d2kg: 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. To implement Open Governance a crucial element is the efficient use of the big amounts of open data produced in the public domain. 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 into 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 developed d2kg ontology is accessible at: http://lpis.csd.auth.gr/ontologies/2022/d2kg/d2kg.owl.

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

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 into documents explicit semantics based on formal knowledge representations such as ontologies.

The motivation of this work is to develop a new ontology d2kg in order to:

- exploit authoritative and standardized resources to represent published decisions and acts from the Greek Programme Diavgeia[2], where Public Organizations are mandated by Law to upload government decisions and acts;
- integrate standard EU ontologies, core and controlled vocabularies following W3C recommendations to exploit publicly available Open data following Linked Data principles;
- demonstrate the integration of different resources entities into a single ontology via a Knowledge Graph based representation of government decisions and acts.

The core objective is to enhance public data re-usability and interoperability at EU level. 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. It is expected that, due to the adoption of core EU legislation via Regulations and Directives in the national law of Members States, commonly used terms in government documents can be represented by using standard ontologies and vocabularies. This implies that a universal approach could be supported in the same manner in a cross-border approach and this is the contribution that the d2kg brings.

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. Finally, Section 7 presents useful conclusions drawn and provides the baseline for future work in the field identifying potential extensions and further enhancements.

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

2. Related Work

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

- 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[6] provides access to, and guidance on the use of Linked Data and Semantic Web technologies;
- Data Europa EU[7] 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[8] 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. Governmental-Public organizations ontologies/vocabularies

Several ontologies related to public organizations have been developed to reflect the specificities of national public organizations and institutions of a given country in terms of their structure and operation.

- The Spanish Public organizations ontology[9] representing the structure of public organizations in Spain;
- The UK Parliament organisation ontology[10].

Others have focused in providing a more complex framework extending beyond the structure of an organization to their respective functions and activities.

- The French National Assembly ontology concerning the legislative procedures and functions of the actors[11];
- The US Federal Enterprise Architecture (FEA) vocabulary[12] a business-based framework for government-wide improvement developed by the Office of Management and Budget that is intended to facilitate efforts to transform the federal to a citizen-centered, results-oriented, and market-based government;
- The electronic model of Public Administration's operation using an ontology, focusing on the case of Human Resources Management in a Greek administrative unit, Region of Central Macedonia [4];
- The Australian Government Records Interoperability Framework (AGRIF, 'the Framework') a system of related semantic ontologies describing the structure, functions and activities of the Australian Government[13];

[3] http://www.legislation.gov.uk
[4] http://data.gov.uk/
[5] https://www.data.gov/
[6] http://semantic.data.gov
[7] https://data.europa.eu/en
[8] https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data
[9] http://purl.org/ctic/infraestructuras/organizacion
[10] https://ukparliament.github.io/ontologies/government-organisation/government-organisation-ontology
[11] https://lov.linkeddata.es/dataset/lov/vocabs/oan
[12] https://lov.linkeddata.es/dataset/lov/vocabs/fea
[13] https://raw.githack.com/agldwg/agrif-ont/master/agrif.html

– Vocabulary for the representation of data on agreements adopted by municipalities with other entities[14].

Nevertheless, these approaches only partially re-use standard ontologies and vocabularies components and mostly construct their own classes and properties which are frequently limited in scope to the requirements of a given organization and country. On the other hand, the best practice would be to focus instead more on re-using commonly agreed widely adopted standards at EU level towards interoperability. Next we delve into this aspect.

2.2. ISA²- core vocabularies

The "Interoperability solutions for public administrations, businesses and citizens (ISA²)" Programme, supporting the development of digital solutions enabling administrations, enterprises and citizens in Europe to benefit from interoperable cross-border and cross-sector public services[15], developed the EU core vocabularies[16]. 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[17].

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

2.3. The Organization ontology

The W3C Organization ontology[18] contributes as the main ontology for organizational structures, since it is designed to allow domain-specific extensions to add classification of organizations and roles. It is designed to enable publication of information on organizations and organizational structures including governmental organizations to provide a generic, re-usable core ontology that can be further extended or specialized. This proves fit for the purpose of information extraction from government decisions/acts to identify the main actors and contact persons.

2.4. 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 % of GDP in Europe). Therefore, the EU is investing significantly on the digitisation of the public procurement process (e-procurement). The pro-

[14]http://vocab.ciudadesabiertas.es/def/sector-publico/convenios/index-en.html
[15]https://ec.europa.eu/isa2/isa2_en
[16]https://ec.europa.eu/isa2/solutions/core-vocabularies_en
[17]https://op.europa.eu/el/web/eu-vocabularies/corevocs
[18]https://www.w3.org/TR/vocab-org/

curement 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 interrelationships. 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 for 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[19]).

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 interoperability and thus its reuse. This creates the need for a 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)[20] 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, to unify existing practices to make it easier to share, access and re-use data[21][22].

