be potentially used for creation of new knowledge. Emerging technologies to exploit knowledge, such as Natural Language Processing, Information Extraction, Data Mining and -primarily- the Semantic Web, have made it possible to develop Knowledge-based Management systems. In the domain of interest, Knowledge-based Management systems such as Knowledge Graphs can be established on appropriate underlying ontologies. Nevertheless, it is still the case that interlinking and interoperability of different national public administration data has not been achieved due to apparent issues stemming either from technical requirements, since there is no harmonization at the level of public/governmental documents produced from Member States in the European Union, or simple facts such as that the information is available in different languages.

To fully benefit from "Open-ing" Data, it is crucial to put information and data in a context that creates new knowledge and enables useful services and applications, a major trait of Knowledge Graphs too. Hence, it becomes evident that in order to achieve highest exploitation, it is necessary to move from Open (Government) data to Linked Open (Government) Data [1]. This is acknowledged at institutional level in the European Union via numerous initiatives to exploit the huge amounts of the Public Sector data of high financial value, known as Public Sector Information (PSI), or Government, data [1]. To overcome the limitations of traditional knowledge representation via public administration documents commonly uploaded in low ranked Open Data quality formats such as pdf, we need to deploy Semantic Web technologies, through embedding onto documents explicit semantics based on formal knowledge representations such as ontologies.

This paper proposes a new ontology which integrates standard EU ontologies, core and controlled vocabularies following W3C recommendations to exploit publicly available Open data following Linked Data principles and thus additionally allow a Knowledge Graph based representation of government decisions and acts. The core objective is to enhance public data re-usability and inter-operability at EU level. The focus is in particular on the Greek Programme Diavgeia [2] where Public Organizations are required mandatory by Law to upload government decisions and acts, as a good showcase. It can also serve as a guideline on how standard ontologies and vocabularies could be employed to represent information included in Public Administration documents at EU level, since it is expected that, due to the adoption of core EU legislation via Regulations and Directives in the national law of Members States, a universal -to a certain extent approach- could be supported in the same manner in a cross-border approach.

The paper is structured as follows: In Section 2 related work is presented in a comprehensive manner. Section 3 builds on the main concepts to establish a methodological framework to develop subsequently an ontology in the field. The new integrated OWL ontology is then presented in Section 4. Section 5 presents representative case studies of how to exploit data from Diavgeia documents and produce additional knowledge. In Section 6 we proceed with an assessment of the developed ontology based on known tools and metrics. Section 7 presents useful conclusions drawn and provides the baseline for future work in the field identifying potential extensions and further enhancements.

2. Related Work

Public Administration and government institutions have widely adopted Open Data mostly through the launch of data portals [2]. A number of best practices of publishing Open Government Data includes portals such as:

[1] https://digital-strategy.ec.europa.eu/en/policies/open-data
[2] https://diavgeia.gov.gr/

- Official UK Legislation: The official government archive [3] of the United Kingdom, managed by the National Archives, providing access to published UK legislation, with available data covering a period of 800 years in time as of 1267;
- The UK official National Open (Government) Data Portal where Central government, Local authorities and Public bodies can publish [4];
- US Government Linked Open Data [5], the US government Open Data project. The Data.gov project's Semantic Community http://semantic.data.gov provides access to, and guidance on the use of Linked Data and Semantic Web technologies;
- Data Europa EU [6] providing access to over 1.4 million public datasets from 36 countries (European Union Member States, the EEA, Switzerland and countries in the EU Neighbourhood Policy Programme).

Concerning, though, the handling of data with regards to documents, decisions and acts, the common approach followed by the majority of Public Organizations is to merely upload documents on the Web, in formats such as PDF, of low ranking according to the 5-star deployment scheme for Open Data quality, not ensuring compliance to the Linked Open Data requirements. To achieve interoperability in the interpretation of administrative procedures and legislation, the integration of data coming from different sources and the effective inter-exchange of information in the context of European Public Services, we need to establish a common conceptual framework [3]. A number of standard ontologies and vocabularies have been developed to accommodate these requirements .

2.1. ISA²- core vocabularies

The "Interoperability solutions for public administrations, businesses and citizens (ISA²)" Programme [4], supporting the development of digital solutions enabling administrations, enterprises and citizens in Europe to benefit from inter-operable cross-border and cross-sector public services[7], developed the EU core vocabularies [8]. The core vocabularies can serve as the tool to harmonize data representation in a comprehensive manner. Core vocabularies are simplified, reusable and extensible data models that capture the fundamental characteristics of a web resource, an entity, such as a Person or a Public Organisation for instance, in a context-neutral manner, abiding by the Linked Data principles. Moreover, core vocabularies promote the use of common identifiers for organisations, people and locations in the form of Uniform Resource Identifiers (URIs), can be easily combined with other Linked Data vocabularies, and are extendable with new classes and attributes to fulfill new domain requirements [9].

The most important core vocabularies developed under ISA² so far are the following:

- Core Person, capturing the fundamental characteristics of a Person, e.g. name, gender, date of birth;
- Core Business, encapsulating the fundamental characteristics of a Legal Entity (e.g. its identifier, activities) which is created through a formal registration process, typically in a national or regional register;
- Core Location, identifying the fundamental characteristics of a location, represented as an address, a geographic name or geometry;
- Core Public Organisation, describing Public Organisations in the EU;
- Core Public Service Vocabulary, capturing the fundamental characteristics of a service offered by public administration, such as the title, description, inputs, outputs, providers, locations, etc. of the public service. An application profile of the Core Public Service Vocabulary (CPSV-AP) has been developed for describing public services and grouping them in business events;
- Core Criterion and Core Evidence, describing the principles and the means that a private entity must fulfil to become eligible or qualified to perform public services

[3] http://www.legislation.gov.uk
[4] http://data.gov.uk/
[5] https://www.data.gov/
[6] https://data.europa.eu/en
[7] https://ec.europa.eu/isa2/isa2_en
[8] https://ec.europa.eu/isa2/solutions/core-vocabularies_en
[9] https://op.europa.eu/el/web/eu-vocabularies/corevocs

2.2. The Organization ontology

The W3C Organization ontology [5] contributes as the main ontology for organizational structures, since it is designed to allow domain-specific extensions to add classification of organizations and roles.

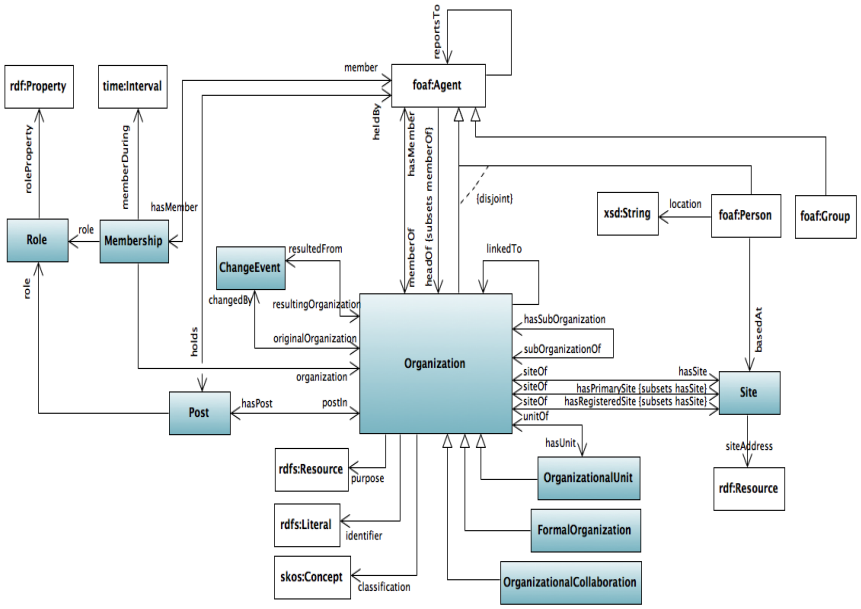

Figure 1. W3C The Organization Ontology-"Copyright © 2012-2014 W3C® (MIT, ERCIM, Keio, Beihang) https://www.w3.org/TR/vocab-org/"

This ontology is designed to enable publication of information on organizations and organizational structures including governmental organizations to provide a generic, reusable core ontology that can be further extended or specialized. This proves fit for the purpose of information extraction from government decisions and acts to identify the main actors and contact persons.

2.3. E-procurement ontology

A significant part of documentation at EU level is related to financial transactions. In this context, the procurement process holds a prominent place (Public procurement represents around 20 percent of GDP in Europe). Therefore, the EU is investing significantly on the digitisation of the public procurement process (referred to as e-procurement). The procurement procedure itself can be quite complex involving many actors and discrete phases end to end, i.e. from notification, through tendering to awarding, ordering, invoicing and payment. This in turn implies variable requirements to cater for different entities and their inter-relationships. This triggered the effort to establish several procedures and standards at EU legislative level (indicatively Directives 2014/24/EU, 2014/25/EU and 2014/23/EU establish rules with respect to public contracts, design contests and concessions, whereas Directive 2014/55/EU defines the requirement for a European standard for electronic invoices, and the Commission Implementing Regulation (EU) 2015/19866 specifies standard forms for the publication of notices in the EU Official Journal[10]).

Given the increasing importance of data standards for e-procurement, there is a number of initiatives driven by the public sector, the industry and academia over the recent years, with a diversity in terms of the vocabularies and the semantics that they are introducing, the phases of public procurement they are covering, and the technologies they are using. These differences hamper data inter-operability and thus its reuse. This creates the need for a

[10]https://eur-lex.europa.eu/oj/direct-access.html

common data standard for publishing procurement data, hence allowing data from different sources to be easily accessed and linked, and consequently re-used. Hence, an Ontology of the Public e-Procurement (ePO) [11] was developed to act as the common standard on the conceptual level, based on the main stakeholders consensus and designed to encompass the major requirements of the e-procurement process complying with the aforementioned EU Directives. Its goal is to formally encode and make available in an open, structured and machine-readable format public procurement data. The e-procurement ontology aims to unify existing practices to make it easier to share, access and re-use data[12] [13].

