SC6DI06692

D5.2.1 - REPORT ON THE PILOT IMPLEMENTATIONS BASED ON CORE VOCABULARIES

Pilot on the Organization Ontology and the Registered Organisation Vocabulary

Deliverable

JOINING UP GOVERNMENTS

DOCUMENT METADATA

Property	Value
Release date	*05/03/2013*
Status:	*For acceptance*
Version:	*0.06*
Authors:	*Phil Archer W3C* *Nikolaos Loutas, PwC EU Services* *Stijn Goedertier, PwC EU Services* *Christophe Colas, PwC EU Services*
Reviewed by:	*Joao Frade, PwC EU Services* Vassilis Michalitsis, *Ministry of Administrative Reform and eGovernance, Greece* Thodoris Papadopoulos, *Ministry of Administrative Reform and eGovernance, Greece*
Approved by:	

DOCUMENT HISTORY

Version	Description	Action
0.01	*Creation*	*Creation*
0.02	*Internal Review and delivered for Review*	*Review*
0.03	*Completion of main text*	*Update*
0.04	*Minor updates throughout*	*Update*
0.05	*Amending root URIs throughout, addition of new section on validating the model.*	*Update*
0.06	*Final version delivered for acceptance*	*Acceptance*

This report was prepared for the ISA programme by:
PwC EU Services

Disclaimer:

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This document describes the Greek Linked Open Government Data Pilot that was carried out in the context of Action 1.1 of the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission.

The pilot demonstrated that:

- the **Organization Ontology** can readily be used as a foundational RDF vocabulary to describe public administrations;
- RDF representations of governments can be **readily and repeatedly produced** using Open Refine[1] (formerly Google Refine) with simple Excel file inputs;
- the arrangement of ministries and their sub units is highly variable between Member States;
- minimal interoperability between Member States is relatively easy to achieve using the ESA95 classification of central and local administrations;
- a high level of interoperability of data between Member States would require the use of not just the Org Ontology and the ESA95 classification but of a controlled vocabulary such as **COFOG** [COFOG];
- use of COFOG or any other common data point requires the active involvement of public authorities and this may not be practicable;
- meaningful commonality between the organisation of governments in different Member States is hard to discover due to **multilingualism**.

In short: the tools exist to create the data in a relatively simple process and the Organization Ontology is well designed to cope with the significant variation between Member States. The more difficult aspect is the sourcing and creation of the detailed classification of different public authorities that is necessary for fully automated interoperability.

1.1 THE ORGANIZATION ONTOLOGY & REGISTERED ORGANIZATION VOCABULARY

The Organization Ontology, ORG, was created initially by Dave Reynolds of Epimorphics [EPI] in 2010[2]. It provides the classes, properties and relationships to describe the structure of an organisation and the reporting structures within it. Since 2010 it has been further developed by the W3C and, at the time of writing, is expected to become a W3C Recommendation during

[1] https://github.com/OpenRefine

[2] For clarity, W3C standards use American English spelling whereas this document uses British English, hence the 'Organization Ontology' (en-us) is used to describe 'organisations' (en-gb)

2013 [ORG]. Figure 1 shows the current version which is now expected to remain stable. The Registered Organisation Vocabulary, originally known as the Core Business Vocabulary [RegOrg] is a profile of ORG with terms required specifically to describe businesses and business units that gain legal entity status through company registration. Although not used directly in this study, RegOrg provides important mechanisms that could be used, for example, to publish data about private sector contractors and other service providers for the public sector.