2.5. Diavgeia Programme

At national level, in Greece, a good showcase is Diavgeia ("Diavgeia" (΄Διαύγεια΄) is the Greek word for Transparency). Diavgeia is a Programme introduced by Law in 2010 obliging Public Organizations to post their decisions and acts on the Web. 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 both at national and European level [5]. 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 thousand decisions per working day [23].

2.5.1. 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 [6].

The corresponding Diavgeia ontology[24] developed incorporates elements from the distinct ontology Nomothesia ("Nomothesia" stands for Legislation in Greek) as concerns the legislation dimension [7]. 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[25]. ELI, one of the actions supported in the frame of the ISA² Programme[26], is a system to

[19] https://eur-lex.europa.eu/oj/direct-access.html
[20] https://github.com/OP-TED/ePO
[21] https://joinup.ec.europa.eu/sites/default/files/document/2017-08/d02.01_specification_of_the_process_and_methodology_v1.00.pdf
[22] https://eprocurement-everis.github.io/
[23] https://diavgeia.gov.gr/en
[24] https://github.com/ThemisB/diavgeiaRedefined
[25] http://legislation.di.uoa.gr/nomothesia.owl
[26] https://eur-lex.europa.eu/eli-register/about.html

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. Diavgeia ontology also integrates Greek Administrative Geography Ontology, a typical ontology to represent the Greek Administrative hierarchical structure[27]. The latter can be deployed to map the Public Organizations issuing decisions and acts depending on the level of administration: Organizations at regional level (Decentralized Administrations, Regions/Regional Units and Municipalities) not including though those at central level of government such as Ministries or other types of public Institutions such as Universities, Hospitals etc.

Diavgeia ontology, which is considered as directly competing to our d2kg ontology, represents an interesting use case that encapsulates diverse individual characteristics following universally adopted standards and promotes 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 documents into an ontology following more or less a strictly 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 manner, it ends up in providing a monolithic approach based on building from scratch an ontology to map decisions and acts rather than re-using available resources with the aim of adopting commonly agreed-used standards and enhancing interoperability.

Our goal is to have decisions that comply to a new integrated ontology replacing the uploading of low ranked PDF files with corresponding RDF ones. The objective is to upgrade 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 by deploying the use of EU common standards, not restricted to fit the requirements of decisions/acts developed in the frame of a single national Programme such as Diavgeia. To this end, this solution can be derived following the concept of a recently proposed intelligent framework, handling both creation and real-time updating of a knowledge graph, while also exploiting domain-specific ontology standards, deploying Diavgeia [8]. In the following section, we elaborate further on the enhancements proposed to build the new integrated ontology d2kg.

3. Methodology

The methodology we followed tries to encapsulate the basic elements of the aforementioned in the related work technologies and principles. We started off from the directly competing 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 EU level. To this end, we applied the basic principles of linked open data development according to the W3C guidelines[28] for establishing the framework for high quality linked open data, we developed our integrated ontology and subsequently built 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 interoperability.

[27] https://pergamos.lib.uoa.gr/uoa/dl/frontend/file/lib/default/data/1324504/theFile
[28] https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

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. The latter applies also for recruitment of personnel.

Use Case 2: Publicity in public spending

A use case focusing mostly on the publicity requirements related to the (pre)award procurement. 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 well-informed 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 to identify where public money is spent, i.e. to which kind of goods, works, consumables, services etc. This can be done by retrieving related Diavgeia Award decisions in order to identify the procurement type via the Common Procurement Vocabulary (CPV) values. In the same manner, one could be interested in specific information concerning personnel appointment/recruitment, such as the type/category of personnel appointed or the frequency of appointments. In terms of its internal functioning, an organization could collect data for statistical reasons (for instance, the average number or duration of public Contracts) to assess the efficiency of its organizational units.

In the Diavgeia platform knowledge extraction is currently performed in two ways: firstly, by a keyword search (either basic or more advanced using multiple criteria) based on the encoded metadata and secondly via an API to query over the metadata. In both cases metadata is encoded by public servants with no uniform approach. This fact leads to significant discrepancies between the decisions/acts textual and metadata information. Therefore, the platform does not allow the translation of use cases such as the above to inputs producing outputs suitable to the end user needs. In our work, we use Competency Questions (CQs) to translate use cases in ontology requirements.

3.2. Competency Questions

Given a set of scenarios related to the application field, developers should be able to place a set of questions stated in natural language representing users demands and translated in SPARQL queries to support the development in:

- Enabling developers to identify the main entities and their relationships to create the ontology vocabulary;
- Providing developers with a simple means to verify requirements' compliance by either knowledge retrieval or by entailment on its axioms and answers checking [9].

Therefore, we identify domains of interest and develop questions, as analysed in Section 5, 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/organizations that are recipients of awarded contracts (within a given time period), i.e. organizations receiving the highest No of contracts?
- CQ2: For a given organization, which are the awarded contracts to a specific economic operator/organization (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 (recruitments in a public organization) for a certain person (within a given time period)?

