

Public spending as LOD: the case of Greece

Michalis Vafopoulos, Marios Meimaris, Ioannis Anagnostopoulos, Agis Papantoniou, Ioannis Xidias, Giorgos Alexiou, Giorgos Vafeiadis, Michalis Klonaras and Vasilis Loumos

*School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece*

Abstract. The PSGR project is the first attempt to generate, curate, interlink and distribute daily updated public spending data in LOD formats that can be useful to both expert (i.e. scientists and professionals) and naïve users. The PSGR ontology is based on the UK payments ontology and reuses, among others, the W3C Registered Organization Vocabulary and the Core Business Vocabulary. RDFized data are linked to product classifications, Geonames and DBpedia resources. Online services contain advanced search features and domain level information (e.g. local government), simple and complex visualizations based on network analysis, linked information about payment entities and SPARQL endpoints. During February 2013, the growing dataset consists of approximately 2 mil. payment decisions valued 44,5 bil. euros forming 65 mil. triples.

Keywords: PSGR, public expenditure, Open data, Big data, Registered Organization Vocabulary, economic LOD.

1. Introduction

The publicspending.gr project (PSGR) is a research effort that has been initiated in 2011 by the Multimedia Technology Laboratory, School of Electrical and Computer Engineering at the National Technical University of Athens, in order to generate, curate, interlink and publish daily updated economic data in LOD formats that are accessible by both the scientific community and mass audience. The online service includes advanced search features and domain level information (e.g. local government), simple and complex visualizations based on network analysis, linked information about payment entities and SPARQL endpoint provision. To the best of our knowledge, PSGR is the first project worldwide to undertake the above tasks in such extent.

The remainder of the paper is as follows. Section 2 refers to the name, URL, versioning, licensing, availability of the dataset, while the next section is about the source for the data and topic coverage. The purpose of the linked dataset is discussed in section 4. Section 5 states the applications using the dataset and other metrics of use. Creation, maintenance and update mechanisms as well as policies to ensure sustainability and stability are analysed in the section 6. The next section is about domain modeling and use of established vocabularies, while section 8 refers

to the quality, quantity and purpose of links to other datasets. Examples and critical discussion of typical knowledge modeling patterns used are provided in section 9. Section 10 refers to known shortcomings of the dataset and the last discusses the way forward.

2. Name, URL, licensing and availability

The PSGR dataset contains information about the public expenditure which originates from the CI@rity program¹ API and business information provided by official Tax Information System (TAXIS)². The RDFized information that is obtained in daily scale from the aforementioned sources consists of approximately 2 mil. payment decisions valued 44,5 bil. euros that have been paid from 3,900 payers to 204,000 payees and form 63 mil. triples (Table 1).

	#decisions	amount (€ bil.)	#triples (mil.)
2010	89K	0,9	2,4
2011	1,031K	12,4	27,9
2012	1,978K	39,3	55,3
2013	2,141K	44,5	63,1

Table 1: aggregated data of PSGR from October 2010 until February 2013.

¹ et.diavgeia.gov.gr/en

² asis.gr

The dataset is publicly available from the publicspending.gr website³. Moreover, at the same domain, a customized SPARQL endpoint⁴ with predefined queries for beginners is offered. For advanced users there is a standard SPARQL endpoint⁵ available. Furthermore, updated dumps of the dataset, divided by quarters are offered in the form of RDF/XML⁶. Data are available under the Creative Commons - Attribution 3.0 license.

3. Source for the data and topic coverage

3.1. *The Clarity program: a major transparency initiative*

Beginning October 1st, 2010, all government institutions are obliged to upload online their decisions with special attention to issues of national security and sensitive personal data. Each document is digitally signed and assigned a unique number certifying that the decision has been uploaded. Furthermore, the decisions can not be implemented if they are not uploaded on the Clarity portal, which introduces an unprecedented level of transparency within all levels of the Greek public administration. Its main objectives concern: (a) the safeguarding of transparency of the governmental actions, (b) the elimination of corruption by exposing it more easily when it takes place, (c) the observance of legality and good administration, (d) the reinforcement of citizens' rights, such as the participation in the Information Society, (e) the enhancement and modernization of existing publication systems of governmental decisions and (f) the availability of all governmental decisions in formats that are easy to access, navigate and comprehend by citizens.