2.4. Diavgeia Programme

At national level, in Greece, a good showcase is Diavgeia ("Diavgeia" (΄Διαύγεια΄) is the Greek word Transparency). Diavgeia is a Programme introduced by Law in 2010 obliging Public Organizations to post their decisions and acts on the Internet. Each document is digitally signed and assigned a unique Internet Uploading Number (IUN) of primary importance, since it operates as a sole reference code certifying that the decision has been uploaded on the respective Diavgeia Portal. Moreover, what makes this effort valuable is that administrative acts and decisions are not considered valid unless published online. This enhances significantly the usability, applicability and role of the Programme in the sphere of Public Administration, which is further exploited as source of data for privately developed applications. The Diavgeia Programme is considered an Open Government Best Practice, received very positively at national, but also at European level [6]. Overall, a significant number of acts and decisions have been published on the Portal, reaching 50 million during its operation to-date, whereas as the rate of uploads has reached 28 million decisions per working day [14].

2.4.1. DiavgeiaRedefined-Diavgeia ontology

A concrete effort to build upon and enhance the public Programme Diavgeia is the open-source development "DiavgeiaRedefined". The project proposes a modular framework using existing ontologies developed in OWL and queried through SPARQL with the aim to modernize and enhance the way that decisions and acts are made public, following the paradigm of other successful efforts in Europe which publish legislative documents as Linked Open Data, applying Semantic Web techniques [7].

The corresponding Diavgeia ontology [15] developed incorporates elements from the distinct ontology Nomothesia ("Nomothesia" stands for Legislation in Greek) as concerns the legislation dimension [8]. Nomothesia is an OWL ontology adopting the ELI framework for modeling the content of Greek legislation documents, along with their accompanying metadata (i.e., title, gazette, publication date, etc.), capturing dynamically how these documents may evolve through time in response to modifications, since this is one of the fundamental issues in the legislation procedure [16]. ELI [9], one of the actions supported in the frame of the ISA² Programme [17], is a system to make legislation available online in a standardised format, so that it can be accessed, exchanged and re-used in a cross-sectoral approach. The ELI ontology is demonstrated as a cornerstone of a 'legal linked data', as it describes relationships between national and European legislative resources, contributing to unification and standardization at European level. It offers also the backbone for legal documentation, making it appropriate for governmental decisions and acts.

A representation of the core of the proposed system, highlighting the grouping of the components from different sources/ontologies (Nomothesia, ELI) [7] is presented in Figure 2. It also integrates Greek Administrative Geography Ontology, a typical ontology to represent the Greek Administrative hierarchical structure [18]. The latter can be deployed to map the Public Organizations issuing decisions and acts depending on the level of administration:

[11] https://github.com/OP-TED/ePO
[12] https://joinup.ec.europa.eu/sites/default/files/document/2017-08/d02.01_specification_of_the_process_and_methodology_v1.00.pdf
[13] https://eprocurement-everis.github.io/
[14] https://diavgeia.gov.gr/
[15] https://github.com/ThemisB/diavgeiaRedefined
[16] http://legislation.di.uoa.gr/nomothesia.owl
[17] https://eur-lex.europa.eu/eli-register/about.html
[18] https://pergamos.lib.uoa.gr/uoa/dl/frontend/file/lib/default/data/1324504/theFile

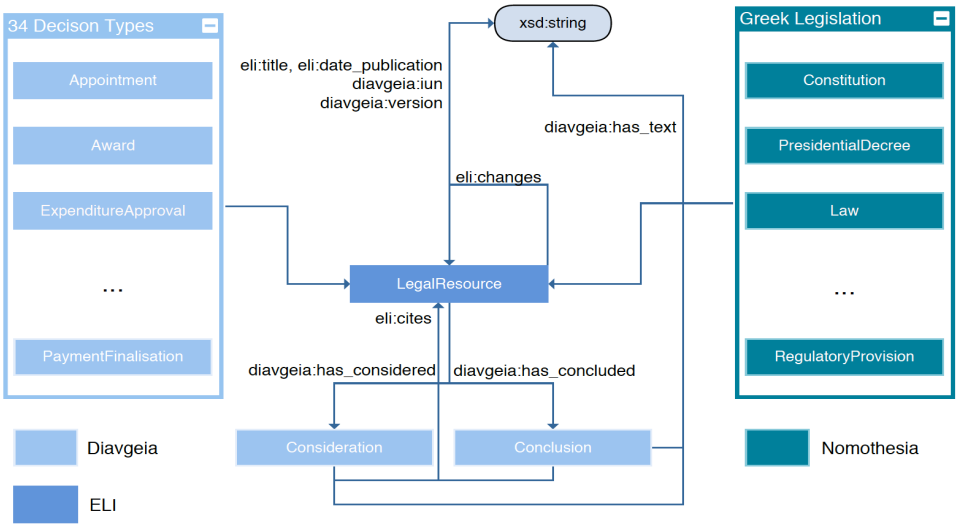

Figure 2. Diavgeia Ontology "Copyright © 2017 Beris Themis[7]"

Organizations at Regional level (Decentralized Administrations, Regions/ Regional Units, and Municipalities) not including though Organizations at the central level of government such as Ministries or other types of Institutions such as Universities, Hospitals etc.

Diavgeia Ontology represents a good use case that encapsulates diverse individual characteristics following universally adopted standards and provides for re-usability of its main elements. Nevertheless, it is built in a rigid manner following the classification of the published Greek Public Administration decisions and acts into certain categories, abiding by a predefined format and the exact meta-data required per decision/act fields. It is oriented towards a rather simplistic translation of the Diavgeia acts into an ontology following more or less strictly a predefined pattern, with certain enhancements as to what concerns more generic properties to accommodate horizontal requirements originating from the different decisions/acts types.

In this respect, it can serve as the starting point for a more complete and universal solution. Our goal is to have decisions that comply to a new integrated ontology, replacing the uploading of low ranked PDF files with the corresponding compressed RDF ones, upgrading decisions/acts issued by the public organizations to Linked data with a "5-star" rating, but in a more systematic and universal format compared to the Diavgeia ontology deploying the use of EU common standards, not restricted to fit the requirements of decisions/acts of a single national Programme. To this end, this solution can be derived following the concept of the recently proposed intelligent framework, handling both creation and real-time updating of a knowledge graph, while also exploiting domain-specific ontology standards, deploying Diavgeia [10]. In the following section, we elaborate further on the enhancements proposed to build the new integrated ontology.

3. Methodology

The methodology developed in the frame of the present work tries to encapsulate the basic elements of the aforementioned in the related work technologies and principles. We take on the Diavgeia ontology, trying to extend it by additional classes, object and data properties with the objective to significantly enhance its functionality and re-usability in terms of widely used standard ontologies and vocabularies already conforming to commonly agreed standards at European level. This constitutes the added value of this work which aims to provide a scalable working solution. To this end, we applied the basic principles of ontology engineering, where applicable in our model, according to the W3C guidelines[19] to establish the framework for high quality linked open data, develop

[19]https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

our integrated ontology and subsequently build a Knowledge Graph in the domain of interest, that is a graph based on knowledge extracted from government decisions and acts as encoded in the Greek Platform Diavgeia.

3.1. Use Cases

The development of Use Cases provides the means to identify the needs of the end user of a Knowledge based system. One should focus on real applicable use cases, independently whether this involves a Knowledge Graph or the respective ontology. The added value we try to bring in is to further elaborate on valuable practical Use Cases for the end user, be it a public servant or a citizen in a wider sense engaging in public policies, combining the deployment of standard ontologies and core vocabularies, following to a great extent the EU standardization so as to enable inter-operability.

Use Case 1: Transparency/Accountability in public money/resources spending

Accountability for the allocation of public money or -in general resources- at national and EU level is the driving force to develop tools for monitoring the money flow. A characteristic use case is to identify the recipient organizations-economic operators of public money.

Diavgeia, as the main repository for decisions related to the procurement procedures in Greece, is an important source of information. Related Diavgeia decisions/acts can help us identify the recipient contractors, the volume of awarded budget, the frequency of awarded contracts to specific economic operators so as to establish potential patterns in the awards or even suspicion of preferential treatment.

Use Case 2: Publicity in public spending

A use case focusing mostly on the publicity requirements related to the (pre)award phase. It is necessary and legally binding in most cases through established procedures at EU and national level that contracting authorities-public organizations announce and publish the calls for tenders to economic operators, citizens and third parties. Essential piece of information consists of the type of procurement procedure, i.e. open/closed tender, selection and award criteria to be fulfilled by the candidates, a potential break down in tender lots, if applicable.

It is also of primary importance that public organizations can be timely and effectively reached to provide feedback on procedural issues. Therefore, Contact Points' information should be available in all possible means of communication (email/telephone/postal address/contact persons) and in this sense modelled by an underlying system (ontology).

Use Case 3: Efficiency of the decision-making processes

In decision-making processes knowledge is the foremost element that contributes to productive results. If we are interested in financial transactions, we could further orient our search accordingly. Thus, to obtain an overview of public resources allocation one could be oriented towards cumulative information. To elaborate on critical financial information, the available data can be further broken down to actual budget categories so as to identify where public money is spent, i.e. to which kind of goods, works, equipment, consumables, services etc. This can be done through retrieving related Diavgeia Award decisions in order to identify the type of procurement via the Common Procurement Vocabulary (CPV) values. In the same manner, one could be interested in specific information concerning personnel appointment procedures by Public Organizations. In this respect, data on the type/category of personnel appointed or the frequency of appointments for a specific public organization could be of interest to the citizens. In terms of its internal functioning, an organization could collect data for statistical reasons for decisions (for instance, the average number or duration of public Contracts) in order to assess the efficiency of its organizational units.

3.2. Competency Questions

The next step is to translate these Use Cases in ontology requirements. In this respect, Competency Questions (CQs) can be used in the course of ontology development. CQs can be employed as the means to determine the ontology requirements. They consist of a set of questions stated and replied in natural language.

Given a set of scenarios related to the application field, developers should be able to place a set of questions representing users demands. These questions support the development process in two ways:

- Enabling developers to identify the main entities and their relationships to create the ontology vocabulary (terminology), and
- Providing developers with a simple means to verify requirements' compliance by either knowledge retrieval or by entailment on its axioms and answers checking [11].

Therefore, we identify domains of interest and develop questions that will drive the identification of the appropriate ontology components and facilitate their implementation into a new ontology.

Use Case 1: Transparency/Accountability in public money/resources spending

- CQ1: For a given organization, which are the top x economic operators that are recipients of awarded contracts (within a given time period)?
- CQ2: For a given organization, which are the awarded contracts to a specific economic operator (within a given time period)?
- CQ3: For a given organization, what are the direct awards (awarded value below a threshold, currently set at 30.000€), not following a tendering procedure (within a given time period)?
- CQ4: What are the top x contracting authorities (Public Organizations)? (their organizational structure, main activities, location data)?
- CQ5: What is the number of appointments for a certain person (within a given time period)?