Use Case 2: Publicity in public spending

- CQ1: Which are the selection(eligibility) 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/recruited (within a given period of time)?
- CQ4a: 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?

This list of CQs is built around the most characteristic thematic categories of decisions and acts. It represents a non exhaustive list that can be further extended to map all properties and classes, customized to the needs of the end user or the organization and thus drive the knowledge extraction process. One can further elaborate on combinations of the above queries to build additional competency questions.

4. The OWL ontology for government decisions and acts

4.1. Diavgeia documents analysis

Document analysis describes the information extraction process via manual annotation to match document data to actual classes and properties to be integrated from standard ontologies and vocabularies or identify the need to create new ones based on published documents on the Greek platform Diavgeia. A reference document (Fig.1) as published on the platform is analysed to identify corresponding classes and properties. It is related to the appointment (recruitment) of personnel to a Greek Ministry. The document is broken down in three main Parts (A, B, C) according to a typical document layout. We highlight the main document elements (in Greek) representing valuable knowledge and provide next to them the corresponding actual related classes/properties from standard ontologies with their prefixes (in English) or the prefix *d2kg:* for our new ontology.

For instance, for this sample document selected (Fig.1) we can extract Organization's relatedata from Part A/upper left corner. It is issued by a *m8g: Public Organization* (class), the 'HELLENIC MINISTRY OF DEVELOPMENT AND INVESTMENTS' (΄ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝΔΥΣΕΩΝ΄ in Greek). We identify the internal hierarchical structure of the Ministry comprising a GENERAL DIRECTORATE (΄ΓΕΝ. Δ/ΝΣΗ ΔΙΟΙΚΗΤΙΚΩΝ ΥΠΗΡΕΣΙΩΝ΄), a DIRECTORATE (΄ΔΙΕΥΘΥΝΣΗ ΔΙΑΧΕΙΡΙΣΗΣ & ΑΝΑΠΤΥΞΗΣ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ΄) and a DEPARTMENT (΄ΤΜΗΜΑ ΚΙΝΗΤΙΚΟΤΗΤΑΣ ΥΠΑΛΛΗΛΩΝ΄) and we see how these are interrelated via appropriate object properties (*org:hasUnit/org:unitOf* of the Organization ontology). Further down within Part A, we

Figure 1: Diavgeia sample in focus

can extract data for the Contact Point (class *epo:ContactPoint*) ΄Πληροφορίες΄ in Greek and how this is reached via the appropriate object properties such as *epo:hasPostalAddress*, and data properties *epo:postCode*, *epo:hasTelephone*, *epo:hasEmail*, *eli:date_document*, *dvg: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, property ΄προσλαμβάνεται΄ in Greek matches the object property *d2kg:appointedIn*, property ΄Βαθμό Β΄ in Greek matches data property *d2kg:staffRank*, etc.

In the concluding part C, the classes *org:Post*, *org:Person* are necessary to identify the Signer introducing object properties such as *epo:isSignedBy*, *epo:heldBy* and data properties such as *epo:hasFullName*, *epo:hasGivenName*, *epo:hasBirthFamilyName* used to retrieve a variety of RDF triples. The document also includes on the right upper corner its unique Identifier assigned by the platform Diavgeia mapped to the data property *dvg:iun*.

4.2. Ontology built-up

Taking into account the prior analysis of a typical document we can practically fit the corresponding entities to the appropriate pieces of information within the document. This implies manual annotation for numerous documents. 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 interrelate and describe them. That means identifying which components of standard ontologies can be integrated to map the information included in the Diavgeia documents and match them. This is the common approach followed for a set of documents through the new integrated ontology built-up.

4.3. The d2kg ontology

A graphical UML representation of the main entities and their relationships of the integrated ontology d2kg is provided in Figure 2. This representation gives an outlook of the developed ontology highlighting characteristic components integrated in the current implementation, that is several of the most frequently used classes, data and object properties.