The most important innovation of the program is the combination of closely interrelated legal choices, operational processes and technological strategies. The technological implementation model that is based on an agile strategy of "open content" enacts the dissemination and re-use of Public Sector Information (PSI), providing the necessary tools for thorough access to it. Users are able to "build" applications with added value using the program's content so as to enhance extra functionality. The

³ publicspending.medialab.ntua.gr

⁴ publicspending.medialab.ntua.gr/en/endpoint

⁵ publicspending.medialab.ntua.gr/sparql

⁶ datahub.io/dataset/public-spending-in-greece

REST-like Open Data API provides access to all decisions and supplementary information issued by the Greek Public Authorities. The data provisioning is compliant with a specific Transparency Law and are available under the Creative Commons - Attribution license (CC BY 3.0 GR). The latest version of the API supports a taxonomic structure of its content that is returned by default in XML format but JSON, atom and RSS are also available.

3.2. *Greek Tax Information System*

The second data source is TAXIS that handles citizens' and legal entities' taxation-related information. TAXIS also provides a web service for querying business entities. The service utilizes the Web Service Definition Language (WSDL) and querying is performed in the form of SOAP calls with the entity's VAT registration number as the reference key. The response contains metadata about the business entity, including contact details, activity descriptions, registration dates and current operational status. Business activities are described with the use of the Classification of Products by Activity (CPA) vocabulary. Within the scope of PSGR, the web service is used for querying legal entities on their first appearance as payment agents and the response data are RDFized and stored as payment agent metadata.

In addition, the Greek Tax Authorities during 2012 have started to publish a list of the biggest debtors to the Greek government, including both physical and legal entities.

3.3. *Topic coverage*

The data from Clarity contain detailed information about payers, payees, payment categories, as well as descriptive data related to the spending action (e.g. amount, official decision documents etc. [14]). The data retrieved from TAXIS create an underlying business register layer providing information that uniquely identifies agents as business entities. These include data such as formal names, addresses, legal status descriptions and contact details.

4. Purpose of the Linked Dataset

The purpose of the dataset is to enable consumption of the provided data by interested parties, either passively or actively. Passive consumption refers to

parties that seek to explore the data in intuitive ways, such as graphs, charts and statistics. With this motivation, a series of daily-updated visualizations and verbalizations are available. We have developed and deployed several applications on the basis of this dataset, the functionality of which is described as follows.

Visualizations: time-series graphs of aggregate payments, bar and pie charts of top 10 measurements of aggregate spending by dimension (e.g. payers, payees and CPV⁷) and by timeframe (day, week, month, year and overall).

Search/Advanced Search functionality: users can search through several dimensions such as payment agents, individual payments, geographical areas, signers and so on, and access static descriptions such as business information, as well as dynamically updated information which includes spending data, linked data mashups (e.g. dbpedia). However, each of these dimensions carries a further and distinct breakdown of characteristics. For instance, payment agents can be searched based on their name, legal form, geographical location, classification of activity and so on, while individual payments can be searched in combination to minimum or maximum payment amount, a range of date and specific procurement category. For this reason, each case was treated separately at the design of the advanced search interface. The relevant queries provide mashups that are in turn visualized in the form of aggregate profiles⁸.

Domain level information: focused analysis has been conducted in the spending domains of local government, education, physical persons, companies, health and energy. Statistics and visualizations are presented in the portal for each domain separately.

Aggregation of spending data (in postal code level): users can see public spending information related to payment agents, either payers or payees, that are registered to operate professionally in the postal code area the user is located.

Crowdsourced payment annotation: it is also provided the ability of discovering and annotating individual payments. Specifically, when a description of a particular payment is shown, the user can indicate errors in the payment metadata, comment and suggest appropriate corrections. These are

usually data entry errors but can also be logical errors, as will be discussed in section 10.

Active consumption refers to educated usage of the data from developers, researchers, policy makers and professionals in order to create novel results, mashups and applications. For these purposes, we have deployed the SPARQL endpoint and semestrial RDF dumps.

5. Applications using the dataset and other metrics of use

The PSGR dataset is directly re-used as data feeds in third party websites and as an input to research (see for instance [4, 6, 7]), data journalism, blogs and political discussions. Innovative visualizations and an important use of PSGR in the official policy assesment are briefly presented.

