



## **Report on the implementation of a LOGD pilot with the Greek public administration**



Date: 29-06-16



**This study was prepared for the ISA Programme by:**

*PwC EU Services*

**Disclaimer:**

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

## Table of Contents

<b>1.</b>	<b>INTRODUCTION .....</b>	<b>1</b>
1.1.	CONTEXT: PUBLICATION OF ORGANOGRAMS OF THE GREEK PUBLIC ADMINISTRATION.....	1
1.2.	OBJECTIVES .....	1
1.3.	STRUCTURE .....	1
1.4.	GLOSSARY.....	2
<b>2.</b>	<b>PILOT DESCRIPTION .....</b>	<b>3</b>
2.1.	BUSINESS NEED .....	3
2.2.	STAKEHOLDERS .....	3
2.3.	CANDIDATE SCENARIOS.....	4
2.4.	PROPOSED SOLUTION FOR SCENARIO 1: ADMINISTRATIVE MAP .....	4
2.5.	EXPECTED BENEFITS .....	5
2.6.	SUCCESS CRITERIA.....	5
2.7.	CONSTRAINTS, LIMITATIONS AND RISKS .....	5
<b>3.</b>	<b>DESCRIPTION OF THE LOGD SERVICE .....</b>	<b>6</b>
3.1.	DATA PREPARATION .....	6
3.2.	METADATA STRUCTURE .....	7
3.3.	URI DESIGN .....	8
3.4.	THE LINKED DATA INFRASTRUCTURE .....	9
<b>4.</b>	<b>DEMONSTRATION OF THE LOGD SERVICE.....</b>	<b>10</b>
4.1.	USE CASE 1: FIND A PUBLIC AGENCY ON THE MAP .....	10
4.2.	USE CASE 2: BROWSE THE HIERARCHY OF PUBLIC AGENCIES IN GREECE .....	11
<b>5.</b>	<b>CONCLUSIONS AND NEXT STEPS .....</b>	<b>13</b>
	<b>REFERENCES.....</b>	<b>15</b>
<b>ANNEX I.</b>	<b>DEPLOYING THE LOGD PILOT ON THE VIRTUOSO UNIVERSAL SERVER.....</b>	<b>16</b>

## 1. INTRODUCTION

This report was prepared in the context of Action 1.1 of the Interoperability Solutions for European Public Administrations (ISA) Programme. It provides an overview of the activities that were carried out for supporting the Hellenic Ministry of Administrative Reform and e-Governance with the development of a Linked Open Government Data pilot demonstrating the feasibility and use of the ISA Core Vocabularies.

### 1.1. Context: publication of organograms of the Greek public administration

National legislation in Greece and other EU Member States makes it mandatory for all public administration agencies to create and open-up their organograms, thus making openly available the organizational structure of an agency as well as its hierarchical relationships to other agencies, e.g. X is supervised by Y.

To further enable interoperability on the semantic level, common vocabularies should be used as much as possible. The ISA Core Vocabularies<sup>1</sup> provide vocabularies for the person, location, business, and public service concepts. Hence, they are well suited for the purpose of publishing organograms of public administration agencies.

### 1.2. Objectives

The ensued LOGD pilot has a dual objective:

- To define consecutive application scenarios that demonstrate the added-value of opening-up high-quality government data, and support the development of at least one of them.
- To demonstrate the use and the extensibility of semantic standards developed by the ISA Programme, such as the e-Government Core Vocabularies, in a national context. Core Vocabularies facilitate the integration of data from heterogeneous sources, providing a common machine-readable representation of basic concepts such as an address or a business.

### 1.3. Structure

The remainder of this report is structured as follows. Chapter 2 presents a basic description of the main aspects of the pilot, including the business need, its scope and objectives, scenarios to be implemented, expected benefits, and anticipated threats

---

<sup>1</sup> The ISA Core Vocabularies are available via the following links:

Core Person vocabulary v1.00, May 2012.

[https://joinup.ec.europa.eu/asset/core\\_person/description](https://joinup.ec.europa.eu/asset/core_person/description)

Registered Organisation vocabulary v1.00, May 2012.

[https://joinup.ec.europa.eu/asset/core\\_business/description](https://joinup.ec.europa.eu/asset/core_business/description)

Core Location vocabulary v1.00, May 2012.

[https://joinup.ec.europa.eu/asset/core\\_location/description](https://joinup.ec.europa.eu/asset/core_location/description)

Core Public Service vocabulary v1.00, March 2013.