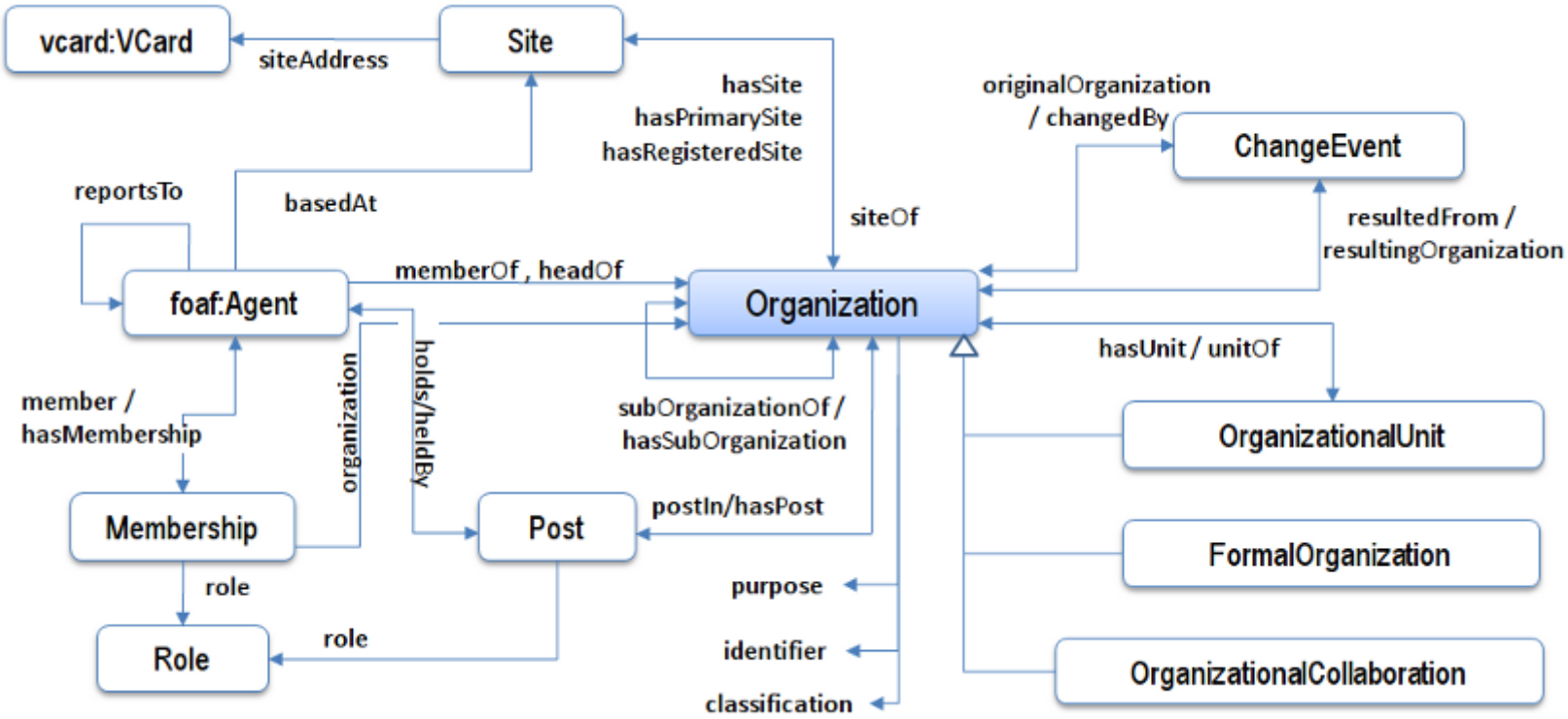

Figure 1 Diagrammatic representation of the Organization Ontology

1.2 STAKEHOLDERS

The table below lists the stakeholders involved in this pilot and their role.

Table 1 – Stakeholders and Roles [SDI-EU]

Stakeholder	Description
Ministry of Administrative Reform and e-Governance ***(Greece)***	The Ministry of Administrative Reform and e-Governance (MAREG), part of the government of Greece, provided limited data including only the organisational units in one section of the ministry. **Role in the pilot**: MAREG provided a sample, participated in coordination meetings, and reviewed the work carried out.
Agenzia per l'Italia Digitale ***(Italy)***	Similar, although more sophisticated work, has been done by Giorgia Lodi and Antonio Maccioni the Agenzia per l'Italia Digitale. This includes a description of many public authorities in Italy as well as the individuals in charge of those administrations. **Role in the pilot**: A benchmark against which the Greek data can be tested. The pilot should show that it is possible to recognise commonalities between ministries in different Member States.
ISA Programme	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 delivery of electronic public services and the availability, interoperability, re-use and sharing of common solutions. **Role in the pilot**: The ISA Programme has initiated the pilot, reviewed the work carried out, and financed it.