Visualizations: [16] argue that public spending data represent a payment network, formulated by the payments coming from public agents (payers) to payees. Nodes are either payers or payees, linked through a payment that is uniquely characterized by its ID, amount, timestamp and category (Figure 1).

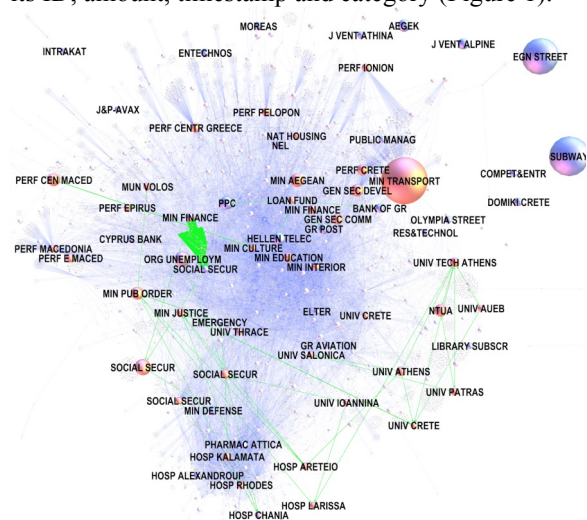


Figure 1: the network of public spending in Greece from October 2010 until February 2013.

Policy assesment: few months after its official launch, the PSGR project attracted the attention of government authorities. In this context, at the time of the writing of this paper, the operator of the Clarity program has already assigned to the PSGR team various tasks in respect to operations optimization i.e. data quality improvements, usage monitoring reports, statistics on dataset slices and custom reports. The

⁷ Common Procurement Vocabulary defines the subject of contracts according to http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm.

⁸ For instance, the profile of Chios is <http://goo.gl/xbfvj>.

outcome of this bi-directional collaboration has proved to be of an important added-value to both parties. First, the official operator took into account the most of our suggestions with regards to the data architecture in the ongoing development of the next version of the program and second, the dataset itself has become even more mature, stable, structured and functional.

6. Creation, maintenance and update mechanisms as well as policies to ensure sustainability and stability

Generation of triples is done using a combination of external data sources, as has been briefly discussed. Each data source is treated separately, as not all types of data exhibit the same degree of dynamicity, the same need for maintenance or the same periodicity of creation. For instance, geopolitical entities, such as regions, do not change and thus their resource representations do not require particular curation, contrary to business registry entities which are updated in an unpredictable way. Decisions are continuously being added to the Clarity program, but the backend mechanism of the PSGR project adds new decisions once a day for load balancing.

Information about businesses is added on first appearance and checked for updates regularly. Ontological schemas and vocabularies are treated and maintained separately. All methods are written in Java and triple models are generated and uploaded using the Jena framework [3]. There are two main types of method invocation for content generation: periodical and on-demand. Periodical generation of Linked Data refers to the process of retrieving and converting data in a timely fashion and following a strict schedule (e.g. new decisions every day), while on-demand generation refers to generating triples from static data whenever appropriate (e.g. maintaining geopolitical resources, CPV/CPA resources, etc.). In the case of routine invocation, requesting, validating and cleaning data, as well as converting to RDF and uploading them to the dedicated triple store is packed in several Java methods that are implemented in a procedural routine way at the beginning of every day in six steps:

1. Decision retrieval. Decisions that are submitted in the Clarity program are retrieved using a RESTful API. Users can request full descriptions of decisions based on several search parameters, such as date of submission, issuing organization, final signer etc. in

XML, JSON, RSS or atom. Not all formats return full descriptions of decisions, and for this reason the created methods handle XML and JSON formats. A mechanism has been deployed that selects between the two, based on current server responsiveness and availability.

2. Response parsing. The responses are parsed and stored in temporary Java class objects that were written to represent individual decisions. Each field value from the response is validated and treated separately during runtime, specifically: (a) string literals are cleaned up from encoding errors and unprintable characters, (b) number literals are formatted and decorated appropriately, (c) errors in number punctuation are identified and corrected and (d) entities are given URIs within the PSGR domain. URI minting is based on uniqueness, and most of the entities have been assigned unique IDs by the data sources they stem from. These include legal entities, individual payment decisions and so on.

3. Triple model creation. A triple model is created using Jena. Within the model, the retrieved information is given structure with the use of appropriate mapping rules to the PSGR ontologies and vocabularies.