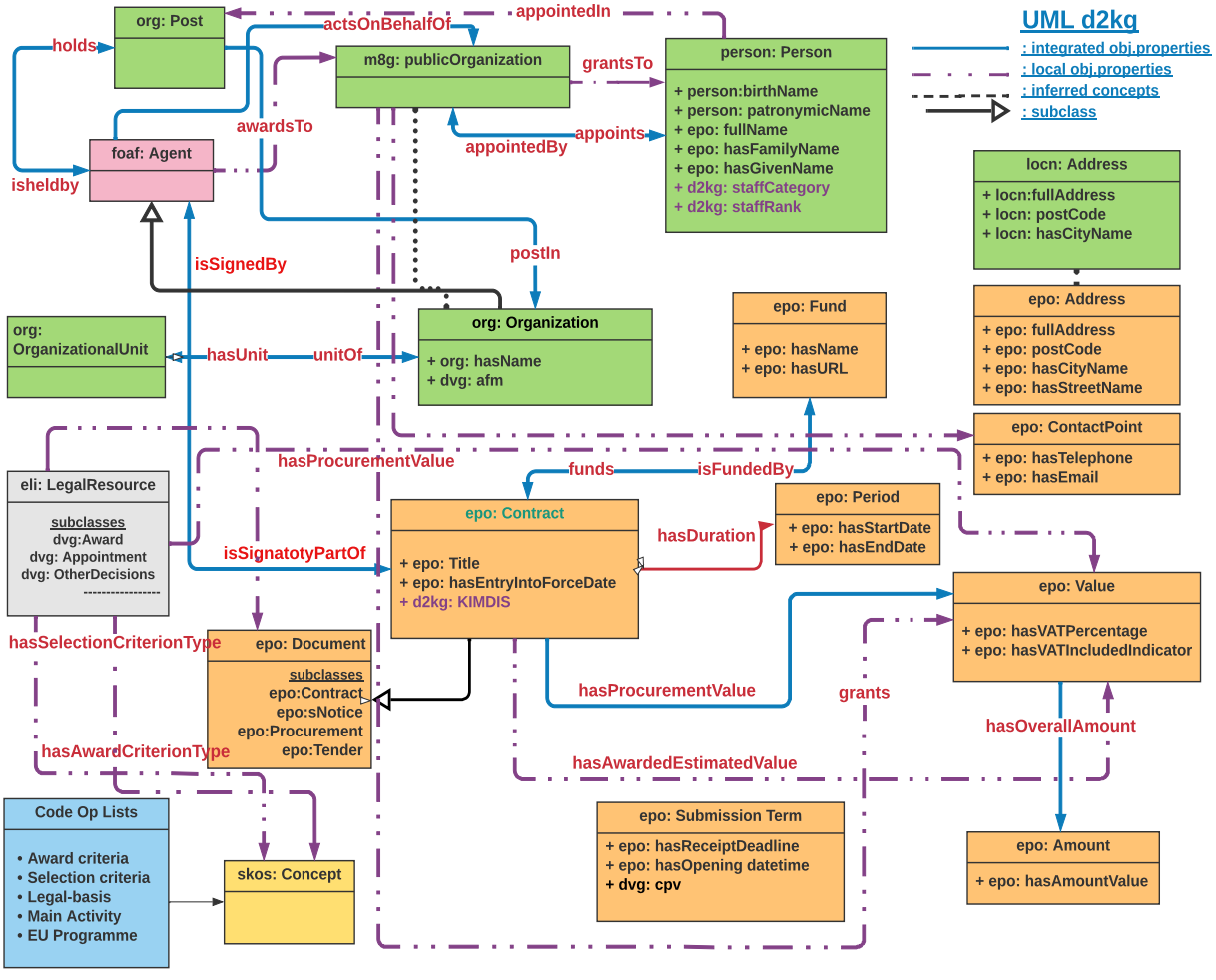

Figure 2: UML representation of the d2kg ontology

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/sub-classes of the Diavgeia ontology in grey, whereas the controlled vocabularies in

blue. The blue continuous lines show existing object properties connecting classes, whereas the purple dashed lines indicate new connections between classes established locally with the re-use of existing or new object properties and black dashed inferred concepts. Data type properties are represented via the preceding + symbol inside a class box. For instance, the property 'hasProcurementValue' connects a Diavgeia decision of type *dvg:Award* with the class *epo:Value*, whereas we can infer that *m8g:Public Organization* is also an *org:Organization*. It is evident that the focus is on the re-use of existing classes, object and data properties from the integrated 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 platform 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 content-wise.

Characteristic entities integrated in d2kg ontology per originating ontology are detailed in the next sections. We describe the re-used per ontology below, introducing the news ones we developed in separate distinct subsections. We note that we mixed components from different formats (OWL, RDF, SKOS) to address use cases (end user) requirements and in this respect it is expected that we adopt the respective different naming conventions for entities representations as well.

4.3.1. d2kg classes

Diavgeia ontology

The Diavgeia ontology (v1.0) classes are extensively analysed in the corresponding repository of the DiavgeiaRedefined Project[29]. For the sake of completeness, we refer to the basic ones used in the context of our work here:

- LegalResource: the core class representing decisions/acts of 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 of the notation of the Diavgeia ontology: Award, Contract, DeclarationSummary, DonationGrant, ExpenditureApproval, OwnershipTransferOfAssets, WorkAssignmentSupplyServicesStudies, PaymentFinalisation, GeneralSpecialSecretaryMonocraticBody, involving a financial aspect (relevant to monetary transactions) which implies the need for a separate class to encode accompanying data such as the involved parties, amount etc;

ePO ontology The epo v1.0 is used. Below there is a summarization of the main entities we re-used[30]:

- Agent: A person, an organization, or a system that act in a procurement or have the power to act in a 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);
- Period: A time interval or a duration, usually consisting of a start and an end date;
- 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.