[https://joinup.ec.europa.eu/asset/core\\_public\\_service/description](https://joinup.ec.europa.eu/asset/core_public_service/description)

and risks. Chapter 3 provides a technical description of the pilot as it is implemented. Chapter 4 shows how the pilot can be used to solve some use cases.

#### 1.4. Glossary

The table below defines the most important terms and acronyms used in the remainder of this report.

**Table 1: Glossary**

Term / Acronym	Description
FOAF	Friend of a Friend – RDF Vocabulary for describing people
HTML	Hypertext Markup Language
ISA Programme	The Interoperability Solutions for European Public Administrations Programme of the European Union
LOGD	Linked Open Government Data
MAREG	Hellenic Ministry of Administrative Reform and e-Governance
MS	Member State
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System – RDF Vocabulary for the representation of key reference data such as code lists, and taxonomies.
SPARQL	SPARQL Query Language for RDF
URI	Uniform Resource Identifier
URI set	A collection of reference data published using URIs, about a single concept, governed from a single source.
W3C	World Wide Web Consortium

## 2. PILOT DESCRIPTION

This section describes the pilot and elaborates on the business case for using the Core Vocabularies for the publication of public data related to the organisations in the Greek public sector. It identifies the business need, stakeholders, candidate scenarios, proposed solution, expected benefits, critical success factors, threats and risks.

### 2.1. Business need

National legislation in Greece and other EU Member States makes it mandatory for all public administration agencies to create and open-up their organograms, thus making openly available the organizational structure of an agency as well as its hierarchical relationships to other agencies, e.g. X is supervised by Y. There are two reasons for wanting to model government structure:

1. **transparency** – so that citizens can see how their government is organised and how responsibility is divided;
2. government **efficiency** – organisation charts (a.k.a. organograms) help people within government find the individual or organisation they need to deal with.

In Greece, up to now, this information is published in html pages, in textual format (e.g. in pdf files) and databases. It is scattered across different systems and is described using different vocabularies. It is therefore incomplete and hard to reuse.

International good practice show the way towards publishing the organograms of public agencies and the organisation of public administration in machine-readable formats, thus enabling their reuse and allowing for flexible visualisations of the information.

### 2.2. Stakeholders

The stakeholders and their roles in the pilot are summarised in the table below.

**Table 2: Stakeholders and Roles**

Stakeholder	Description
Hellenic Ministry of Administrative Reform and e-Governance <sup>2</sup>	The Ministry of Administrative Reform and e-Governance (MAREG), part of the government of Greece, provided data on the organisation of public administration in the country coming from different sources. <b>Role in the pilot:</b> Data provider and owner.
EC, ISA Programme <sup>3</sup>	The Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission supports and facilitates efficient and effective cross-border electronic collaboration between European public administrations. The programme aims at enabling the

---

<sup>2</sup> <http://www.ydmed.gov.gr/?lang=en>

<sup>3</sup> <http://ec.europa.eu/isa/>

Stakeholder	Description
	delivery of electronic public services and the availability, interoperability, re-use and sharing of common solutions. <b>Role in the pilot:</b> The ISA Programme has initiated the pilot, reviewed the work carried out, and co-financed it as a demonstration of the use of Core Vocabularies.

### 2.3. Candidate scenarios

The stakeholders of the pilot agreed on the following three proposed successive scenarios, each of which builds upon its previous one(s).

**Scenario 1 (administrative map):** This scenario will deliver a queryable administrative map of Greece. Public administration agencies will be placed on the map and basic information about each agency, e.g. type, address, contact details and dependencies with other agencies) will be made available.

**Scenario 2 (administrative map with staff information):** Scenario 2 extends scenario 1 by adding also information about the public servants that are serving in each of the public administration agencies described previously. Both individual data per public servant and aggregate data will be supported – omitting in any case private data. In the case of aggregate data it will be possible to organise public servants in different categories, e.g. by level of education, by field of education, by type of work-contract, by age, by sex etc.

**Scenario 3 (administrative map with staff information – including aggregate salary information):** Scenario 3 extends scenario 2 by including the salary information of public administration staff. This can be done either by collecting aggregate data about the salaries of public servant at a particular function or within a particular agency; or by including the actual salaries of every public servant, as part of the information that is opened-up by them.