1.3 BUSINESS NEED

National legislation in Greece, the UK 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, the posts within that agency and the person per post (including also contact information).

In Greece, up to now, this information is published in html pages, in textual format (e.g. in pdf files) or as pictures (e.g. in jpg or png format), which hampers its reuse, see for example [3] and [4]. Good practice from the UK government show the way towards publishing the organongrams of public agencies in machine readable formats[5], thus enabling their reuse and allowing for flexible visualisations of the information.

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.

The latter point holds at all levels of government from local to international. Someone in a role in one EU Member State, for example, can quickly identify their opposite number in another Member State thus facilitating greater cooperation and sharing of solutions to common problems.

1.4 PROPOSED SOLUTION

The two participating ministries have tackled the pilot very differently. One has populated a triple store with abundant data about the entire public sector[6], the other has provided a small amount of data in an Excel spreadsheet. Such diversity is to be expected of course and it should be emphasised that *neither* approach is complete in meeting the business case. What we have is a beginning and a roadmap for what is needed if that business case is to be met. As a result of the pilot, we can foresee a relatively simple process for non specialists as follows:

1. create the data in an Excel spreadsheet;
2. load the data into an instance of Open Refine [OR] with the RDF extension created by DERI installed [RDFR];
3. copy the 'RDF skeleton' into Refine;
4. export the RDF and load into a triple store;
5. run a SPARQL Update query to generate inferred triples.

[3] http://kwz.me/4D
[4] http://www.tpd.gr/?page_id=792
[5] http://data.gov.uk/organogram/cabinet-office
[6] http://spcdata.digitpa.gov.it/dataIPA.html

The pilot shows that this process worked for the Greek data and can be repeated. The key step towards truly interoperable data is step 1.

The figure below summarises the architecture model of the pilot.

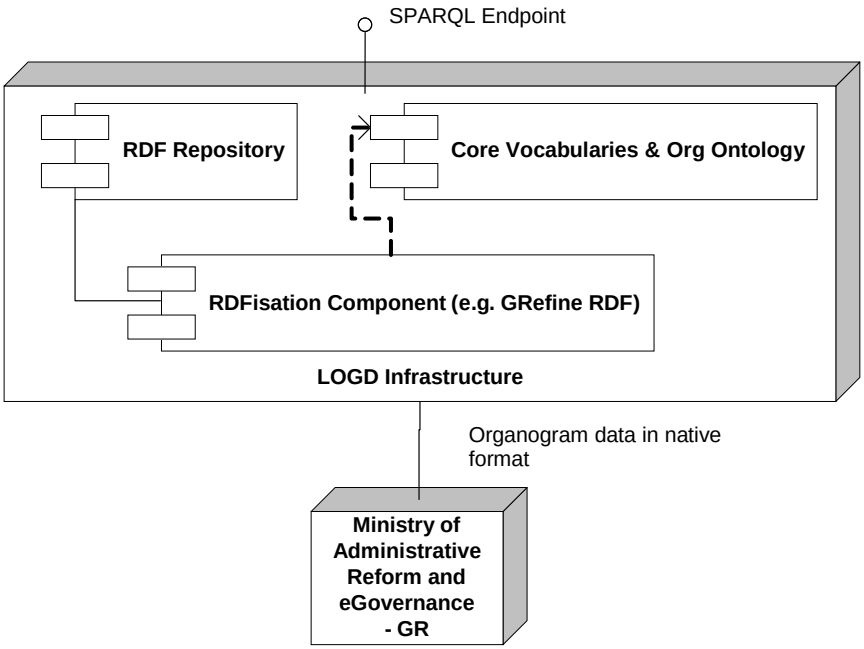

Figure 2: Architecture model – GR MAREG

Having modelled the Greek ministry, it was possible to compare the model with the Italian one, which also uses the Core Vocabularies and Org Ontology, and assess the differences and similarities.