[29] https://github.com/ThemisB/diavgeiaRedefined/tree/master/rdf
[30] https://github.com/OP-TED/ePO

Organization ontology

We re-use the following classes of the W3C Recommendation of 16.01.2014[31]:

- 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 allows us to identify the Organization structure and hierarchical relationships involved in Diavgeia docs. 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 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.

e-Government core vocabularies

The release v2.0.1 was used in the frame of our implementation[32]

Core Public Organization vocabulary

- PublicOrganization: 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 we re-used/integrated.

Diavgeia ontology

- has_expense: has expense links a certain decision type with an Expense;
- signed_by: links a Legal Resource with a Signer; there is an equivalent property in ePO ontology.

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 'isFundedBy';

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

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

- 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: the place where the tenders will be publicly opened. Important for the sake of transparency to be communicated, since candidate contractors can be present during the tenders' opening (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 contracts 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 'isSignatoryPartOf'.

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.

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 types 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) an Organization receives;
- staff: represents the personnel/staff of an Organization;

4.3.3. d2kg data type properties

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

New data type properties

It has been judged appropriate to introduce specific data type 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: Α, Β, 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 these properties. This is the point where the EU vocabularies are introduced[33].

4.4.1. Authority tables

The Authority tables[34] 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[35] provides the list of conditions 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[36].

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[37] with the list of rules to be taken into account for the award decisions. The initial values are those foreseen in public procurement directives (2014/23/EU, 2014/24/EU, 2014/25/EU)[38].

EU Programme

The EU Programme Authority Table (AT)[39] provides the list of programmes created, coordinated by and financially supported by the EU 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[40] 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[41], explicitly falling within the sectoral directive (2014/25/EU Art. 8 - Art. 14).

[33] https://op.europa.eu/en/web/eu-vocabularies
[34] https://op.europa.eu/en/web/eu-vocabularies/authority-tables
[35] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/selection-criterion
[36] https://docs.peppol.eu/pracc/espd/codelist/CriteriaTypeCode/
[37] https://op.europa.eu/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/award-criterion-type
[38] https://eur-lex.europa.eu/oj/direct-access.html
[39] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eu-programme
[40] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/main-activity
[41] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_the_functions_of_government_(COFOG)

Procurement Procedure type

This set[42] 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[43] based on the legal acts used for a given public procurement procedure, as provided by the EU Publications Office.

CPV

To make public procurement more transparent and efficient, European Commission drafted a single classification system, the Common Procurement Vocabulary(CPV)[44], aimed at standardising the references used by contracting authorities and entities to describe the subject of procurement contracts (Regulation (EC) 213/2008[45]).

4.5. URIs

The format of the URIs is differentiated depending on the actual usage of the entity. Concerning the main entity Legal Resource, we adopted the Persistent URIs approach of the integrated Diavgeia ontology to comply as much as possible with common requirements[46]. The decisions/acts are structured according to the ELI rationale template: **http://www.diavgeia.gov.gr/eli/iun/version**. Modifications of a decision result to a new URI with the same iun and a new version number. Regarding other types of entities, though, the Greek Diavgeia platform has determined a way to uniquely represent them according to an XML schema[47] including the most important data and providing an available HTTP URL to look it up. The representation per the most important entities is shown below:

- Organizations: **https://diavgeia.gov.gr/opendata/organizations/uid**, where uid stands for the Organization unique identification number;
- Person: **https://diavgeia.gov.gr/opendata/signers/uid**(uid identifying the Signer based on Diavgeia records).