*It was finally agreed that only scenario 1 would be implemented in the context of this pilot.*

### 2.4. Proposed solution for scenario 1: administrative map

The Organization (ORG) Ontology<sup>4</sup>, the Registered Organization Vocabulary<sup>5</sup> (RegOrg) and the Core Location Vocabulary<sup>6</sup> were used in order to describe public administration agencies (e.g. their type, address, contact details, hierarchical relationships between agencies) and visualise this data on a map and/or otherwise.

---

<sup>4</sup> <http://www.w3.org/TR/vocab-org/>

<sup>5</sup> <http://www.w3.org/TR/vocab-regorg/>

<sup>6</sup> [https://joinup.ec.europa.eu/asset/core\\_location/asset\\_release/core-location-vocabulary-100](https://joinup.ec.europa.eu/asset/core_location/asset_release/core-location-vocabulary-100)

## 2.5. Expected benefits

This pilot aimed at demonstrating the following potential benefits both to data publishers and data consumers (public administrations / businesses and the general public):

- Standard Web interfaces (such as HTTP and SPARQL) can greatly simplify the use of public data, especially for machines.
- Linked data is an efficient technology for integrating heterogeneous data scattered in different data sources.
- Shared unique identifiers in the form of HTTP URIs can be used for integrating and cross-querying fragmented and heterogeneous data;
- Linking public data helps identify underlying data quality issues; hence, an increased use of public data may lead to improvements of quality and completeness of the source data. This will also enrich the meaning of the data by adding information about the context (of (re)use).
- Possibility to develop new data-driven services and applications, thus creating value for citizens, businesses and public administration, and return on investment from the data.

## 2.6. Success criteria

The success of the pilot has been determined by the following indicators:

- The pilot can be applied with small changes to production level.
- The stakeholders perceive a national benefit of the pilot and are willing to invest in it, taking it to production.
- Cost benefit criterion: the consolidation and linking of existing data from disparate sources and languages can be implemented within limited resources and time using existing technologies.

## 2.7. Constraints, limitations and risks

The pilot is a *proof-of-concept* demonstrating that Linked Data technologies can efficiently support the integration of heterogeneous datasets from different sources.

The pilot is limited to scenario 1, because data about staff and salary was not available at the time of writing.

### 3. DESCRIPTION OF THE LOGD SERVICE

This section describes the process of transforming raw data into a Linked Open Government Data service using the ISA Core Vocabularies. The raw data consists of several loosely linked datasets. Data curing and linking is thus the first step of the process. Next, we decide on how to structure the data with the help of Core Vocabularies. URIs are used to identify objects and concepts, therefore their design is discussed. Finally, we cover how to set up the infrastructure providing the service.

#### 3.1. Data preparation

Table 3 shows the datasets provided by MAREG. The census dataset lists the offices of public administration agencies that have completed a nation-wide census. The hierarchy of those agencies are provided by the hierarchy dataset. The transparency dataset is derived from finance databases and contains the VAT numbers of agencies. The syzefxis dataset contains a contact log for some agencies, linking them to phone numbers. The kep dataset lists the Citizen Service Centres (KEPs).

The main goal has been to combine information from all these different datasets, structure it according to the ISA Core Vocabularies and make it available in RDF as linked data.