Use Case 2: Publicity in public spending

- CQ1: Which are the selection criteria for a tender?
- CQ2: Which are the award criteria for a tender?
- CQ3: Which is the full information for the Contact Point for a decision/act (the designated organizational units/person)?
- CQ4: What is the submission deadline, date/time for a tendering procedure?
- CQ5: What is the opening date/time for a tendering procedure?

Use Case 3: Efficiency of the decision-making processes.

- CQ1: For a given CPV (type of procured assets), what is the number of contracts awarded and the total amount awarded?
- CQ2: For a given organization what is the most popular type of awards based on the CPV (type of procured assets)?
- CQ3: For a given organization, what is the number of persons appointed (within a given period of time)?
- CQ4: What is the budget per year awarded though certain type of procurement procedure/CQ4b:What is the budget per year awarded though funding by European Funds?
- CQ5: What is the average duration of contracts awarded?

One can further elaborate on combinations of the above queries to build additional competency questions customized to the needs of the end user or the organization. These drive the knowledge extraction process.

4. The OWL ontology for government decisions and acts

4.1. Diavgeia documents analysis

A reference document from the Diavgeia Programme is analysed to identify the corresponding classes and properties to be integrated from standard ontologies and vocabularies. The document is broken down in three main Parts (A, B, C) according to a typical document layout.

For instance, for the sample document selected (Fig.3) we can extract Organization's related data from Part A. It is issued by a 'Public Organization' (class), i.e. the 'Greek Ministry of Development and Investments' (΄ΥΠΟΥΡΓΕΙΟ ΕΣΩΤΕΡΙΚΩΝ΄ in Greek). We identify the internal hierarchical structure of the Ministry comprising a General Directorate (΄ΓΕΝ. ΔΙΕΥΘΥΝΣΗ΄), a Directorate (΄ΔΙΕΥΘΥΝΣΗ΄) and a Department (΄ΤΜΗΜΑ΄) and we see how

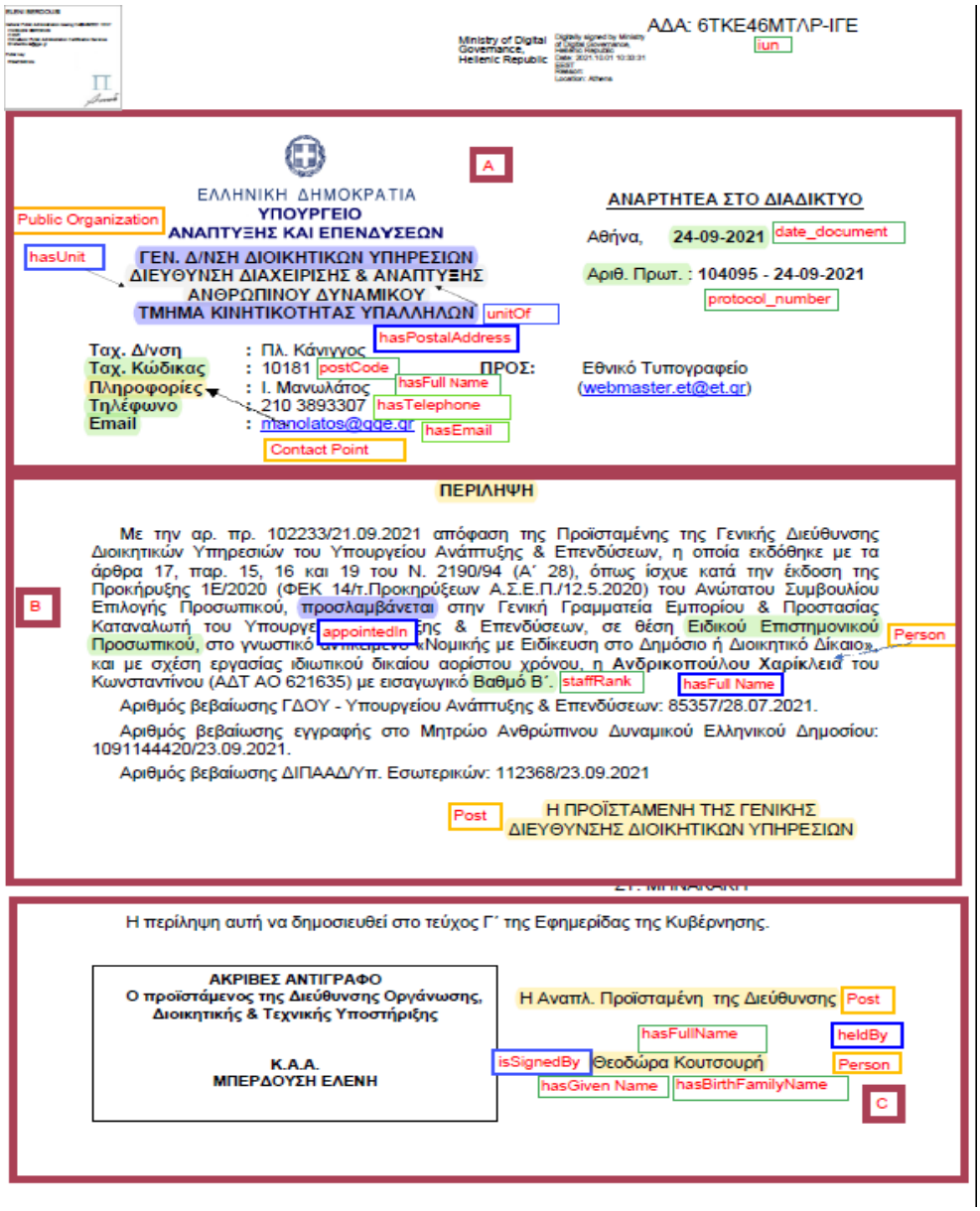

Figure 3. Diavgeia sample in focus

these are related via the appropriate object properties (’hasUnit’/’unitOf’ of the e-procurement ontology in this case). Further down within Part A, we can extract data for the Contact Point (class ContactPoint) ΄Πληροφορίες΄ and how this is reached via the appropriate object properties such as ’hasPostalAddress’, and the data properties ’postCode’, ’hasTelephone’, ’hasEmail’, ’date_document’, ’protocol_number’ etc. It is interesting to note that in most cases the classes and properties are self-explained in the sense that their naming is quite straightforward.

Part B constitutes the main part of the document comprising the vital information. For this appointment decision, object properties such (΄προσλαμβάνεται΄) matches the property (’appointedIn’), data property (΄Βαθμό Β΄) matches (’staffRank’) etc.

In the concluding part C, the classes ’Post’, ’Person’ are necessary to identify the Signer introducing the object properties ’isSignedBy’, ’heldBy’ and data properties ’hasFullName’, ’hasGivenName’, ’hasBirthFamilyName’ to

retrieve a variety RDF triples. Besides the information retrieved in the three main parts, the document includes the unique Identifier of the Diavgeia on the right upper side retrieved by the data property 'iun'.

4.2. Ontology built-up

Taking into account the prior analysis of the typical document we can practically fit the corresponding entities to the appropriate pieces of information within the document. We perform the translation of the main document elements in ontology terms, i.e. via their classes and the object and data properties used to inter-relate and describe them. That means identifying which components of standard ontologies can be used to map the information included in the Diavgeia documents and then match them. This is the common approach followed for a set of documents through the new integrated ontology built-up.

4.3. d2kg ontology

The final integrated ontology d2kg includes the appropriate ontologies, core and controlled vocabularies. A graphical UML representation of the main entities and their relationships is provided in Figure 4 below.