4. Business registry information retrieval. For each payment agent (either payer or payee) that appears in the triple store for the first time, their VAT registration number is used to form a request, which is in turn sent to TAXIS web service for legal entities and freelancers. This returns as XML, the business information associated with the particular agent. The response is parsed and cleaned-up and the appropriate triples are created and stored in the model. For names and addresses, the string values are transliterated from Greek to English using the ISO 843⁹. The agents are linked to CPA and geographical resource URIs accordingly.

5. Triple upload. The resulting triple model is uploaded to the triple store using the Virtuoso Jena provider methods. These act like a bridge between a local Jena runtime environment and a remote Virtuoso store.

6. Calculation of statistics. At the end of the routine, several measures are derived that are used as references to the dataset's qualitative and quantitative characteristics. These include aggregated amounts of payments by timeframe, number of payers and payees, number of decisions etc. These are stored in a dedicated named graph for statistics.

⁹www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=5215

In the case of generation from static sources, the process is similar to the above, but does not require the same degree of automation, as there is always some control over the process and its execution. The appropriate methods are invoked on demand and the generated data are uploaded after being checked for the consistency in a manual way.

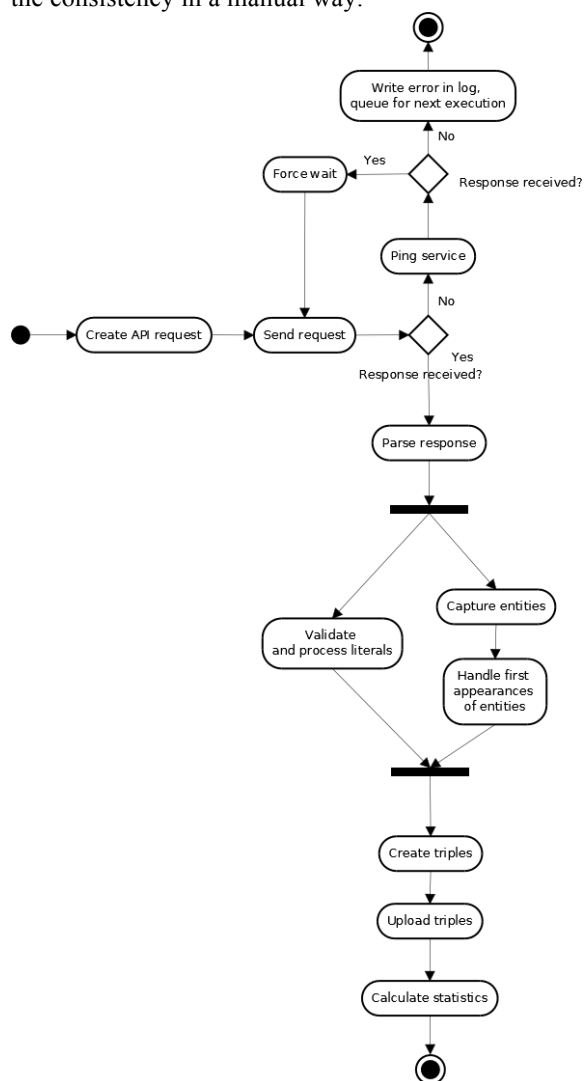


Figure 2: UML activity diagram of the daily routine

7. Domain modeling and use of established vocabularies

Essentially, three distinct ontology schemas have been deployed, as the data were distinguished in three domains, namely spending, organizations and geography. The domain modelling process of each

one will be described in brief in the following, mainly focusing on the *spending ontology* as it forms the backbone of the whole model schema.

The PSGR spending ontology has been designed as an extension of the UK Payments ontology¹⁰, adjusted for the Greek public domain. Public decisions are represented by the *psgr:Decision* core class, instances of which are associated with the decision metadata extracted from Clarity, following a similar approach to the UK ontology. Such metadata include dates, subject descriptions, submission timestamps etc. These metadata apply to all different types of decisions. Since payments are not the only decision type available, we adopt an open-world approach on how decision instances are associated with their type.