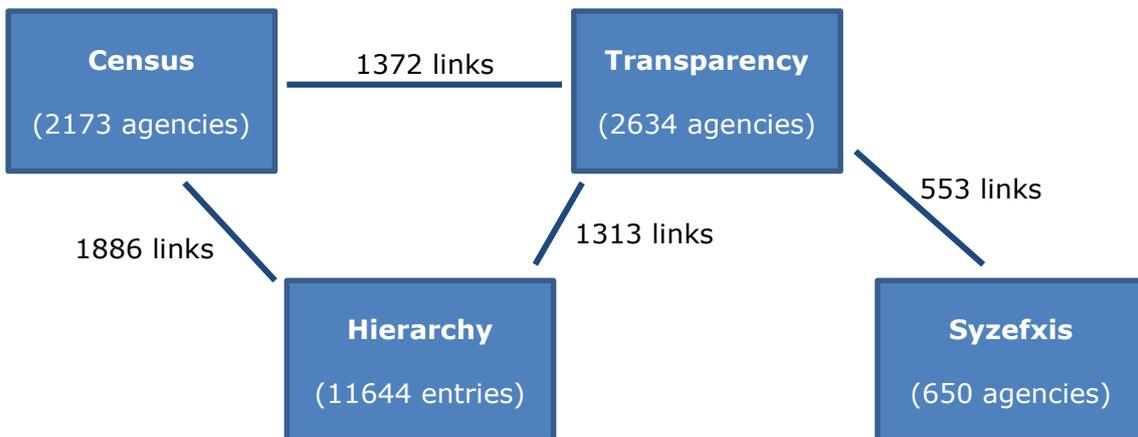
**Table 3: Datasets provided by MAREG**

Name	Description
<b>census</b>	Agency names and addresses.
<b>hierarchy</b>	Hierarchy of agencies, reusing the identifiers of the census dataset. The dataset also contains a category code.
<b>hierarchy_types</b>	Labels for the category codes in the hierarchy dataset.
<b>transparency</b>	Agency names, VAT numbers, organizations types, and supervising agency. Some entries are also linked to identifiers of the census dataset.
<b>syzefxis</b>	Agency names, VAT numbers, phone numbers, and addresses.
<b>kep</b>	Agency names, addresses, phone numbers, and e-mail addresses. The dataset is not linked with the other datasets.

The datasets were converted into RDF data using a Python script. The script is available at <https://github.com/SEMICEu/gr-pilot>. The datasets were converted to UTF-8 encoded CSV files to feed to the script.

The extractor script performs the following cleansing tasks:

- Some Greek letters were encoded as their ASCII counterparts, e.g., the Greek capital alpha (Α) is often encoded as the Latin capital A. The script converts Latin letters to their Greek counterparts.
- For some entries in the hierarchy dataset, columns were shifted. The script tries to recover the correct ordering.
- Empty names and invalid VAT numbers are discarded.
- Some entries in the transparency dataset do not mention a corresponding identifier in the census dataset, while they do appear in either the census or the hierarchy dataset. The script discovers such additional links by matching the legal names (ignoring case, punctuation, and diacritics).



**Figure 1: Overview of the links between the datasets**

Ideally, an agency would be identified by its VAT number. However, many agencies have unknown VAT numbers and do not appear in the transparency dataset. On the other hand, some agencies have more than one VAT number. Hence we focus on the census dataset and identify agencies by their identifiers in that dataset.

The census dataset contains location information about 2173 agencies. Out of those, 87 % have a corresponding entry in the hierarchy dataset, adding information about the category and the supervising agency.

63 % of the census dataset have a corresponding entry in the transparency dataset, adding information about the VAT numbers and organization type. We do not consider hierarchy information of the transparency dataset, as it seems to contain wrong relations. We do however use the VAT numbers to further link an agency to entries in the syzefxis dataset, adding information about phone numbers (33 % of the census dataset).

The kep dataset was not used as it is not linked in any way to the other datasets. Figure 1 shows an overview of the datasets and the links between them.

### 3.2. Metadata structure

The public agencies are described mainly with the W3C Organization Ontology and the ISA Core Location vocabulary. Each agency is an instance of the `org:Organization` class. Each address is an instance of the `locn:Address` class.

The table below shows the properties that were used to describe public agencies. Some properties were borrowed from the “Friend of a Friend” (FOAF)<sup>7</sup> and “Registered Organization” (RegOrg) vocabularies. Categories and types are described as SKOS concepts<sup>8</sup>.

**Table 4: Properties of an agency (i.e., an instance of `org:Organization`)**

Property	Vocabulary	Description
<code>rdfs:label</code>	RDF Schema	Legal name
<code>org:identifier</code>	Org	VAT number
<code>rov:orgType</code>	RegOrg	Organization type (a SKOS concept)
<code>org:classification</code>	Org	Category (a SKOS concept)
<code>org:subOrganizationOf</code>	Org	Supervising agency
<code>locn:address</code>	Core Location	Address
<code>foaf:phone</code>	FOAF	Phone number

### 3.3. URI design

Drawing on the recommendations of the [PURI] study, URIs that are used as identifiers in the data follow the pattern

`http://data.ydmed.gov.gr/id/{concept}/{reference}`

where `{concept}` is the concept type of the resource, and `{reference}` a unique identifier whose meaning depend on the concept type. The following table shows the concept types.

**Table 5: Concept keywords used in URIs**

Concept	Class	Description
<code>organization</code>	<code>org:Organization</code>	Public agency, using the identifier of the census dataset.
<code>address</code>	<code>locn:Address</code>	Address, using the address identifier of the census dataset.
<code>category</code>	<code>skos:Concept</code>	Agency category, using the code of the hierarchy dataset.
<code>type</code>	<code>skos:Concept</code>	Organization type, using a simple incremental number as identifier.