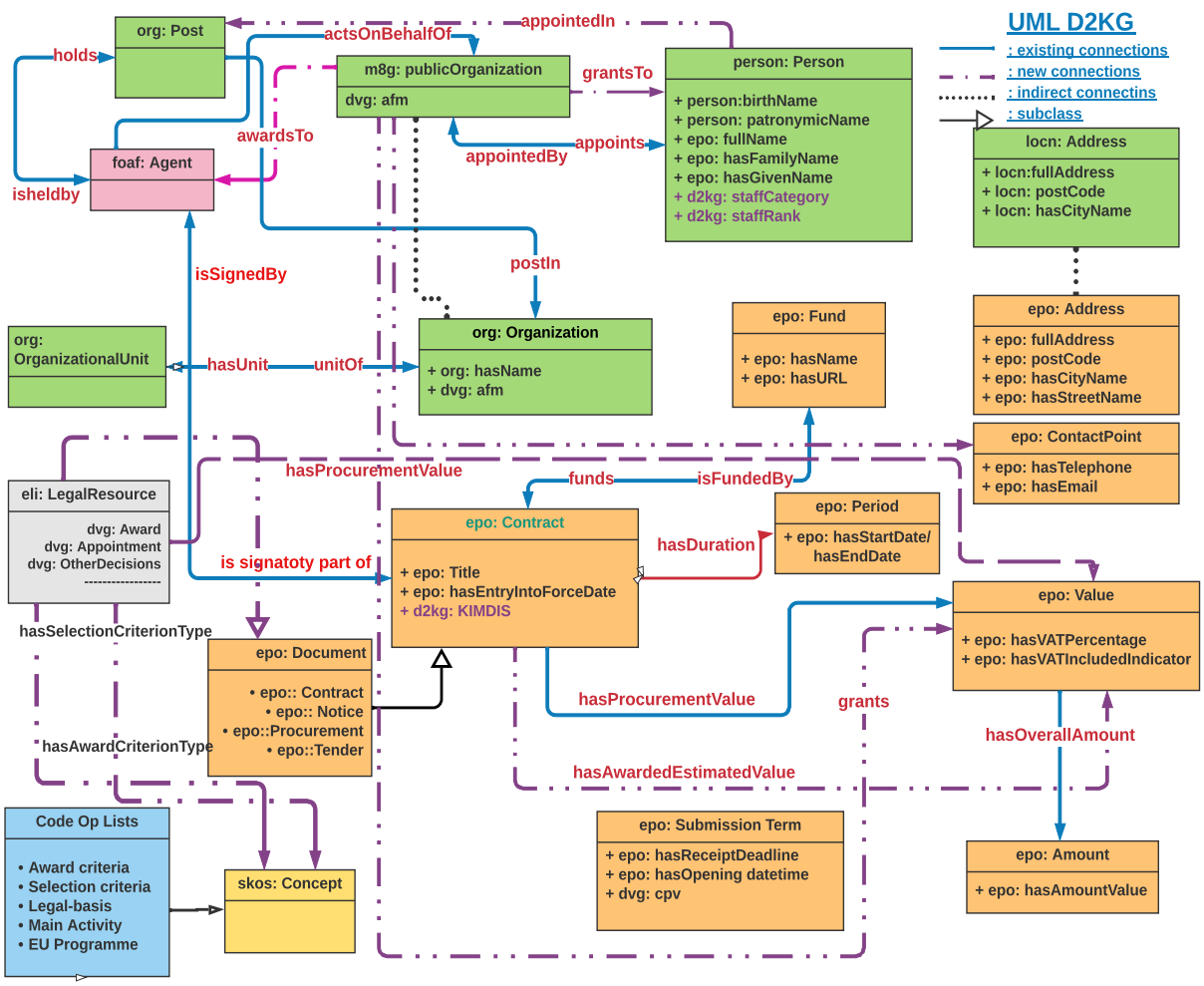

Figure 4. d2kg Ontology representation

This representation gives an outlook of the developed ontology highlighting the different components integrated in the current implementation, along with several of the more frequently used data and object properties. At the level of the main ontologies and vocabularies, the ISA core vocabularies are shown in green, the ePO in orange and the main classes of the Diavgeia ontology in grey, whereas the controlled vocabularies in blue. The blue continuous lines show existing connections between different classes via their properties, whereas the purple non-continuous lines indicate potential new connections that can be established with the re-use of existing properties. For instance, the property 'hasProcurementValue' can connect a Diavgeia decision of type dvg:Award with the class epo:Value. It is evident that the focus is on the re-use of existing classes, object and data properties from the imported ontologies, along with additional ones for the purpose of extracting valuable information from Diavgeia decisions and acts. The majority of data and object properties derives from the ePO and the Diavgeia ontology.

The d2kg ontology, in this respect, is a unique integration of existing ontologies combined with core and controlled vocabularies developed based on EU standards. It provides a customized solution to abide by the requirements of the Greek Programme Diavgeia, extending significantly the respective Diavgeia ontology and proposing at the same time a solution to encode government and administrative decisions/acts that could be universally adopted to integrate public documents produced by other EU Member States, with certain adjustments contentwise.

Characteristic entities integrated in d2kg ontology are detailed in the next sections.

4.3.1. d2kg classes

The ontology is built on re-used classes of the imported individual ontologies. We will describe the commonly deployed per ontology below.

Diavgeia ontology

The Diavgeia ontology classes are extensively analysed in the corresponding repository of the DiavgeiaRedefined Project [20] (Figure 1 serves as reference too). For the sake of completeness, we refer to the basic ones mostly used in the context of this work here:

- LegalResource: the core class representing the decisions/acts of Diavgeia based on their formal classification according to the Diavgeia Programme;
- Expense: the most common entity to represent financial transactions; it is used by the following decision types following the notation of the Diavgeia ontology: Award, Contract, DeclarationSummary, DonationGrant, ExpenditureApproval, OwnershipTransferOfAssets, WorkAssignmentSupplyServicesStudies, PaymentFinalisation, GeneralSpecialSecretaryMonocraticBody. All the above involve a financial aspect (relevant to monetary transaction) which implies the need for a separate class to encode accompanying data such as the involved parties, amount etc;

ePO ontology

This is a summarization of the main entities introduced [21]:

- Agent: A person, an organization, or a system that act in procurement or have the power to act in procurement; This is the respective class from the FOAF ontology, as integrated in ePO;
- ContactPoint: Details used to reach an organisation: a role, email address, telephone number, etc. This is the respective class from schema.org integrated in the ePO. It can prove very useful in the current implementation, as the decisions/acts normally have a Contact Person (Point) to be reached by the citizens;
- Fund: A financial resource used to support the procurement. In the context of EU, funds can be divided into Programmes, Actions and Projects. Examples of EU funds are: the European Structural and Investment Funds, European Social Fund (ESF), the Connecting Europe Facility (CEF) Programme, or the ISA2 Programme and its actions (e.g. Action 2016.05 European Public Procurement Initiative, which supports the e-Procurement Ontology). Funds may change between the lot and the contract, for example in the case of an emergency crisis, a contract may e financed by a budget that was not foreseen in the call;

[20] https://github.com/ThemisB/diavgeiaRedefined/tree/master/rdf
[21] https://github.com/OP-TED/ePO/tree/v2.0.1/v2.0.1

- Period: A time interval or a duration, usually consisting of a start and an end date;
- Purpose: The description of the objectives related to a procurement;
- Tender: Information submitted by the economic operator to specify its offer regarding one or more lots or the whole procedure, in response to the call for tender;
- Value: Value of an asset, normally expressed as Amount.

Organization ontology

In the context of the present implementation, we deploy the following classes [22]:

- Organization: representing a collection of people organized together into a community or other social, commercial or political structure, often de-composable into hierarchical structures;
- Post: representing some position in the organization that may or may not be currently filled. It is a vital element in the Public Admin sphere, as Posts enable reporting structures and organization charts to be represented independently of the individuals holding those posts.

The above effectively allow us to identify the Organization structure and hierarchical relationships involved in Diavgeia acts. Additionally, it provides us with the possibility to identify Persons related to the organizational structures at a given time; for instance the person holding a certain post at a given period and its role in a procedure. On top of that, if combined with certain properties introduced by the E-procurement ontology, such as *"Acts on behalf of"*, legal relationships can be derived which could be quite important in case of transactions between different entities.

In terms of the e-Government core vocabularies we introduce the following per core vocabulary:

e-Government core vocabularies [23]

Core Public Organization vocabulary

- Public Organization: a class that represents the Organization. One Organization may comprise several sub-organizations and any organization may have one or more organizational units. Each of these is described using the same properties and relationships; In the context of this implementation, we use this class for decisions/acts issued by Public Administration (this could be interchangeably used with the class Organization of the Organization Ontology in a wider context; In the RDF release of the CPOV, hasUnit is equivalent to org:hasUnit and unitOf is equivalent to org:unitOf).

Core Person vocabulary

- Person: An individual person who may be dead or alive, but not imaginary. It is that restriction that makes person: Person a sub class of both foaf:Person and schema:Person which both cover imaginary characters as well as real people.

Core Location vocabulary

- Address: Its properties are closely bound to the INSPIRE data model for addresses. The Location core vocabulary does borrow the fullAddress property from VCard as a means of providing the full text of the address as a literal;
- Location: dcterms:Location class fully represents the ISA Programme Location core vocabulary class of Location.

4.3.2. d2kg object properties

We can investigate on the most important object properties per ontology introduced:

Diavgeia ontology

- has_expense: has expense links a certain decision type with an Expense;

[22] https://www.w3.org/TR/vocab-org/

[23] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/release/201

– signed_by: signed by links a Legal Resource (decision type) with a Signer; there is an equivalent property in ePO ontology as well.

ePO ontology

– appointedBy: used in acts related to appointment of new staff to organizations/inverse of 'appoints';
– funds: represents the relationship between the Funding source (source of funding, i.e. European or National Budget) and the recipient organization /inverse of 'is_funded_by';
– hasAwardCriterionType: the determining criterion for awarding the tender to a candidate contractor (lowest price, cost, quality); important to be communicated to candidate contractors;
– hasProcurementValue: used in the context of Contractual binding agreements; It refers to the initially set value at the time the tender is announced. At contract time, this procurement value may be different from the Procurement Value of a Lot or a Procedure that was announced. Associated with the class 'Value';
– hasAwardedValue: The value of the procurement provided by the Award decision, i.e. the actual value awarded to the contractor when the procurement is concluded;
– hasOpeningPlace: it provides information on the location where the quotations are opened and evaluated; The place where the tenders will be publicly opened. Important for the sake of transparency to be widely communicated, since candidate contractors can be present during the opening of the tenders (range: Address);
– hasOverallAmount: relates the classes Value and Amount to link the generic concept of Value with a corresponding Amount when the asset is expressed as monetary value;
– hasMainClassification: provides the Common Procurement Vocabulary (CPV) values/can be used interchangeably with the data property dvg:cpv of the Diavgeia ontology;
– hasPostalAddress: the postal address predicate connecting the entity Location with the class Address (to further be used to encode the actual address as data property of the Address class);
– hasProcedureType: related to the activities leading to the conclusion of public contracts used in public procurement according to the legislation- identifies the type of procedure: 'Open', 'Competitive Dialogue', 'Closed' etc.;
– isCreatedBy: to identify the issuing Organization (creator) of a document (decision/Act in this context);
– isSignedBy: identifies the Signer/inverse of 'is signatory part of'.

Organization ontology

– hasSubOrganization: to represent hierarchical structures within an Organization, important to identify the organizational units issuing a decision;
– holds: Indicates a post held by some Agent/inverse of 'heldBy';
– postIn: Indicates the Organization in which the post exists.

e-Government core vocabularies

Core Location ontology

– location: The location property links any resource to the Location class. Asserting the location relationship implies only that the domain has some connection to a Location in time or space. It does not imply that the resource is necessarily at that location at the time when the assertion is made.

4.3.3. New object properties

Apart from the integrated properties from standard ontologies it was needed to create new ones to meet requirements not covered by existing properties. The necessity for these new properties comes from the specific type of data that can be retrieved from Diavgeia decisions/acts. They mostly represent relationships between an Organization and another entity (Organization or Individual/Person).

– appointedIn: expresses the relationship between the staff/personnel and the post where the person/individual is appointed in an organization;
– awardsTo: represents the property relationship between the funding organization and the recipient organization/inverse of 'isAwardedBy';

- grantsTo: used to define the relationship between an Organization Sponsor and the Sponsored Organization inverse of 'receivesGrantsBy';
- receivesGrantOf: defines the type of asset (e.g. amount) one Organization receives;
- staff: represents the personnel of an Organization; relates an Organization to the personnel in appointments;

4.3.4. d2kg data properties

In this section, we decided not to go extensively through existing properties, but highlight instead only the new data properties required to encode vital information in decisions/acts.