A payment object holds enough semantics to be considered as a separate entity. It represents a payment transaction that is validated by an appropriate decision and it is associated with a decision instance via the *psgr:refersTo* property. Therefore, while a payment instance holds information on the payment itself, its associated decision object holds metadata that is shared among different types of decisions. This approach aims at extendibility and scalability, given that other types of decisions will be added to the dataset in the future. Payment instances are represented with the *psgr:Payment* class, which contains information that relates a payment with a payer and a payee, as well as a CPV code and a net payment amount. In order to retrieve the payer and the payee associated with a particular payment instance, the *psgr:payer* and *psgr:payee* properties must be used. Furthermore, the *psgr:paymentAmount* property yields the total amount of the payment (in €), typed as an *xsd:double* literal. CPV codes are associated with a particular payment instance via the *psgr:cpv* property.

It is important to note that since payers, payees and CPV codes refer to resources (as opposed to a payment amount which, in principle, is a literal), querying for these will return the URIs associated with the particular resources, which then have to be further described in order for the information to be human-consumable. The ontology can be accessed from the web from its IRI.

Concerning the organizations, *psgr:PaymentAgent* is the main class in PSGR ontology. All payers and payees fall into this class and are identified uniquely by their VAT registration number. However, apart from the agents' unique identification, name and

¹⁰ data.gov.uk/resources/payments

association with payment decisions that are drawn from Clarity, we also use external sources (TAXIS) to draw information that is independent from payments.

Among business entities, the main classification factor is their legal form. A lot of work has been done for creating reference vocabularies for business entities, but these are generic as countries use their own legal classifications. This being exactly the case of PSGR as all the types of legal entities retrieved from TAXIS are plain string literals. To this end an organizations ontology¹¹ was designed in order to reflect the legal reality for business entities in Greece also providing mappings with the Core Business Vocabulary (CBV)¹² and the W3C vocabulary for registered organizations¹³ at the ontological level in the form of class subsumption.

Within PSGR, business entities are represented as formal organizations, as defined in the W3C vocabulary for organizations. However, as there exists a distinction between physical and non-physical (i.e. legal) business entities, this issue is covered by differentiating between the two with the use of appropriate classes [11]. It should be noted that a physical entity is no different than the *foaf:Person* class of the FOAF vocabulary [5]. Furthermore, the class *LegalEntity* is designed as equivalent with *Legal Entity* from the CBV. The main types of business entities have been modeled for Greece with their English labels. Mappings from literals to classes were performed with the use of Jena and some basic natural language processing methods on the string descriptions of the legal forms.

Last but not least comes the GeoData ontology¹⁴. Within the PSGR dataset each payment agent is associated with a postal code. Given that, a class that describes an area which is defined by its postal code is created. Postal code areas belong to broader areas which can be municipalities, regional units, regions and the country. This pattern created the need for an ontology that could efficiently describe it based on existing geospatial ontologies. All the classes of the PSGR ontology are interconnected via the *psgrGeo:isPartOf/hasPart* property. Each class provides a simple and generic set of properties in order to describe the area that is referred to. These sets of properties include names, postal codes, LAU/NUTS¹⁵ level, latitude and longitude. Since the

¹¹ publicspending.medialab.ntua.gr/organizationsOntology.owl

¹² joinup.ec.europa.eu/asset/core_business/description

¹³ www.w3.org/TR/vocab-org/

¹⁴ publicspending.medialab.ntua.gr/geoOntology.owl

¹⁵ geovocab.org

ontology only describes the structure that is mentioned above, we connected our dataset, using the *owl:sameAs* property, with DBpedia [2] and Geonames in order to further extend it.

8. Quality, quantity and purpose of links to other datasets

As has been mentioned, each payment is associated with a particular CPV and each business agent is assigned a unique CPA by the Greek taxation authority, which is available in the PSGR dataset. However, procurement classifications do not follow a unique standard. For instance, the European Union uses the CPV, while the United States follow the North American Industry Classification System and the United Nations the Standard Products and Services Code. The MOLDEAS project [1] is an effort to align this plethora of classification systems and publish them as RDF with controlled semantics, following SKOS [13] modelling principles. URIs representing CPV and CPA codes in our dataset are linked to their corresponding codes within the MOLDEAS dataset, paving the way to comparisons on different classifications of spending (Table 2).