For example: **https://diavgeia.gov.gr/opendata/organizations/50205** represents the Public Organization "Decentralized Administration of Macedonia-Thrace" (΄ΑΠΟΚΕΝΤΡΩΜΕΝΗ ΔΙΟΙΚΗΣΗ ΜΑΚΕΔΟΝΙΑΣ-ΘΡΑΚΗΣ΄ in Greek) and points to a URL with administrative data on this organization.

```
−<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 3: Standard XML representation of a URI for Organization in Diavgeia

[42] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/procurement-procedure-type

[43] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/legal-basis

[44] https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/cpv

[45] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R0213&qid=1646558739181&from=EN

[46] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris

[47] https://github.com/diavgeia/opendata-client-samples-java/blob/master/src/main/xsd/diavgeia.xsd

5. Case studies

The applicability of the developed d2kg ontology is 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 Diavgeia typical appointment document we used already in Section 4.1. can be visualized in the form of a Knowledge Graph via Ontotext GraphDB [48].

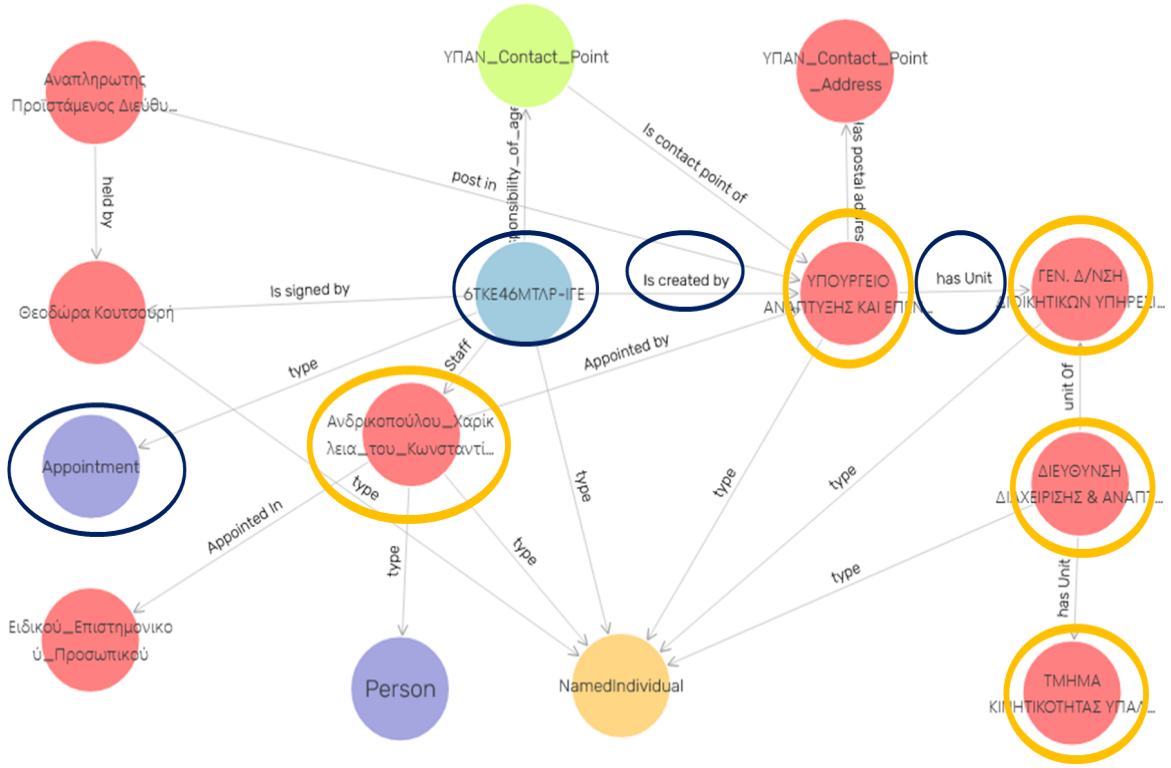

Figure 4: GraphDB visualization for the Diavgeia Sample in focus

In Fig.4 we can identify all core elements of this sample decision (properties with their labels). Indicatively:
- the corresponding decision ID (IUN): ΄6ΤΚΕ46ΜΤΛΡ-ΙΓΕ΄ in the center;
- the decision type: 'Appointment'- in purple on the left hand side;
- the Organization which creates/issues the decision and appoints the Person, right next to the decision ID in red: the 'HELLENIC MINISTRY OF DEVELOPMENT AND INVESTMENTS'(ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝ. in Greek). The object property labeled 'is created by' is used to identify the issuing Organization;
- the Organizational Units with their hierarchical structural relationship of the Ministry in the right hand side in red: ΓΕΝ. Δ/ΝΣΗ ΔΙΟΙΚΗΤΙΚΩΝ ΥΠΗΡΕ., ΔΙΕΥΘΥΝΣΗ ΔΙΑΧΕΙΡΙΣΗΣ & ΑΝΑΠΤ., ΤΜΗΜΑ ΚΙΝΗΤΙΚΟΤΗΤΑΣ ΥΠΑΛΛ in Greek;
- the Person appointed Andrikopoulou Charikleia (Ανδρικοπούλου_Χαρίκλεια in Greek) in red just below the decision ID on the left side.

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

5.2. SPARQL queries

We deploy SPARQL for a systematic and targeted extraction of knowledge, for certain 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/organizations that are recipients of awarded contracts (within a given time period)

The query provides the list of operators/contractors ranked according to the highest number of contracts awarded by the HELLENIC MINISTRY OF INTERIOR (id: 100054492) after 01.01.2017 (publication date). We identify the integration of different ontologies (epo: for E-procurement, dvg: for Diavgeia, eli: for ELI ontology).

```
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 2
```

The output (Table 1) comprises the name of the Contractors and the number of contracts they received.

Table 1: Results for UC1/CQ1

a/a	Contractor	number_of_contracts
1	Κοινωνικό_Συνεταιρισμό_Περιορισμένης_Ευθύνης_Δυτικού_Τομέα:	2
2	ΠΑΛΑΙΟΧΩΡΙΝΟΣ_ΠΑΝΑΓΙΩΤΗΣ	1

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 to have data related to the issuing organization of a decision/act. The focus is on the Core Public Organization ontology via its main entity 'Contact Point' to retrieve data of the Responsible agent issuing the act.