New data properties

It has been judged appropriate to introduce specific data properties as well to accommodate valuable and sometimes critical information encoded in certain decisions, as follows:

- kIMDIS: This stands for the central electronic register of public contracts reference (΄ΚΗΜΔΗΣ΄ in Greek) (rdfs:Literal);
- staffCategory: The staff Category of the personnel (rdfs:Literal);
- staffRank: The staff rank of the personnel (data range: corresponding to four different ranks: A, B, Γ, Δ);
- SAE: This property corresponds to the decision type issued for taking over financial commitments at the expense of the Public Investments programme budget (΄ΣΑΕ΄ in Greek) (rdfs:Literal);

4.4. d2kg controlled vocabularies

Apart from the appropriate classes, object and data properties, it is significant to introduce re-usability with regards to the terms used by the actual data incorporated via instances. This is possible through the integration of controlled vocabularies ensuring a standardized approach concerning the terms that correspond to predefined values for the properties. This is the point where the EU vocabularies are introduced [24].

EU vocabularies

The developed ontology introduces the publicly available arrangements in different formats as presented in the sub-sections that follow.

4.4.1. Authority tables

The Authority tables [25] is the structure that provides the consistent information to harmonise and standardise the codes and associated labels used in various environments (web platforms, systems and applications) and to facilitate data exchanges between the institutions involved in decision-making process and more.

Selection criterion type

In the domain of public procurement, selection criteria are normally based on a specific legal framework. This table [26] provides the list of conditions that are concerned for evaluation purposes in terms of the criteria that the candidate contractors should fulfil. It is common that these form elements referred in public administration documentation. This codelist is a subset of the ESPD codelist Criterion Taxonomy [27].

Award criterion type

In public procurement, it is important to make available in a standardized manner the award criteria types. This is normally part of the relevant decisions concluding the procedures and announcing formally the results. It conforms to the transparency requirements with regard to public resources allocation as it concerns not only the selected contractors, but the ones not chosen following a procurement procedure, and the wider public.

This is made possible through the authority table [28] with the list of rules to be taken into account for the award decisions. The initial values are those foreseen in the public procurement directives of 2014 (Directives 2014/23/EU, 2014/24/EU and 2014/25/EU) [29].

[24] https://op.europa.eu/en/web/eu-vocabularies
[25] https://op.europa.eu/en/web/eu-vocabularies/authority-tables
[26] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/selection-criterion
[27] https://docs.peppol.eu/pracc/espd/codelist/CriteriaTypeCode/
[28] https://op.europa.eu/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/award-criterion-type
[29] https://eur-lex.europa.eu/oj/direct-access.html

EU Programme

The EU Programme Authority Table (AT) [30] is a controlled vocabulary providing the list of programmes created and coordinated by and financially supported by the European Union or, in a few cases, by the contributions from the Member States. It has been developed specifically for the EU Budget as open linked data project. It indicates the authority code and start-use date of each concept and gives labels in all official EU languages. It provides useful insights when used in the context of an ontology to identify sources of funding for instance.

Main activity

A list of values [31] to classify the main activities of the buyers. The codes associated with contracting authorities are derived from the top level of the Classification of the functions of the government (COFOG) from the United Nations Statistics Division [32]. The codes associated with contracting entities are derived from sectors explicitly falling within the sectoral directive (2014/25/EU Art. 8 - Art. 14).

Procurement Procedure type

This set [33] identifies the procurement type selected (open, close, competitive dialogue etc), providing significant information on the procedure requirements to candidate contractors.

Legal basis

The legal basis [34] based on the legal acts used for a given public procurement procedure, as provided by the EU Publications Office.

4.4.2. Taxonomies

A taxonomy [35] is a controlled vocabulary in which all the terms belong to a single hierarchical structure and have parent/child or broader/narrower relationships to other terms, sometimes referred to as a 'tree'.

CPV

To make public procurement more transparent and efficient, European Commission drafted the Common Procurement Vocabulary (CPV) [36]. The Common Procurement Vocabulary (CPV) establishes a single classification system for public procurement aimed at standardising the references used by contracting authorities and entities to describe the subject of procurement contracts. It is adopted by Regulation (EC) No. 213/2008 [37].

4.5. URIs

The format of the URIs is differentiated depending on the actual usage of the entity. Concerning the main entity Legal Resource, the Persistent URIs approach is followed to comply as much as possible with common requirements [38]. The decisions/acts are structured according to the template: http://www.diavgeia.gov.gr/eli/iun/version of the Diavgeia Ontology. Modifications of a decision result to the generation of a new URI with the same iun and a new version number. This approach provides a flexible solution to accommodate decisions/acts following the ELI rationale.

Regarding other types of entities, though, it was noted that the Greek Diavgeia Programme has determined a way that these are uniquely represented, i.e. Diavgeia has already included the most important data and provided an available URI to look it up. More specifically, the representation per the most important entities is presented as follows:

-Organizations: https://diavgeia.gov.gr/opendata/organizations/uid, where uid stands for the Organization unique identification number. For example: https://diavgeia.gov.gr/opendata/organizations/50205 represents the

[30] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eu-programme
[31] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/main-activity
[32] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_the_functions_of_government_(COFOG)
[33] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/procurement-procedure-type
[34] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/legal-basis
[35] https://op.europa.eu/en/web/eu-vocabularies/taxonomies
[36] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/cpv
[37] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R0213&qid=1646558739181&from=EN
[38] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris

"Decentralized Administration of Macedonia-Thrace" (Public Organization) and points to a URL with administrative data on this organization (see Figure 5)


```
−<organization>
    <uid>50205</uid>
    <label>ΑΠΟΚΕΝΤΡΩΜΕΝΗ ΔΙΟΙΚΗΣΗ ΜΑΚΕΔΟΝΙΑΣ – ΘΡΑΚΗΣ</label>
    <latinName>admt</latinName>
    <status>active</status>
    <category>ADMINISTRATIVEREGION</category>
    <vatNumber>997612629</vatNumber>
    <fekNumber>235</fekNumber>
    <fekIssue>fektype_A</fekIssue>
    <fekYear>2010</fekYear>
    <odeManagerEmail>diavgeia@damt.gov.gr</odeManagerEmail>
    <website>http://www.damt.gov.gr</website>
    <supervisorId>22887</supervisorId>
    <supervisorLabel>ΠΕΡΙΦΕΡΕΙΕΣ</supervisorLabel>
    <organizationDomains/>
  </organization>
```


Figure 5. URI for an Organization

The same principle applies for a Person which fulfils the role of Signer/Final Signatory: https://diavgeia.gov.gr/opendata/signers/uid (uid identifying a signer based on Diavgeia records).


```
−<signer>
    <uid>100031111</uid>
    <firstName>ΙΩΑΝΝΗΣ</firstName>
    <lastName>ΨΩΜΑΣ</lastName>
    <active>true</active>
    <activeFrom>2017-04-06T03:00:00+03:00</activeFrom>
    <activeUntil>3000-01-01T00:00:00.921+02:00</activeUntil>
    <organizationId>100029495</organizationId>
    <hasOrganizationSignRights>false</hasOrganizationSignRights>
  −<units>
    −<unit>
        <positionId>POS_10128</positionId>
        <positionLabel>Α/Α (Αντι Αυτού) Προϊσταμένου Διεύθυνσης</positionLabel>
        <uid>100029671</uid>
      </unit>
    </units>
  </signer>
```


Figure 6. URI for a Signer/Final Signatory (Person)

5. Case studies

The applicability of the developed d2kg ontology can be demonstrated via the deployment of Knowledge Graphs to visualize actual government decisions and acts.

5.1. A Knowledge Graph representation for the Greek Programme Diavgeia

The developed ontology allows to produce data visualization via a Semantic Graph Database which is compliant with W3C Standards. The Diavgeia typical reference document, derived from the newly developed ontology, can be visualized in the form of a Knowledge Graph via Ontotext GraphDB [39], considered a highly efficient and robust Graph database with RDF and SPARQL support [12].

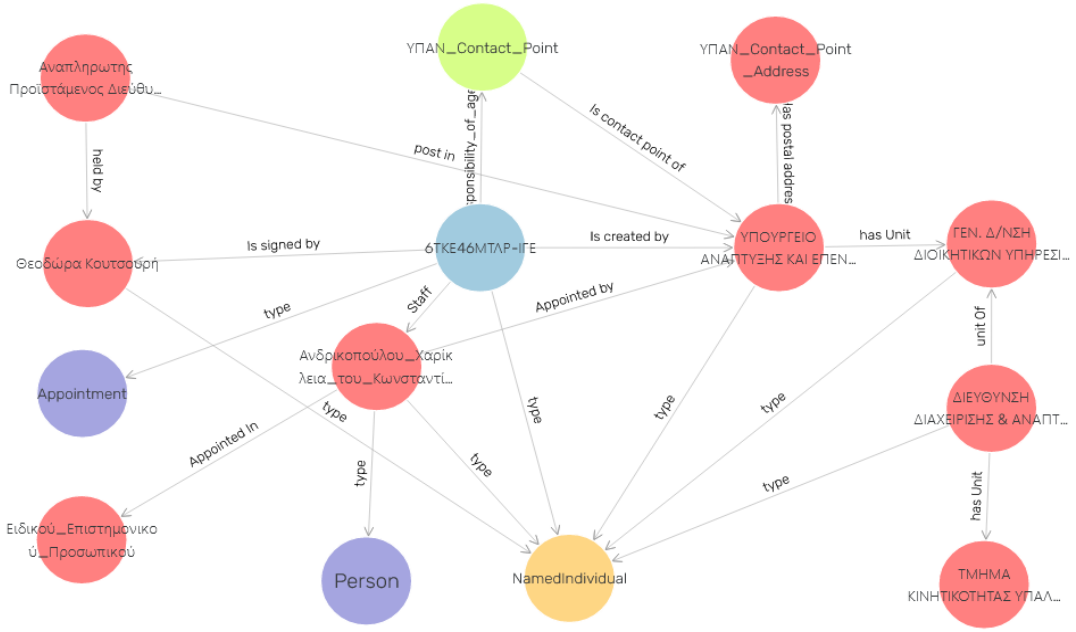

Figure 7. GraphDB visualization for the Diavgeia Sample in focus

In Fig.7 we can identify all core elements of this sample decision in focus. We can identify:
- the corresponding decision IUN: ΄6ΤΚΕ46ΜΤΛΡ-ΙΓΕ΄;
- the decision Type: 'Appointment';
- the Person appointed ΄Ανδρικοπούλου Χαρίκλεια΄
- the Organization which issues the decision and appoints the Person, the Greek Ministry of Development and Investments, ΄ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝΔΥΣΕΩΝ΄ in Greek;
- the Organizational Units of the Ministry: ΄ΓΕΝ. Δ/ΝΣΗ ΔΙΟΙΚΗΤΙΚΩΝ ΥΠΗΡΕΣΙΩΝ, ΔΙΕΥΘΥΝΣΗ ΔΙΑΧΕΙΡΙΣΗΣ & ΑΝΑΠΤΥΞΗΣ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ, ΤΜΗΜΑ ΚΙΝΗΤΙΚΟΤΗΤΑΣ ΥΠΑΛΛΗΛΩΝ΄ and their hierarchical structural relationship.