	MOLDEAS	Geonames	DBpedia
number	9,454	402	804
purpose	product classifications	geonames resources	payment agents

Table 2: number and purpose of PSGR interconnections until February 2013

The second area of interlinking refers to geopolitical information that is associated to business register entries. Given that each business agent is associated with geospatial information (in the form of associations with postal code areas), we have identified and linked the relevant geopolitical resources (regions, regional units, municipalities, country) with their representations in DBpedia and Geonames (Table 2). In particular, we fulfilled the mapping between postal codes and postal code areas (i.e. geopolitical areas uniquely defined by postal codes in a 1:1 fashion) in order to create postal code area resources and, in turn, link to geopolitical resources at different hierarchical levels. This enables the association of spending and business register data with resources that exist within DBpedia and Geonames and to form queries such as finding statistical information of spending by area, or spending by population become trivial. Furthermore,

as we traverse the DBpedia graph we can associate the PSGR dataset with a large amount of external resource nodes with a linking path connecting them semantically, giving answers to potentially interesting queries. Currently, we are building direct connections of the payment agents to Greek [10], English and Live DBpedia [8]. The linking is done in the geoinformation level since we decided not link the authority, but the geopolitical region that the authority serves (e.g. distinction between the authority “Municipality of Athens” and the region that is governed by this authority). This procedure has been completed for local government authorities and is ongoing for the payees coming from the private sector. Already, they have been identified more than 1000 private agents having a lemma.

In addition, we initiated the linking with data related to the Greek economy. Local Tax Authorities during 2012 have started to publish an updated list of the biggest debtors to the Greek government, including both physical and legal entities. However, the data was published as plain HTML with no actual URIs or web resource identifiers of any sort. For this reason, we had to mint our own URIs and generate triples to represent this information in our dataset. A cross search between payees and debtors of the public concluded that 416 entities are both payees and debtors to the public during February 2013¹⁶.

9. Examples and critical discussion of typical knowledge modeling patterns used

Being “pragmatic” within the context of the aforementioned methodologies, in other words following simple approaches as compared to “heavy” ones, commonKADS [12] being such an example, was the chosen path. Adopting “heavy” ontology engineering within LOD initiatives, not only hinders Web Science [9, 15] driven researchers to provide mere outcomes but it is the authors’ belief that will affect the performance, scalability and flexibility of the underlying schemas.

However, a strong motivation for economic open data is for their respective datasets to be able to support interoperability, not just between authorities in the same country, but at an international level as well, so that comparisons can be feasible and meaningful. This means that the established vocabularies and data should be structured in ways that are descriptive enough to support the linkages.

For instance, there is no unambiguous 1:1 matching of legal entity types between countries.

Registered Organization Vocabulary¹⁷ is a higher level vocabulary, but perhaps too high for this purpose, as its expressiveness is limited to representing organizations that are recognized by some official authority, and registered within it. An interesting research topic for future work is to create a mapping schema of close matches between different types of legal entities, or a vocabulary that aggregates most commonly used close-matching legal entity types from different countries under higher-level superclasses (e.g. a generic class to represent all company types that bear close resemblance, such as PLC, SA, A.E.). Without this step, there is no direct way of bringing companies from different countries down to a common denominator, in order to make useful comparisons.

10. Known shortcomings of the dataset

Since the first initiation of the Clarity project, it is estimated that more than 30K public organizations in Greece use the web application to upload every single decision, which is digitally signed.

Throughout the investigation we noticed that there are several errors derived mainly in the data entry phase due to the lack of appropriate restrictions and validation mechanisms. Serious limitations have to do with *data format errors* (e.g. numbers as strings), *incorrect entries* (e.g. VAT or payee name) and *missing values* (mainly CPV codes).

Having pointed the major data format errors, we have developed a series of automatic error correction mechanisms. As a result nearly 200K and 4K decisions with erroneous VAT entries and CPV entries respectively, have been fixed. These corrections reflect a total amount of expenditures that reaches the level of 10 billion euros.

Syntactic or semantic types of errors mainly drive incorrect entries. Syntactic errors on the VAT or payee name are corrected by calling the TAXIS service. Semantic errors are much harder to handle since they involve manual detection and correction. Since the launching of PSGR, it was noticed that for several physical entities (individuals), the amount of received payments was enormous and in the majority of the cases unreasonable. A closer investigation revealed a significant problem, which arose due to an erroneous procedure that many users follow during

¹⁶ See the relevant SPARQL query results at bit.ly/14hrhx8.