1.5 USE CASES

The use cases for organograms are centred on the need to **identify the relevant/responsible governmental unit for a particular topic**.

For example, a government worker in Athens may wish to contact their opposite number in Rome; a citizen may want to find out the department responsible for a particular function; someone at the EC might want to identify who they need to contact to ask about implementation of a Directive and so on.

In some countries, notably the UK, organograms include the names of individuals in specific posts with senior staff's salaries also disclosed. This not only aids government efficiency but also greatly increases transparency.

1.6 TARGET BENEFITS

This pilot aims at *demonstrating* – not sustaining – the potential benefits both to data publishers and data consumers (public administrations and businesses). The specific benefits envisaged are:

- improved transparency towards the citizen;
- the use of common vocabularies and URIs as common identifiers make fragmented and heterogeneous organisational data **interoperable** and **linkable,** within the same country but also across borders;
- how the use of standard Web interfaces (such as HTTP(S) and SPARQL) can greatly **simplify the use of data** representing the internal structure of public administration and public posts;
- an increased use of this data leading to **improved quality** of address data, e.g. by crowd-sourcing improvements to data quality;
- opening the possibility of developing **new data-driven services and applications**, thus creating value and ROI from the data.

2. URI SETS FOR GREEK PUBLIC ORGANIZATIONS, UNITS AND PUBLIC POSTS

On any linked data implementation, it is important to define a URI structure and stick to it. A study of the ISA programme on good practices for persistent URIs[7] as well as UK *Government's Designing URI Sets for the UK Public Sector* [UKURI] converge on the following pattern:

`http://{domain}/{type}/{concept}/{reference}`

where {type} is one of

- `id` – where the URI identifies a non information resource (a real world object);
- `doc` – where the URI identifies a document, including one that describes a non information resource;
- `def` – a concept definition;
- `set` – a data set.

{concept} is a word or string to capture the essence of the real-world 'thing' that the URI identifies e.g. school.

{reference} a string that is used to identify an individual instance of concept.

Taking this advice we can construct a URI for the Greek Ministry of Administrative Reform and eGovernance thus:

http://org.testproject.eu/mareg/id/ministry/ministry-of-administrative-reform-and-e-governance

The domain, http://org.testproject.eu/ and initial path segment /mareg/ were used for the pilot but would be replaced by something like http://reference.data.gov.gr if implemented in Greece. The 'reference' sub domain would indicate that this is reference data, i.e., a way to refer to something that is likely to be used in many different datasets - like a ministry - rather than something much more intangible like financial results, specific policies and other data produced by or that is about the ministry.

The URI identifies a real world thing and so the type is 'id.'

The concept here is of a 'ministry' and the specific ministry is 'ministry-of-administrative-reform-and-e-governance' so the URI becomes more specific as we move from left to right.

[7] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

Other concepts that are apparent in the Greek data are:

- directorate general;
- directorate;
- office;
- department;
- unit; and
- managing authority.

Given these types of government unit it is also possible to mint URIs for them following the same pattern so that the URI for the concept of 'a ministry' is:

http://org.testproject.eu/mareg/def/orgunit/ministry
http://org.testproject.eu/mareg/def/orgunit/directorategeneral
etc.

These URI identify the *definition* of a ministry, a directorate general etc. and so the type is `def`, not `id`.