The above URIs identify real-world objects and concepts which cannot be downloaded from the internet. However, it is still useful to get metadata about the object/concept

---

<sup>7</sup> <http://xmlns.com/foaf/spec/>

<sup>8</sup> <http://www.w3.org/TR/skos-reference/>

when requesting a URI. Such metadata can be downloaded from URLs following the pattern (basically replacing "id" by "doc"):

```
http://data.ydmed.gov.gr/doc/{concept}/{reference}{extension?}
```

Without extension, content negotiation mechanisms will be used to determine which content to return. Web browsers will get a human-readable representation of the object/concept, while specialized applications may request raw data in RDF/XML or Turtle format. The optional extension allows requesting a particular representation format, ignoring content negotiation mechanisms.

When a user requests an "id" URI, he gets the HTTP 303 response code ("See also") with a link to the corresponding "doc" URL. Web browser handle such redirections seamlessly. The redirection approach keeps a clean separation between URIs used as identifiers and URLs used for downloading metadata or human-readable representations.

### 3.4. The linked data infrastructure

The pilot is available at <http://data.ydmed.gov.gr>. Section 4 shows the behaviour of the pilot.

The LOGD service is implemented on the Virtuoso Universal Server<sup>9</sup>, available under the GNU General Public License Version 2. All code is available on <https://github.com/SEMICeu/gr-pilot> under the EUPL.

The service provides the following features:

- A front page with a simple keyword search engine and a list of agency categories to start browsing;
- Resolving URIs to machine-readable RDF data or human-readable visualizations<sup>10</sup>;
- A SPARQL endpoint, available at <http://data.ydmed.gov.gr/sparql>.

Annex I contains detailed steps on how to deploy the pilot on the Virtuoso Universal Server.

---

<sup>9</sup> <http://virtuoso.openlinksw.com/>

<sup>10</sup> Under the hood, visualization of the data is implemented with Virtuoso Server Pages and XML Stylesheets. The `description.vsp` page handles the responses for "doc" URLs. It executes a SPARQL CONSTRUCT query and applies an XSL Transformation on the RDF/XML result.

## 4. DEMONSTRATION OF THE LOGD SERVICE

This section provides an overview of the end-user interface of the LOGD service. The service describes 2173 agencies in 21,445 locations (addresses). The underlying triple store contains 437,372 triples. While the user interface has been written in English, the data was only provided in Greek. Hence, the examples shown in this section are also given in Greek.

### 4.1. Use case 1: Find a public agency on the map

As an example, we want to find ΕΘΝΙΚΟ ΣΥΜΒΟΥΛΙΟ ΡΑΔΙΟΤΗΛΕΟΡΑΣΗΣ (National Board for Radiotelevision) on the map. The front page of the service provides a search box as depicted below.