[39] https://www.ontotext.com/products/graphdb/

5.2. SPARQL queries

We deploy SPARQL for a systematic and targeted extraction of knowledge, building characteristic Competency Questions for the Use Cases we have identified.

Use Case 1: Transparency/Accountability in public money/resources spending

CQ1: For a given organization, which are the top x economic operators/contractors that are recipients of awarded contracts (within a given time period)?

A simple self-explanatory query to retrieve the top winning contractors for a given Organization (id:100054492).

```
SELECT (?Org AS ?Contractor) (COUNT(distinct(?contract)) AS ?number_of_contracts) where {
        ?contract a dvg:Award;
                epo:isCreatedBy dvgo:100054492;
                eli:date_publication ?pub_date;
                dvg:has_sponsored ?Org;
        FILTER (?pub date ⩾ "2017 − 01 − 01" ∧ ∧xsd : date)
        } group by ?Org order by desc(?number_of_contracts) LIMIT x
```

Use Case 2 : Publicity in public spending

CQ3: Which is the full information for the Contact Point for a decision/act (the designated organizational units/person)?

To enhance accountability it is necessary have data related to the issuing organization of a decision/act.

```
SELECT distinct ?doc ?URL ?full_name ?Email ?Telephone where {
        ?doc
                epo:isCreatedBy dvgo:100054492;
                epo:hasURL ?URL;
                eli:responsibility_of_agent ?Contact_Point.
        ?Contact_Point epo:hasFullName ?full_name;
                epo:hasEmail ?Email;
                epo:hasTelephone ?Telephone.
        }
```

Use Case 3: Efficiency of the decision-making process

CQ3: For a given organization, what is the number of persons appointed (within a given period of time)?

The following query invokes specific data on the actual number of employees appointed and their Staff Category that might be of interest to know for Human Resources Management within a (Public) Organization.

```
SELECT (COUNT(distinct(?doc)) AS ?number_docs) ?Staff_Category ?Staff ?Post where {
        ?doc a dvg:Appointment;
                eli:date_publication ?pub_date;
                d2kg:staff ?Staff;
        ?Staff d2kg:staffCategory ?Staff_Category;
                d2kg:appointedIn ?Post;
                person:birthName ?birthName;
                epo:appointedBy dvgo:99221922;
        FILTER (?pub date ⩾ "2015 − 01 − 01" ∧ ∧xsd : date)
        } group by ?Staff_Category ?Staff ?Post order by desc(?number_docs)
```

The queries results are shown in the following Figures 8, 9 and 10 respectively.

The query provides the list of operators/contractors ranked according to the highest number of contracts awarded by the Greek Ministry Of Interior (organization id: 100054492) after 01.01.2017 (date of publication). The reader can identify the integration of different ontologies (epo: for E-procurement, dvg: for Diavgeia, eli: for ELI ontology).

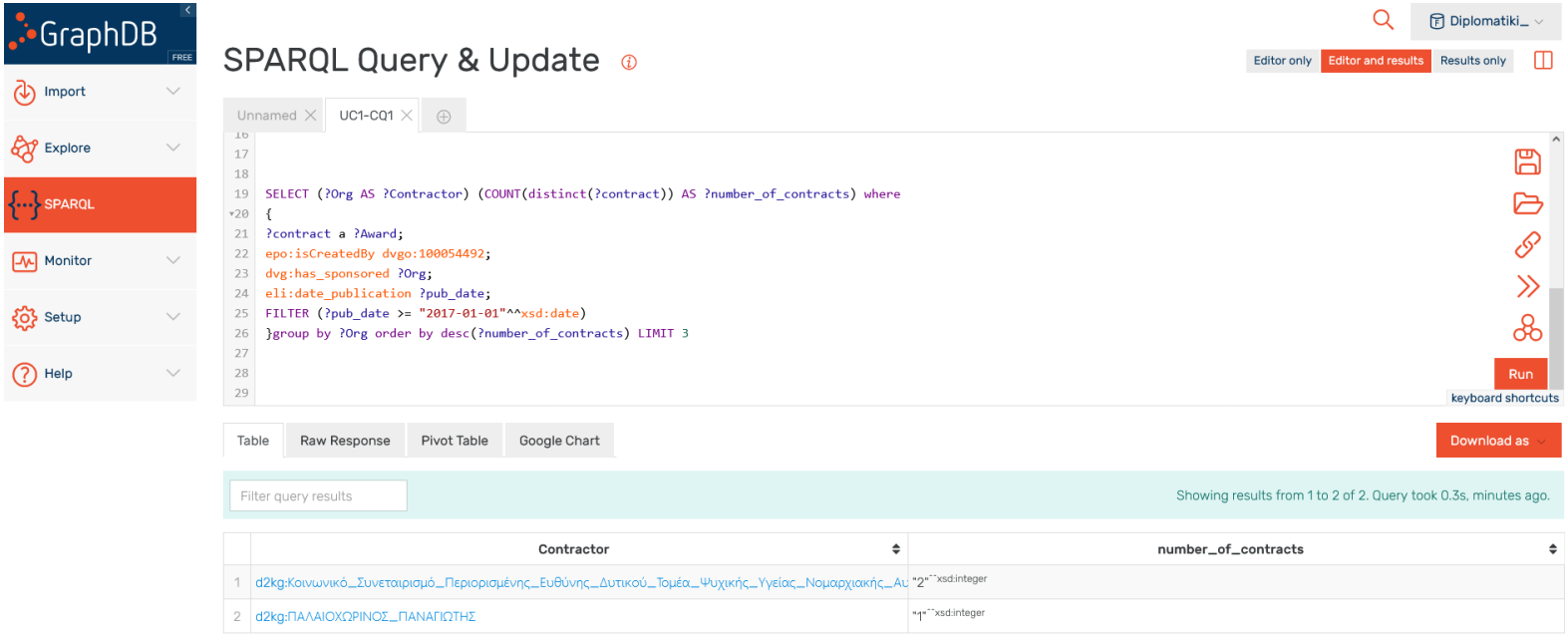

Figure 8. UC1-CQ1 results

The focus is on the Core Public Organization ontology via its main entity 'Contact Point' as introduced in the ePO ontology to retrieve data of the Responsible agent issuing the act.

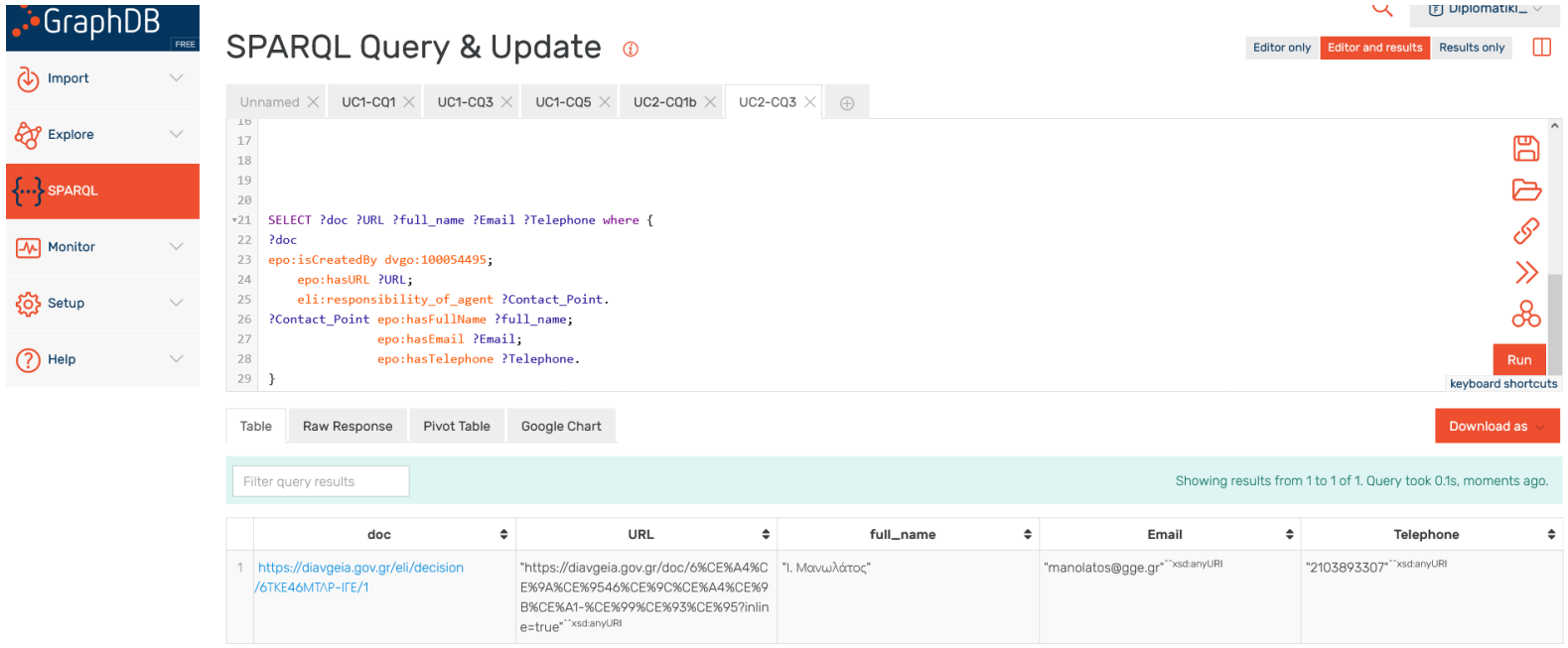

Figure 9. UC2-CQ3 results

This query re-uses the properties of Diavgeia ontology combined with ePO, eli and d2kg ontologies to provide a targeted view on a specific organization appointments procedure. For demonstration purposes we selected a Public Organization-Greek General Hospital of Corinth (id:99221922).