These patterns are followed in the Italian data except for the {type} element. For example, the Agenzia delle Dogane has the identifier:

http://spcdata.digitpa.gov.it/Amministrazione/agdogane

The URI structure tells humans that Agenzia delle Dogane is an Administrazione. Machines need to follow the links to discover that:

http://spcdata.digitpa.gov.it/Amministrazione/agdogane is an instance of the class identified by http://spcdata.digitpa.gov.it/Amministrazione which is the class of Italian public administrations. Italian public administrations are defined as a sub class of ORG's Formal Organization class.

3. MAPPING THE SAMPLE GREEK DATASET TO THE ORGANIZATION ONTOLOGY

Through Thodoris Papadopoulos and Vassilis Michalitsis of the Informatics Development Agency (part of the ministry), we were able to obtain a spreadsheet that partially described the ministry of Administrative Reform and e-Governance.

1	Ministry of Administrative Reform and e-Governance				
2	Minister	Deputy Minister	Secretary General	Directorate General of Financial and Administrative Services	Directorate of Budget and Financial Reports
3					Directorate of Procurement,
4					Directorate of Administrative Services
5				Directorate General of Administrative Reform and e-	Directorate of Strategic Planning for Administrative Reform
6					Directorate of e-Governance
7					Directorate of Public Services Provision
8				Directorate General of Human Resources Management	Directorate of Planning and Human Resources Development
9					Directorate of Human Resources
10			Managing Authority of the Operational Programme "Administrative Reform		
11		Department for Parliamentary Control			
12		Department for Legislative Work			
13		Department of Internal Auditing			

Figure 3 The original data received

The concept behind the data shown in Figure 3 is that each unit is a sub unit of the one immediately to the left in the spreadsheet, so that, for example, the Directorate of Budget and Financial Reports is a sub unit of the Directorate General of Financial and Administrative Services. This makes sense and is the kind of data that a member of staff within a public administration could reasonably be expected to supply. A matching spreadsheet was provided with the same data in Greek.

The data posed a number of limitations that had to be curated. First we observed that the data mixes organisational units and the individuals who are responsible for them. The Minister, Deputy Minister and Secretary General are individuals whereas the Directorate General of Financial and Administrative Services and the Managing Authority of the Operational Programme "Administrative Reform 2007-2013" are organisational units. With the agreement of Thodoris Papadopoulos and Vassilis Michalitsis the data was altered to talk about *the offices of* the Deputy Minister and Secretary General.

Then the hierarchy is not sufficiently explicit in the spreadsheet data for machine processing. For example, just looking at the spreadsheet it is not clear that the Directorate of Human Resources is a sub unit of Directorate General of Human Resources Management which is a sub unit of the Secretary General's Office which is a sub unit of the Deputy Minister's Office since the cells to the left of Human Resources are all blank. For processing, those cells need to be filled in, however repetitive that may be.

The is a common occurrence and so Open Refine, the tool we use to generate RDF from the spreadsheet data, has an option to 'fill down' – i.e. repeat the current cell's contents in all those below until a non-empty cell is encountered. The problem with using the fill down function is that it very quickly introduces unintended consequences. For example, the Directorate General of Human Resources Management is rendered as a sub unit of the Managing Authority of the Operational Programme “Administrative Reform 2007-2013” (which it is not). The fill down tool is very useful but needs to be used with care.

The actual conversion of the data is carried out when Open Refine executes the rules encoded in the 'RDF Skeleton' (See Figure 4).