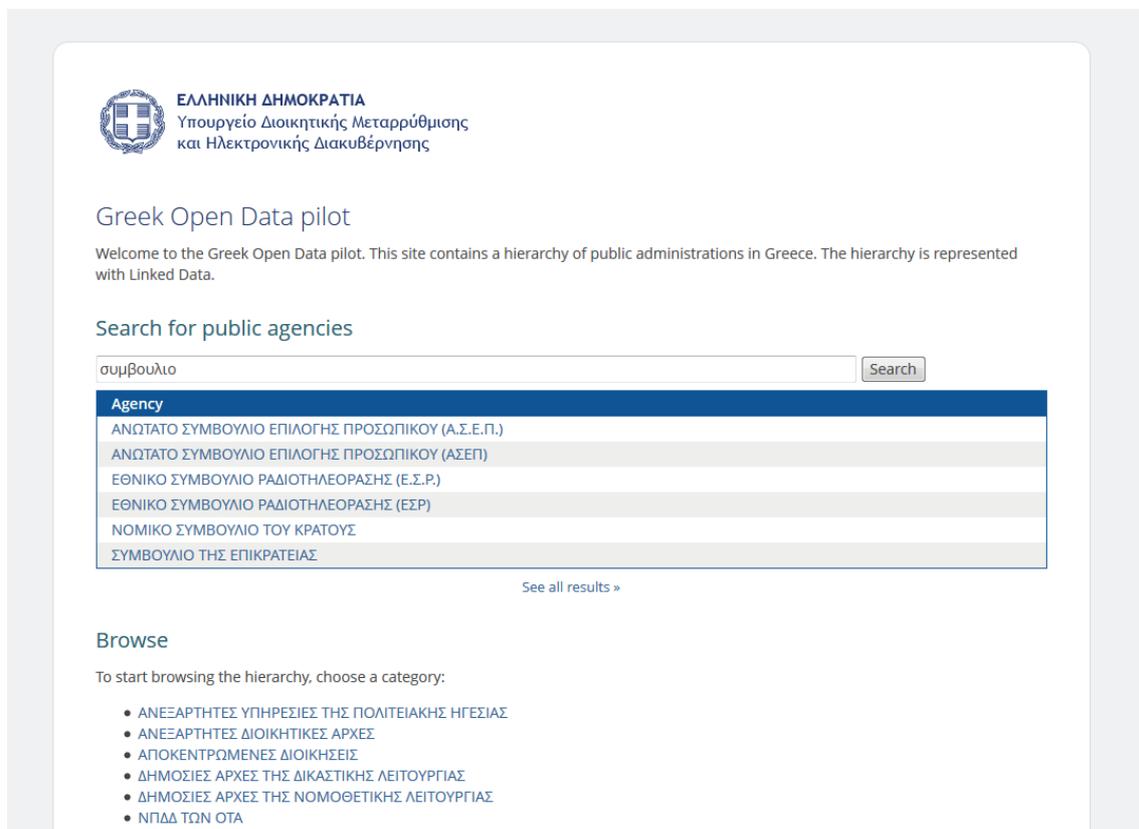


Figure 2: The front page allows to search for an agency

Clicking on “ΕΘΝΙΚΟ ΣΥΜΒΟΥΛΙΟ ΡΑΔΙΟΤΗΛΕΟΡΑΣΗΣ (Ε.Σ.Ρ.)” brings us to a description of the agency along with a map (see Figure 3). The page contains the following information:

- A unique URI to be used by other applications using the data;
- The VAT number taken from the transparency dataset;
- The category taken from the hierarchy dataset;
- Locations taken from the census dataset;
- Phone numbers taken from the syzefxis dataset.

On the page, download links are also available for machine-readable representations of the same data in RDF/XML or Turtle format. Such raw data, along with unique URIs for identifying agencies and addresses, allow third parties to integrate the data in their applications.

ΕΘΝΙΚΟ ΣΥΜΒΟΥΛΙΟ ΡΑΔΙΟΤΗΛΕΟΡΑΣΗΣ (Ε.Σ.Ρ.)

URI: <http://data.ydmed.gov.gr/id/organization/9261>

Type: Organization

Raw data: [HTML](#) | [RDF/XML](#) | [Turtle](#)

VAT number: 999369227

Category: [ΑΝΕΞΑΡΤΗΤΕΣ ΔΙΟΙΚΗΤΙΚΕΣ ΑΡΧΕΣ](#)

**Locations**

- ΑΜΕΡΙΚΗΣ 5, ΑΘΗΝΑ 10564, ΕΛΛΑΔΑ
- ΠΑΝΕΠΙΣΤΗΜΙΟΥ, ΑΘΗΝΑ 10564, ΕΛΛΑΔΑ

**Phone numbers**

- +302131502301
- +302131502302
- +302131502304
- +302131502305
- +302131502307
- +302131502378
- +302131502379
- +302131502380
- +302131502381
- +302131502382

[Show more](#)

**Map**

Figure 3: The description page of an agency shows its locations on a map

## 4.2. Use case 2: Browse the hierarchy of public agencies in Greece

The description page of an agency shows the supervising and supervised agencies in the top-right corner. Such information is extracted from the hierarchy dataset. For example, ΔΗΜΟΣ ΝΕΑΣ ΙΩΝΙΑΣ (Municipality of Nea Ionia) is supervised by ΥΠΟΥΡΓΕΙΟ ΕΣΩΤΕΡΙΚΩΝ (Ministry of Interior), and supervises two agencies:

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
Υπουργείο Διοικητικής Μεταρρύθμισης  
και Ηλεκτρονικής Διακυβέρνησης

ΔΗΜΟΣ ΝΕΑΣ ΙΩΝΙΑΣ

URI: <http://data.ydmed.gov.gr/id/organization/11105>

Type: Organization

Raw data: [HTML](#) | [RDF/XML](#) | [Turtle](#)