¹⁷ www.w3.org/TR/2013/WD-vocab-regorg-20130108/.

the data entry. For being more specific, let us consider the case where a social security authority (payee) receives X euros in order to provide the welfare allowance for Y beneficiaries. Each one of the Y beneficiaries has a distinct VAT registration number. Nevertheless, in most of the cases in the Clarity program, a sole physical entity that probably is the legal representative of the payee or even in some specific cases a random representative out of all beneficiaries (usually the person whose name appears first in alphabetical order) is stated erroneously as the corresponding payee. This procedure occurs as a “solution”, mainly because there is no multiple VAT registration number provision to such kind of payments. As a result, some physical persons along with their VAT registration numbers are “loaded” with huge payments, which are totally irrelevant with their income. Trying to provide a solution, we followed a semi-automatic procedure that helped us clean the respective data. Through SPARQL queries we gathered the top 2,000 physical persons (records) who were labelled as payees from the beginning of the Clarity program up to the end of September 2012. Then for each separate record, we reviewed and cleaned their top 10 payments. The total amount of updated payments reached the level of 550M euros.

11. To the future

Having stabilized the daily update process, our main ongoing effort has moved to interlink with foreign datasets (e.g. US and UK) in order to enable cross-country comparisons in the public expenditure domain.

12. References

- [1] ALVAREZ, J. and LABRA, J. 2012. Towards a pan-european e-procurement platform to aggregate, publish and search public procurement notices powered by Linked Open Data: the MOLDEAS approach. *International Journal of Software Engineering and Knowledge Engineering*. 22, 3 (2012), 365–383.
- [2] Auer, S. et al. 2007. Dbpedia: A nucleus for a web of open data. *The Semantic Web*. (2007), 722–735.
- [3] Carroll, J. et al. 2004. Jena: implementing the semantic web recommendations. *13th international World Wide Web conference* (2004).
- [4] CHARBONNEAUX, J. and GKOUSKOU-GIANNAKOU, P. 2013. Le journalisme de “données”, une pratique d’investigation? Discours allemand et grec en regard: routines renouvelées, propos identitaire pérennisé? *Mejor, Natal (Rio Grande do Norte), 7-10 mai 2013* (2013).
- [5] FOAF Vocabulary Specification 0.98: 2010. <http://xmlns.com/foaf/spec/>. Accessed: 2013-03-14.
- [6] Galiotou, E. and Fragkou, P. 2012. Applying Linked Data Technologies to Greek Open Government Data: A Case Study. *The 2nd International Conference on Integrated Information* (2012).
- [7] Greek Legal Entities Faceted Browser: <http://linkeddata.ihu.edu.gr:8080/rdp-browser/>.
- [8] Hellmann, S. et al. 2009. DBpedia Live Extraction. *On the Move to Meaningful Internet Systems: OTM 2009*. R. Meersman et al., eds. Springer Berlin Heidelberg. 1209–1223.
- [9] Hendler, J. et al. 2008. Web science: an interdisciplinary approach to understanding the web. *Communications of the ACM*. 51, 7 (2008), 60–69.
- [10] Kontokostas, D. et al. 2011. Towards linked data internationalization-realizing the greek dbpedia. *ACM WebSci’11, June 14-17 2011, Koblenz, Germany* (2011), 1–4.
- [11] Papantoniou, A. et al. 2012. A Methodological Approach on Linked Open Data Production: The Publicspending.Gr Project. *Social Science Research Network Working Paper Series*.
- [12] Schreiber, G. and Wielinga, B. 1994. CommonKADS: A comprehensive methodology for KBS development. *IEEE Expert*. 9, 6 (1994), 28–37.
- [13] Skos simple knowledge organization system primer: 2008. <http://www.w3.org/TR/2008/WD-skos-primer-20080829/>. Accessed: 2013-03-02.
- [14] Vafoopoulos, M. et al. 2012. Publicspending. gr: interconnecting and visualizing Greek public expenditure following Linked Open Data directives. *USING OPEN DATA policy modeling, citizen empowerment, data journalism* (2012).
- [15] Vafoopoulos, M. 2011. The Web economy: goods, users, models and policies. *Foundations and Trends® in Web Science*. 3, 1-2 (2011), 1–136.
- [16] Vafoopoulos, M.N. et al. 2013. “Storytelling” in the Economic LOD: The Case of Publicspending.gr.