```
SELECT distinct ?doc ?full_name ?Email ?Telephone where {
        ?doc
                epo:isCreatedBy dvgo:100054495;
                eli:responsibility_of_agent ?Contact_Point.
        ?Contact_Point epo:hasFullName ?full_name;
                epo:hasEmail ?Email;
                epo:hasTelephone ?Telephone. }
```

Table 2: Results for UC2/CQ3

a/a	doc	full_name	Email	Telephone
1	https://diavgeia.gov.gr/eli/decision/6ΤΚΕ46ΜΤΛΡ-ΙΓΕ/1	Ι.Μανωλάτος	manolatos@gge.gr	2103893307

Use Case 3: Efficiency of the decision-making process

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

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

```
SELECT (COUNT(distinct(?doc)) AS ?number_docs) ?birthName ?Staff_Category ?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 output (Table 3) is that Mrs EVANGELIA KARKOYLA (ΕΥΑΓΓΕΛΙΑ ΚΑΡΚΟΥΛΑ in Greek) has been appointed as PHARMACIST (ΠΕ_ΦΑΡΜΑΚΟΠΟΙΩΝ in Greek) in a vacant post (κενή_οργανική_θέση in Greek) and the appointment of this specific person has taken place once since 01-01-2015.

Table 3: Results for UC3/CQ3

a/a	number_docs	birthname	StaffCategory	Post
1	1	ΕΥΑΓΓΕΛΙΑ ΚΑΡΚΟΥΛΑ	ΠΕ_ΦΑΡΜΑΚΟΠΟΙΩΝ	κενή_οργανική_θέση

6. Ontology assessment

6.1. Debugging

To assess the d2kg ontology we deploy initially Protégé, the tool used to develop it. We assess the possible faults in the ontology via the 'Pellet' reasoner and the installed 'Debugger' plug-in. The debugging shows that no faults occur during this validation process and no repairs are suggested for the ontology. The ontology proved to be both consistent and coherent. The default configuration for the reasoner is selected, which computes inferences on class, object and data property hierarchy, along with class, object and data property assertions delivering an error-free ontology.

OOPS! also provides an indicator for each pitfall, according to the classification of possible negative consequences (critical, important, minor) [10]. In the case of d2kg, OOPS! detected several pitfalls; however, all critical or important ones derived from the integrated ontologies implementation (for instance, missing domain or range in properties, missing annotations, defining wrong inverse relationships etc) and not related to our implementation.

6.2. Reasoning

Pellet and Hermit reasoners were used to retrieve additional knowledge through inference. Inferences provide an insight of the data encapsulated in Diavgeia documents. A concrete example of a Declaration Summary docu-

ment (iun:΄ΨΥ1Υ46ΜΤΛ6-ΛΩ3΄) is used. Apart from the predefined classification of the document as Declaration Summary, we infer the following class types based on domain/range constraints.

-Award Criterion: derived from the object property "hasAwardCriterion" present in the doc text and related to the value 'Cost'. Therefore, the Reasoner yields the corresponding class type 'Award Criterion' for this property.

-Contract: implying the assignment of a Contract as a result of the object property: "hasContractNature type" and its corresponding value 'Supplies', revealing the class type 'Contract'.

-Procedure Type: The procurement procedure type is identified as 'Open' through the use of the respective object property: "hasProcedureType" encoded in the document.

-Opening Term: the information on the Opening date and time is provided via the respective data property: "hasOpeningDateTime" and its corresponding value '2022-01-2022T15:00:00', helping us to identify this class type.

The inferred concepts (types) produced after Reasoning proved to be sound, complete and meaningful, since it is anticipated that for this type of act to retrieve relevant information for the type of contract that is going to be assigned, its award criterion, the type of the procurement procedure deployed to award the contract etc.

6.3. Ontology Metrics

In terms of the main ontology metrics we initially proceed with Protégé metrics. A summary is depicted in Table 4(a) below, with the figures being representative of the ontology's size. These figures show the collection of a significant number of classes, properties and a high number of axioms (since it includes the combined logical and non-logical axiom count) as a result of the integration of numerous standard ontologies and vocabularies and the high number of populated instances.

Table 4: Ontology metrics

(a) Protégé metrics

Metric	Value
Axioms	281800
Logical Axioms	21780
Class Count	205
Object Property count	361
Data Property count	333
Individual count	9852
Annotation Property count	61

(b) Ontometrics

Metric	Value
Attribute Richness	1.62439
Inheritance Richness	1.809756
Relationship Richness	053913
Average Population	48.058537
Class Richness	0.170732
DL expressivity	SROIN(D)

To evaluate also the domain coverage we deployed the online platform OntoMetrics [11]. An overview of the main metrics is presented in Table 4(b). The assessment is based on ontology's accuracy and conciseness. The first three metrics refer to the ontology's accuracy, while the other two refer to its conciseness. We note the following:

- A high attribute richness value (the average number of attributes per class, giving an indication of both the ontology design quality and the amount of information pertaining to instance data). The high value is anticipated due to the integration of numerous ontologies in a single schema;

- A good coverage in the range of concepts illustrated via the inheritance richness value (the average number of sub-classes per class describing the distribution of information across different levels of the ontology's inheritance tree). This measure can differentiate a horizontal ontology (where classes have a large number of direct sub-classes) from a vertical ontology (where classes have a small number of direct sub-classes). In our case, a balance is achieved to cover on the one hand good range of concepts without, on the other hand, going into a highly detailed analysis of their sub-traits ;

- A balanced relationship richness close to 50% (the ratio/percentage of the number of (non-inheritance) relationships divided by the total number of relationships which reflects the diversity of the types of relations). An ontology containing only inheritance relationships conveys less information than an ontology that contains a diverse set of relationships.