VAT number: 090127044

Org. type: ΟΤΑ Α ΒΑΘΜΟΥ

Category: ΟΤΑ Α ΒΑΘΜΟΥ (ΚΑΛΛΙΚΡΑΤΗΣ)

ΥΠΟΥΡΓΕΙΟ ΕΣΩΤΕΡΙΚΩΝ

ΔΗΜΟΣ ΝΕΑΣ ΙΩΝΙΑΣ

ΚΕΝΤΡΟ ΒΡΕΦΟΝΗΠΙΑΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΦΡΟΝΤΙΔΑΣ ΟΙΚΟΓΕΝΕΙΑΣ (ΚΕ.ΒΡΕ.ΦΟ) ΔΗΜΟΥ ΝΕΑΣ ΙΩΝΙΑΣ Ν. ΑΤΤΙΚΗΣ

ΟΡΓΑΝΙΣΜΟΣ ΠΟΛΙΤΙΣΜΟΥ ΑΘΛΗΤΙΣΜΟΥ ΚΑΙ ΝΕΟΛΑΙΑΣ Ν. ΙΩΝΙΑΣ (Ο.Π.Α.Ν.)

Figure 4: The description page of an agency shows its supervising and supervised agencies

Clicking on a supervising or supervised agency brings the visitor to the description page for that agency:

The screenshot shows a web page for the organization 'ΟΡΓΑΝΙΣΜΟΣ ΠΟΛΙΤΙΣΜΟΥ ΑΘΛΗΤΙΣΜΟΥ ΚΑΙ ΝΕΟΛΑΙΑΣ Ν. ΙΩΝΙΑΣ (Ο.Π.Α.Ν.)'. At the top left is the logo of the Hellenic Republic and the text 'ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ Υπουργείο Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης'. The organization's name is displayed prominently. Below it, there are details: URI: http://data.ydmed.gov.gr/id/organization/16181, Type: Organization, and Raw data: HTML | RDF/XML | Turtle. Further down, VAT number: 997736036, Org. type: Ν.Π.Δ.Δ., and Category: ΝΠΔΔ ΤΩΝ ΟΤΑ are listed. A 'Locations' section contains two entries: 'ΛΕΩΦΟΡΟΣ ΗΡΑΚΛΕΙΟΥ 264, ΝΕΑ ΙΩΝΙΑ ΑΤΤΙΚΗΣ 14231, ΕΛΛΑΔΑ' and 'ΛΕΩΦΟΡΟΣ ΗΡΑΚΛΕΙΟΥ 268, ΝΕΑ ΙΩΝΙΑ ΑΤΤΙΚΗΣ 14231, ΕΛΛΑΔΑ'. On the right, there is a 'Map' section with a street map of Nea Ionisia, Greece, showing a blue location pin and a callout box with the organization's name. The callout box also includes the text 'ΔΗΜΟΣ ΝΕΑΣ ΙΩΝΙΑΣ'.

Figure 5: Clicking on an agency in the hierarchy leads to the description page of that agency

## 5. CONCLUSIONS AND NEXT STEPS

The LOGD pilot with the Greek public administration has demonstrated the *technical* feasibility of the following aspects:

- The ISA Core Vocabularies can be used as a foundational vocabulary to homogenise organisation data that originates from disparate data sources as described in Section 3.1.
- The ISA Core Vocabularies can be directly used and easily extended with other standard vocabularies as described in Section 3.2.
- Standard Web interfaces (such as HTTP and SPARQL) greatly simplify the use of public data, especially for machines. The graphical user interface developed in the context of this pilot (Section 3.4) has made use of these standard Web interfaces (in combination with XSLT, HTML, and JavaScript). Other machines could just as easily consume the Linked Data Service set up in the context of this pilot.
- Linking public data helps identify underlying data quality issues via data triangulation techniques. For example, entities such as public administrations mentioned in one datasets but absent in another allow to identify missing records. Also, an increased use of public data may lead to improvements of quality and completeness of the source data. This will also enrich the meaning of the data by adding information about the context (of (re)use).
- Possibility to develop new data-driven services and applications, thus creating value for citizens, businesses and public administration, and return on investment from the data. One example of this would be to link the now obtained register of public administrations to information on the public services offered by them. All this information could feed into the content management system of a central government access portal. Examples of such achievements in other member states are the gov.uk or the Danish borger.dk government access portals.
- In order to ensure the sustainability of the pilot, a manual has been created and published on Github<sup>11</sup>, which allows MAREG to easily deploy the software on their own servers following a step-by-step approach. MAREG plans to migrate the pilot to their own infrastructure. A technical meeting is set up to initiate this migration. Before taking the pilot into production, MAREG needs to further clean its data.