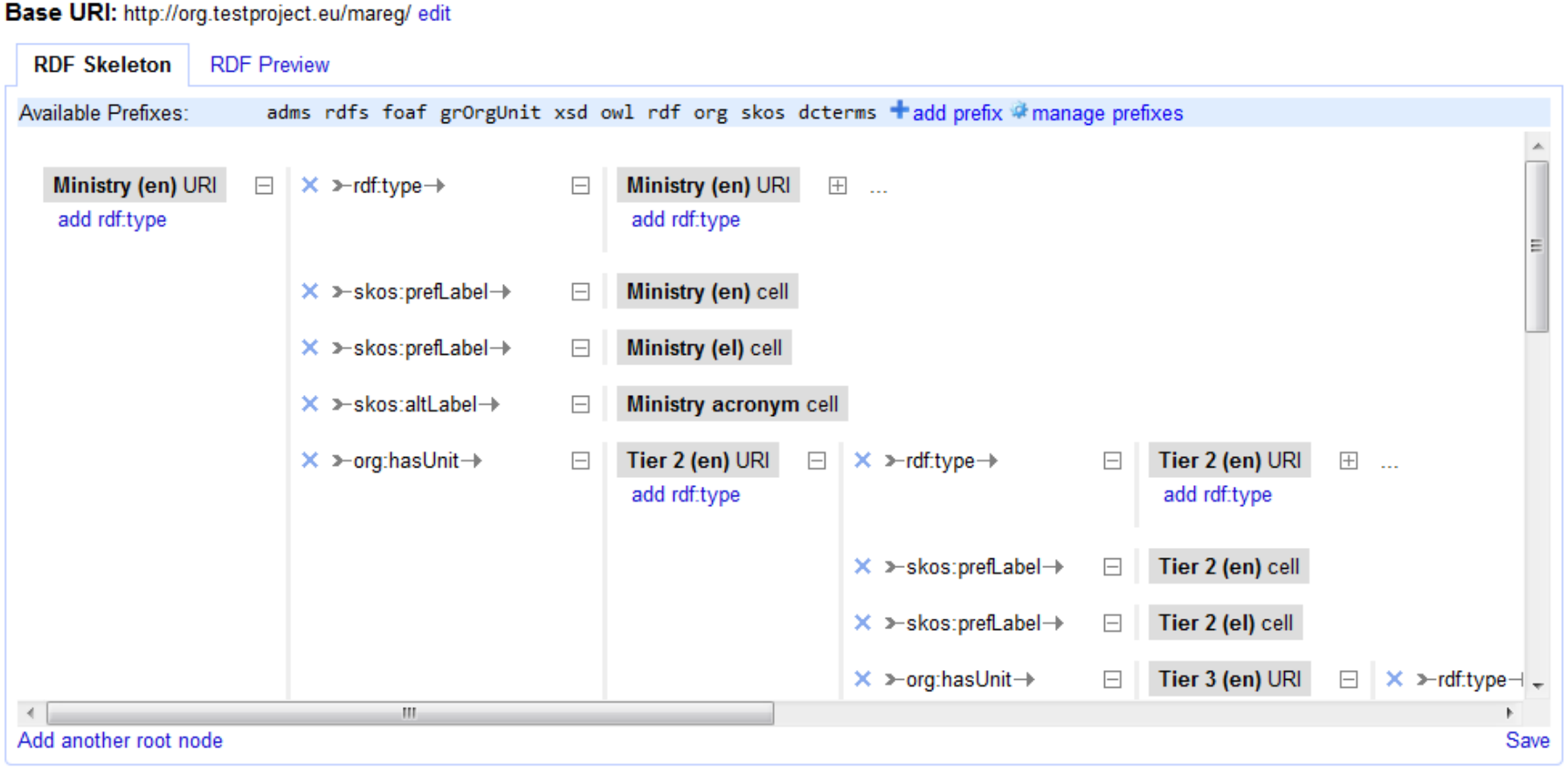

Figure 4 Part of the RDF Skeleton for Open Refine that generates the output data from the Excel spreadsheet

More detail of the RDF Skeleton is provided in section 3.1 but for now it's important to recognise that this is a machine process - one that can be repeated indefinitely.

After an iterative process of experimentation with the RDF Skeleton in Open Refine and making changes to the input data, the final input data has:

- column headings that reflect the structure of the ministry;
- both languages in a single spreadsheet.

Figure 5 shows some of the reorganised input data.