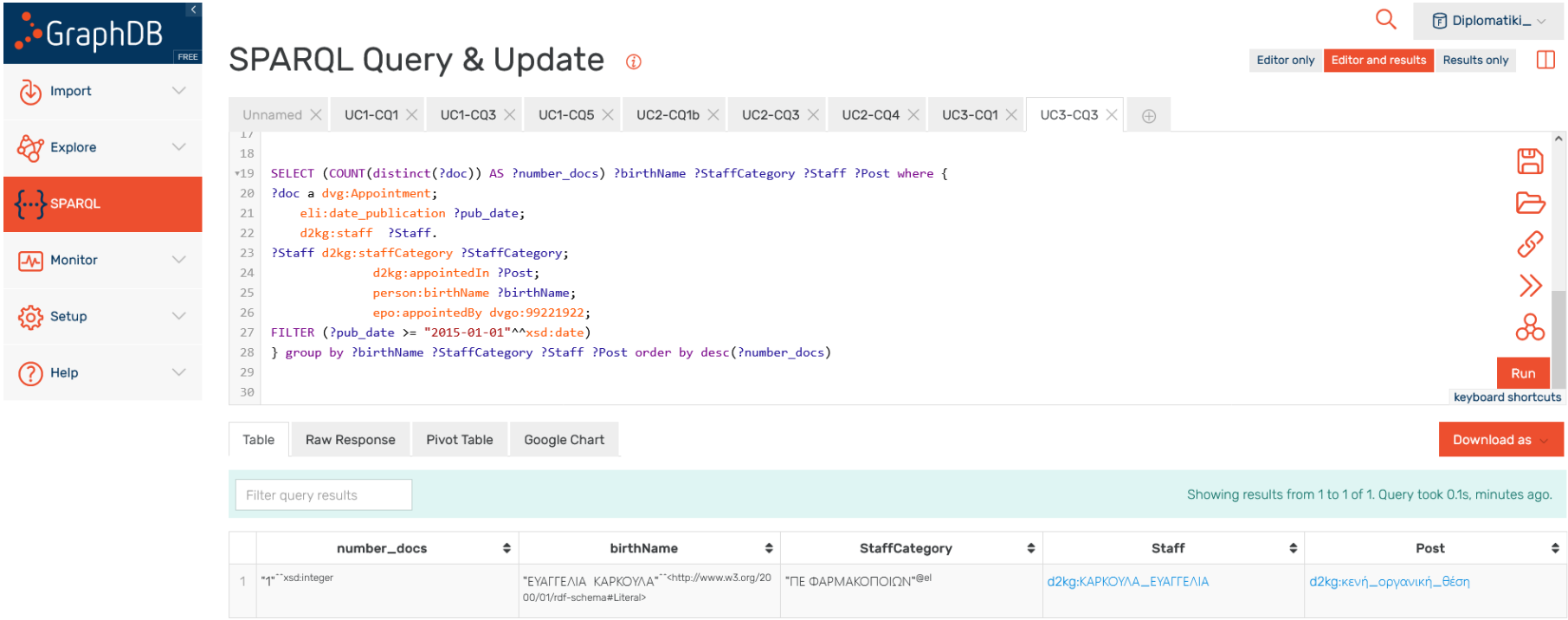

Figure 10. UC3-CQ3 in results

6. Ontology assessment

6.1. Debugging

To assess the ontology we initially deployed Protégé, the tool used to develop the ontology. We assess the possible faults in the ontology via the 'Pellet' Reasoner built in Protégé and the installed 'Debugger' plug-in. The debugging results that no faults occur during this validation process and no repairs are suggested for the ontology.

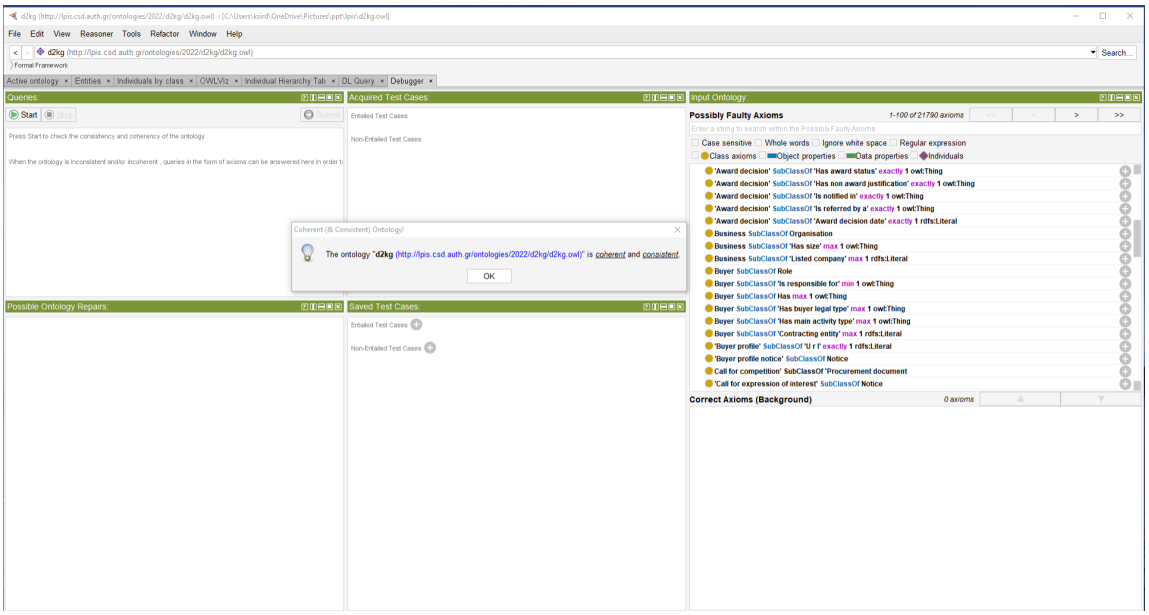

Figure 11. Debugging in Protégé

6.2. Reasoning

Reasoning can be used to test the validity of the concept approach and retrieve additional knowledge through the inferred concepts. Additional concepts can occur that could provide an insight of the data encapsulated in Diavgeia documents. This works also if there is no prior classification of an entity into a certain class, as Reasoning can deliver a classification.

A concrete example of a Declaration Summary Diavgeia act in the developed ontology is presented in Fig.12.

Apart from the classification of the document as Declaration Summary, we can infer that the decision/act involves (the properties are listed with their Labels, as shown in Protégé):

- Award Criterion, since the respective object property: "hasAwardCriterion" is extracted from the decision text and assigned the value 'Cost'. Therefore, the deployment of the Reasoner yields the corresponding 'Domain' as 'Award Criterion' for this property.

In the same manner, based on Domain/Range constraints, we gain additional knowledge. Certain inferred concepts are presented below:

- Contract: This means that a Contract is involved, as a result that we have extracted from the decision the respective property: "hasContractNature type" and its corresponding value 'Supplies'.
- Opening Term: This provides information on the Opening date and time since the respective property: "hasOpeningDateTime" and its corresponding value '2022-01-2022T15:00:00' are present in the decision.
- Procedure Type: The procurement procedure type is identified as 'Open' through the use of the respective property: "hasProcedureType" encoded in the decision.

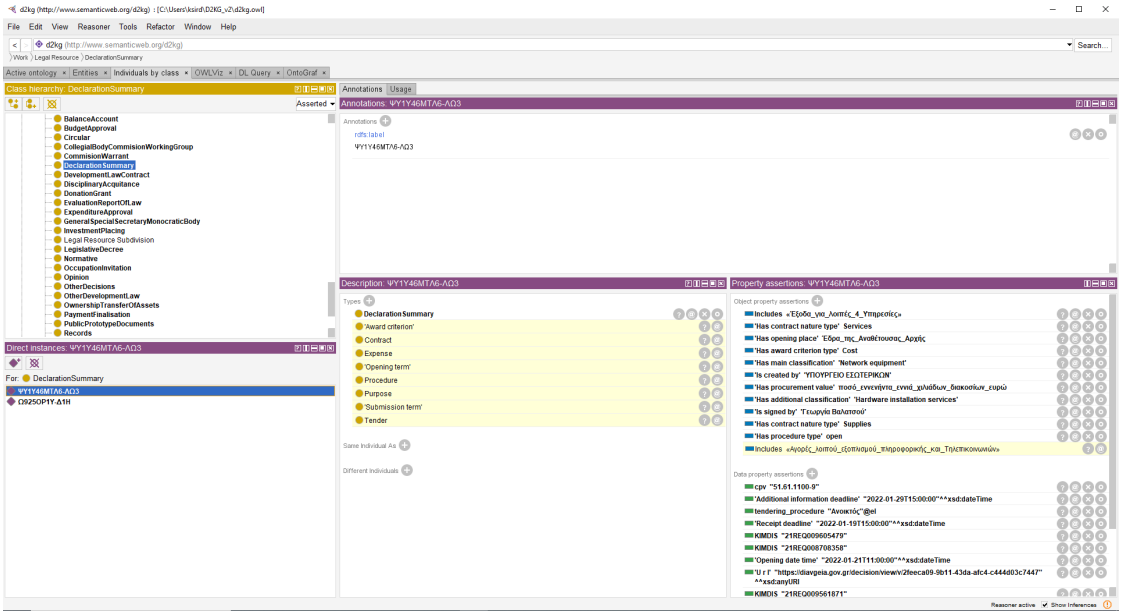

Figure 12. Reasoning in Protégé

The inferred concepts, despite not being straightforward but produced after Reasoning is applied, are meaningful. It is anticipated that for this type decisions we are expected to retrieve relevant information for the type of contract that is going to be signed for procurement, the type of the procurement procedure to be deployed to award the contract etc.

6.3. Metrics

In terms of the main metrics we initially proceed with Protégé metrics. A summary is depicted in Table (a) below, with the figures being representative of the ontology's size. It is expected that these figures show the collection of a high number of classes, properties and axioms as a result of the integration of the numerous standard ontologies and vocabularies and the number of populated instances.