- A satisfactory value for the average population. Formally, the average population is defined as the number of instances of the knowledge base divided by the number of classes defined in the ontology schema, providing an indication of the ontology population quality;

- A low 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 ontology. This property is related to how instances are distributed across classes. The number of classes that have instances in the knowledge base is compared with the total number of classes to show how well the knowledge base utilizes the knowledge modelled by the schema classes (a low value implies the knowledge base does not have data that exemplifies all the class knowledge, whereas a high value would indicate that the data represents most of the knowledge in the schema);

The DL expressivity is classified as SROIN(D), derived via both Ontometrics and Protégé, which indicates the support of Transitivity, Role chains, Nominals, InverseOf and Cardinality restrictions. The d2kg ontology corresponds to the OWL 2 DL profile.

7. Conclusions and Future Work

This paper focuses on the development of the d2kg integrated ontology appropriate to represent government decisions and acts. The Greek Programme Diavgeia was selected as a representative case study since it is a public administration repository of substantial amounts of governmental/administrative documents. The ontology d2kg was built to highlight the capacity to represent public domain data as interrelated Linked Open Data that can be easily exploited to extract knowledge. Characteristic classes and properties of the standard EU core and controlled vocabularies following W3C recommendations were used together with new ones in the d2kg ontology to demonstrate how to overcome the shortcomings of publishing open data in formats, such as PDF, of lesser value and quality ranking and surpass the limitations imposed by the Greek Diavgeia platform that requires data and metadata to be encoded in a rigid manner. The ontology allows targeted extraction of knowledge from formal documents issued by the Greek public administration via numerous use cases. Furthermore, it is highlighted how underlying data in documents published in Diavgeia platform can feed and enrich effectively a Knowledge Graph.

In terms of future work, the d2kg ontology itself can be extended to exploit other available object and data properties of the integrated standard ontologies, matching them to additional pieces of information included in government decisions and acts. Moreover, the investigation of integrating additional ontologies widely used in the field of EU public administration remains a challenge. 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 conforming to EU standards, so as to be able populate the ontology with instances containing sets of predefined standardized values. This ontology serves, additionally, as a good practice for similar efforts that could be undertaken at national level of other EU member states to encode publicly available data in administrative documents. In conjunction with "ISA²-Interoperability solutions for Public Administrations, businesses and citizens" and other collaborative efforts, there is significant potential of promoting knowledge creation in the area of e-Governance abiding by best practices for interoperability and re-usability.

One could also plan to actively involve the actors in the field, public servants and administrators, to collaborate in a systematic and regulated manner to identify actual valuable knowledge. We could 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 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 [8]. 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] I. Savvas, N. Bassiliades, K. Kravari and G. Meditskos, An Ontological Business Process Modeling Approach for Public Administration: The Case of Human Resource Management, in: *Handbook of Research on E-Business Standards and Protocols: Documents, Data and Advanced Web Technologies*, IGI Global, 2012, pp. 725–753.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] B. Lantow, OntoMetrics: Putting Metrics into Use for Ontology Evaluation., in: *KEOD*, 2016, pp. 186–191.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] T. Schneider and M. Šimkus, Ontologies and data management: a brief survey, *KI-Künstliche Intelligenz* **34**(3) (2020), 329–353.

[19] 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.

[20] E. Huaman, E. Kärle and D. Fensel, Knowledge graph validation, *arXiv preprint arXiv:2005.01389* (2020).

[21] A. Blumauer, H. Nagy and H. Nagy, *The Knowledge Graph Cookbook*, edition mono/monochrom, 2020.

[22] L. Ehrlinger and W. Wöß, Towards a Definition of Knowledge Graphs., *SEMANTiCS (Posters, Demos, SuCCESS)* **48**(1–4) (2016), 2.

[23] M.-L. Alvite-Díez, Linked open data portals: functionalities and user experience in semantic catalogues, *Online Information Review* (2021).

[24] 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.

[25] 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.

[26] S.A. Theocharis and G.A. Tsihrintzis, Open data for e-government the Greek case, in: *IISA 2013*, IEEE, 2013, pp. 1–6.

[27] L. Ding, V. Peristeras and M. Hausenblas, Linked open government data [Guest editors' introduction], *IEEE Intelligent systems* **27**(3) (2012), 11–15.