Nonetheless, the pilot is only a *proof-of-concept*. Further study is required to analyse and deal with the following challenges before the pilot can be transformed into a production-level solution.

- Cleansing and curating the data is a tedious and expensive task, which is often hard to automate, as it requires domain knowledge and expertise, which is hard to codify. Additionally, it may also require changes to the way (structure

---

<sup>11</sup> <https://github.com/SEMICeu/gr-pilot>

and format) that the data sources expose their data.

- Matching agencies from various data sources is error-prone and difficult to fully automate, due to abbreviations and variations in agency names. It is however an essential task before attributing URIs to unambiguously identify agencies.
- Linking to external data sources is required for adding context to the data and enriching their meaning and expressivity. For example municipality URIs could be linked (wherever possible) to their counterparts in DBpedia.
- For keeping the system up-to-date, the communication between the data sources and the Linked Data infrastructure should be automated, so that updates are automatically cascaded.
- Managing the data lifecycle is a challenging task, especially in a case like this one, where data from numerous sources is integrated. Mechanisms for handling updates and deletions in the data should be devised.
- The pilot is implemented on top of well-known open-source software. However, before taking this to a production level, further testing is required, mostly to evaluate scalability, performance, and security. Some additional infrastructure might be needed, such as a reverse proxy in front of Virtuoso.

Next steps also include implementing the following scenarios:

- Scenario 2: add information about the public servants that are serving in each of the public administration agencies.
- Scenario 3: include the salary information of public administration staff.

## REFERENCES

- [ISA] Interoperability Solutions for European Public Administrations,  
<http://ec.europa.eu/isa/>
- [Location] Core Location Vocabulary  
[http://joinup.ec.europa.eu/asset/core\\_location/description](http://joinup.ec.europa.eu/asset/core_location/description)
- [PURI] D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC  
<https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>
- [Virtuoso] OpenLink Virtuoso  
<http://docs.openlinksw.com/virtuoso/>

## Annex I. DEPLOYING THE LOGD PILOT ON THE VIRTUOSO UNIVERSAL SERVER

Here are the steps to deploy the LOGD service on a Virtuoso server running on the Debian operating system.

1. Download all the files from <https://github.com/SEMICeu/gr-pilot> to `/var/lib/virtuoso/vsp/gr-pilot`.
2. Run `./configure.sh` to set up the base URI and installation directory.
3. Put the extracted CSV files in the `data` subdirectory, and generate the RDF data with `python -m extractor`.
4. Upload `data.ttl` into the Virtuoso graph `http://data.ydmed.gov.gr/`.<sup>12</sup>
5. In Virtuoso Conductor, go to "Web Application Server" > "Virtual Domains & Directories". Create a virtual domain `data.ydmed.gov.gr`, and edit the `/` path with the following parameters:
  - a. Physical path: `/gr-pilot/www/`
  - b. Default page: `index.vsp`
  - c. VSP user: `dba`
6. Set up the following rewrite rules for the `/` path. All rules are to be set as "First matching".

**Table 6: Rewrite rules for Virtuoso**

Source pattern	Destination	Response code
<code>^/search(.*)\$</code>	<code>/search.vsp\$s1</code>	Internal
<code>^/about/([^/]+)/(.*)\$</code>	<code>/description.vsp?format=\$U1&amp;uri=\$s2</code>	Internal
<code>^/doc/([^/.]*) (?:/([^/.]*)?) (?:\.(.*))?\$</code>	<code>/description.vsp?type=\$U1&amp;id=\$U2&amp;format=\$U3</code>	Internal
<code>^/id/(.*)\$</code>	<code>/doc/\$s1</code>	303

<sup>12</sup> See <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtRDFInsert> for details.

7. In Virtuoso Conductor, go to “Linked Data” > “Namespaces”, and add the following namespaces:

**Table 7: Namespaces for Virtuoso**

Prefix	URI
dcterms	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>
foaf	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>
locn	<a href="http://www.w3.org/ns/locn#">http://www.w3.org/ns/locn#</a>
org	<a href="http://www.w3.org/ns/org#">http://www.w3.org/ns/org#</a>
rov	<a href="http://www.w3.org/ns/regorg#">http://www.w3.org/ns/regorg#</a>
skos	<a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a>