(a) Protégé metrics

Metric	Value
Axioms	281730
Logical Axioms	21790
Class Count	210
Object Property count	361
Data Property count	333
Individual count	9856
Annotation Property count	60

(b) Ontometrics

Metric	Value
Attribute Richness	1.585714
Inheritance Richness	1.7666667
Relationship Richness	059513
Average Population	46.93333
Class Richness	0.185714
DL expressivity	SROIN(D)

To evaluate also the domain coverage we deployed the online platform OntoMetrics [13]. The assessment is based on ontology's accuracy and conciseness. An overview of the main metrics is presented in Table (b). We note the following:

- A high attribute richness value for the developed ontology, which is anticipated due to the integration of numerous ontologies in a single schema;

- A good coverage in the range of concepts with regards to the distribution of information in sub-classes per class, illustrated via the inheritance richness value;

- A balanced relationship richness in terms of the ratio of relationships non-inherited to the total number of relationships;
- A satisfactory value in terms of the number of instances compared to the number of classes as denoted by the average population value;
- A lower value for class richness, which is reasonable as we have manually populated the ontology with instances, not covering the entirety of the class knowledge contained in the schema.

The DL expressivity is classified as SROIN(D), derived with the use of both Ontometrics and Protégé.

7. Conclusions and Future Work

This paper focuses on the development of an integrated ontology considered fit to represent government decisions and acts. The Greek Programme Diavgeia was selected as a representative case study since it is a national Public Administration repository where substantial sources of governmental/administrative public documents reside. The implementation identified the added value of developing a new ontology by integrating standard ontologies in the field of knowledge extraction from Public Administration decisions/acts. Characteristic entities complying with EU core and controlled vocabularies and standards, following W3C recommendations were used together with newly developed ones. It is evident that in conjunction with "ISA²-Interoperability solutions for Public Administrations, businesses and citizens" and other collaborative efforts, there is significant potential of promoting knowledge creation. More specifically, the underlying data "hidden" in "Diavgeia" docs and how this can feed and enrich effectively a Knowledge Graph is highlighted.

In the context of the present work a core ontology was built as the "skeleton" of a project to highlight the capacity to present public domain data as inter-related Linked Open Data that be easily exploited. The proposed solution shows how to overcome the obstacles of uploading the data in formats, such as PDF, of lesser value and quality ranking and surpass the limitations imposed by the Greek Diavgeia Programme when publishing data and metadata. It provides more flexibility and can orientate the end user towards more custom-based approach depending on Use Cases of interest.

The benefit is evidently the scalability of the ontology and its characteristic to serve as a good practice for similar efforts undertaken at national level to encode publicly available data in the form of administrative acts. Nevertheless, it remains open and challenging to further exploit all available object and data properties of the imported ontologies and match them to additional pieces of information included in government decisions and acts. It would be also interesting to investigate whether a new set of terms could be put together in the fashion of a new controlled vocabulary matched to user needs and conforming to EU standards. Moreover, the investigation of integrating additional ontologies widely used in the field of EU Public Administration remains a challenge. In this manner, additional Use Cases could be formulated to drive knowledge extraction, via appropriate developed Competency Questions.

One could also plan to actively involve the actors in the field, public servants and administrators, to collaborate in a systematic and regulated manner in this effort to identify actual valuable knowledge. This could also build up on the developed solution to provide an automated tool to encode the decisions/acts directly in a user friendly application tool to be systematically used by public servants so as to ensure that the use of Linked Open Data of high quality is promoted. The counter-benefit would obviously be the need to accommodate the transition from the existing Greek Diavgeia Programme to a new solution without the risk of losing the uploaded data.

The current effort could also be significantly promoted if automatic extraction of knowledge, in the form of RDF triples, is made possible in an automated manner via exploitation, for instance, of Natural Language Processing and Machine Learning techniques so that it can feed-in with a sufficient amount of data the developed ontology. The latter could additionally enable a validation process of the data integrity in the sense of a "Proof of Concept" procedure for the imported data. This validation process could be further enhanced and should be ideally established via the use of an automated framework in the direction of providing sound sources of information prior to exploitation [10]. Furthermore, it is imperative that used techniques are customized to the national language and more importantly to the special terminology used in the frame of the Public Administration.

References

[1] F. Bauer and M. Kaltenböck, Linked open data: The essentials, *Edition mono/monochrom, Vienna* **710** (2011).
[2] L.-D. Ibáñez, I. Millard, H. Glaser and E. Simperl, An assessment of adoption and quality of linked data in European open government data, in: *International Semantic Web Conference*, Springer, 2019, pp. 436–453.
[3] K. Bovalis, V. Peristeras, M. Abecasis, R.-M. Abril-Jimenez, M.A. Rodriguez, C. Gattegno, A. Karalopoulos, I. Sagias, S. Szekacs and S. Wigard, Promoting Interoperability in Europe's E-government, *Computer* **47**(10) (2014), 25–33.
[4] ISA, ISA² - Interoperability solutions for public administrations, businesses and citizens, 2022. https://ec.europa.eu/isa2/isa2_en.
[5] W3C, The Organization Ontology, 2022. https://www.w3.org/TR/vocab-org/.
[6] A. Gritzalis, A. Tsohou and C. Lambrinoudakis, Transparency-Enabling Systems for Open Governance: Their Impact on Citizens' Trust and the Role of Information Privacy, in: *International Conference on e-Democracy*, Springer, 2017, pp. 47–63.
[7] T. Beris and M. Koubarakis, Modeling and preserving Greek government decisions using semantic web technologies and permissionless blockchains, in: *European Semantic Web Conference*, Springer, 2018, pp. 81–96.
[8] I. Chalkidis, C. Nikolaou, P. Soursos and M. Koubarakis, Modeling and querying greek legislation using semantic web technologies, in: *European Semantic Web Conference*, Springer, 2017, pp. 591–606.
[9] ELI, ELI Ontology, 2022. https://op.europa.eu/documents/3938058/7399866/eli-diagrams.pdf.
[10] N. Stylianou, D. Vlachava, I. Konstantinidis, N. Bassiliades and V. Peristeras, Doc2KG: Transforming Document Repositories to Knowledge Graphs, *International Journal on Semantic Web and Information Systems (IJSWIS)* **18**(1) (2022), 1–20.
[11] C. Bezerra, F. Freitas and F. Santana, Evaluating ontologies with competency questions, in: *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 3, IEEE, 2013, pp. 284–285.
[12] GraphDb, GraphDB, 2022. https://graphdb.ontotext.com/.
[13] B. Lantow, OntoMetrics: Putting Metrics into Use for Ontology Evaluation., in: *KEOD*, 2016, pp. 186–191.
[14] S.C.J. Palvia and S.S. Sharma, E-government and e-governance: definitions/domain framework and status around the world, in: *International Conference on E-governance*, Vol. 5, 2007, pp. 1–12.
[15] A. Hadhiatma, Improving data quality in the linked open data: a survey, in: *Journal of Physics: Conference Series*, Vol. 978, IOP Publishing, 2018, p. 012026.
[16] D. Cheng, F. Yang, X. Wang, Y. Zhang and L. Zhang, Knowledge graph-based event embedding framework for financial quantitative investments, in: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 2221–2230.
[17] M. Vafopoulos, M. Meimaris, A. Papantoniou, I. Anagnostopoulos, I. Xidias, G. Alexiou, G. Vafeiadis and V. Loumos, Intelligent and semantic real-time process of the Greek LOD for enhancing citizen awareness in public expenditures, in: *International Conference on Web Information Systems Engineering*, Springer, 2012, pp. 808–811.
[18] A. Hasnain and D. Rebholz-Schuhmann, Assessing FAIR data principles against the 5-star open data principles, in: *European Semantic Web Conference*, Springer, 2018, pp. 469–477.
[19] A. Hasnain and D. Rebholz-Schuhman, *Assessing FAIR Data Principles Against the 5-Star Open Data Principles: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, 2018, pp. 469–477. ISBN 978-3-319-98191-8. doi:10.1007/978-3-319-98192-5_60.
[20] T. Schneider and M. Šimkus, Ontologies and data management: a brief survey, *KI-Künstliche Intelligenz* **34**(3) (2020), 329–353.
[21] P.-Y. Vandenbussche, G.A. Atemezing, M. Poveda-Villalón and B. Vatant, Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web, *Semantic Web* **8**(3) (2017), 437–452.
[22] E. Huaman, E. Kärle and D. Fensel, Knowledge graph validation, *arXiv preprint arXiv:2005.01389* (2020).
[23] A. Blumauer, H. Nagy and H. Nagy, *The Knowledge Graph Cookbook*, edition mono/monochrom, 2020.
[24] L. Ehrlinger and W. Wöß, Towards a Definition of Knowledge Graphs., *SEMANTiCS (Posters, Demos, SuCCESS)* **48**(1–4) (2016), 2.
[25] M.-L. Alvite-Díez, Linked open data portals: functionalities and user experience in semantic catalogues, *Online Information Review* (2021).
[26] N. Kertkeidkachorn and R. Ichise, T2kg: An end-to-end system for creating knowledge graph from unstructured text, in: *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
[27] D. Dessì, F. Osborne, D.R. Recupero, D. Buscaldi and E. Motta, Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain, *Future Generation Computer Systems* **116** (2021), 253–264.
[28] S.A. Theocharis and G.A. Tsihrintzis, Open data for e-government the Greek case, in: *IISA 2013*, IEEE, 2013, pp. 1–6.
[29] L. Ding, V. Peristeras and M. Hausenblas, Linked open government data [Guest editors' introduction], *IEEE Intelligent systems* **27**(3) (2012), 11–15.
[30] L.O.D. Essentials, Linked Open Data:The Essentials, 2022. https://www.reeep.org/LOD-the-Essentials.pdf.
[31] ISA, How Linked Data is transforming eGovernment, 2022. https://joinup.ec.europa.eu/sites/default/files/inline-files/D4.3.2_Case_Study_Linked_Data_eGov.pdf.
[32] LOV, Linked Open Vocabularies (LOV), 2022. https://lov.linkeddata.es/dataset/lov/.
[33] T. Kravaris, Constructing INSPIRE Compliant Ontologies for Geospatial Data, 2022. https://pergamos.lib.uoa.gr/uoa/dl/frontend/file/lib/default/data/1324504/theFile.
[34] Diavgeia, Diavgeia, 2022. https://diavgeia.gov.gr/.
[35] M. Poveda-Villalón, A. Gómez-Pérez and M.C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, *International Journal on Semantic Web and Information Systems (IJSWIS)* **10**(2) (2014), 7–34.