Tier 4 (en)	Tier 4 (el)	Tier 5 (en)	Tier 5 (el)
Directorate General of Financial and Administrative Services	Γενική Διεύθυνση Οικονομικών και Διοικητικών Υπηρεσιών	Directorate of Budget and Financial Reports	Διεύθυνση Προϋπολογισμού και Δημοσιονομικών Αναφορών
Directorate General of Financial and Administrative Services	Γενική Διεύθυνση Οικονομικών και Διοικητικών Υπηρεσιών	Directorate of Procurement, Infrastructure and Assets Management	Διεύθυνση Προμηθειών, Υποδομών και Διαχείρισης Υλικού
Directorate General of Financial and Administrative Services	Γενική Διεύθυνση Οικονομικών και Διοικητικών Υπηρεσιών	Directorate of Administrative Services	Διεύθυνση Διοικητικών Υπηρεσιών
Directorate General of Administrative Reform and e-Governance	Γενική Διεύθυνση Διοικητικής Μεταρρύθμισης & Ηλεκτρονικής Διακυβέρνησης	Directorate of Strategic Planning for Administrative Reform	Διεύθυνση Στρατηγικού Σχεδιασμού Διοικητικής Μεταρρύθμισης
Directorate General of Administrative Reform and e-Governance	Γενική Διεύθυνση Διοικητικής Μεταρρύθμισης & Ηλεκτρονικής Διακυβέρνησης	Directorate of e-Governance	Διεύθυνση Ηλεκτρονικής Διακυβέρνησης

Figure 5 Screenshot of some of the rearranged input data

There are several points to make about this revised spreadsheet.

1. As noted, for any given cell that carries data, all cells to its left are filled in, even though that means many cells on the left of the spreadsheet contain exactly the same data.
2. The Greek names of the different units within the ministry are in columns immediately to the right of the English equivalents.
3. All the names of the organisational units within the ministry begin with the key word or keywords that define the type of unit. For example, 'Directorate General of Financial and Administrative Services' begins with the keywords 'Directorate General' and not, 'the' or any other term. It is these keywords that the machine processing of the data picks up when creating the data. The Open Refine template used recognises the terms

- Ministry;
- Office;
- Directorate; and
- Managing.

and if the next words are 'general' or 'authority' then they are also recognised.

4. With the key words for each organisational unit included in the data cells themselves, what is important from a structural point of view is the tier it sits in within the structure of the ministry. Column headings therefore are Tier1, Tier 2 etc.

5. The English words are used to generate the URIs for two reasons. Firstly because the individual working on the project doesn't speak Greek and so simply finds it a lot easier to work with English. More importantly however is that, although non-ASCII characters are certainly allowed in URIs in terms of the technical specification, not all linked data tools work reliably with Internationalised Resource Identifiers [RFC3987]. In principle though there is no reason at all for URIs to contain only the letters a-z.

3.1 THE RDF SKELETON

A brief look at the Greek data reveals that there are specific types of organisational unit within ministries: departments, directorates general and so on. The first thing to do when planning the RDF data was to create a brief ontology of these.

The Greek Organisational Unit ontology is very short and simple and can easily be extended. For the purposes of the pilot we used the prefix `grOrgUnit` and assigned it a namespace of `http://org.testproject.eu/mareg/def/orgunit/`. It defines 6 classes:

```
grOrgUnit:Ministry
grOrgUnit:Office
grOrgUnit:Department
grOrgUnit:ManagingAuthority
grOrgUnit:DirectorateGeneral
grOrgUnit:Directorate
```

`grOrgUnit:Ministry` is defined as a sub class of the ORG ontology's Formal Organization but there is no relationship drawn between the classes within the ontology.

That done we are able to say that the Ministry of Administrative Reform and e-Governance is an instance of the class `grOrgUnit:Ministry`, the Office of the Deputy Minister for Administrative Reform and e-governance is an instance of the class `grOrgUnit:Office` and

so on. With that done, and with the spreadsheet set up as discussed in the previous section, we can now set up the rules within Open Refine to generate the RDF.

Just two rules do most of the work as described below.

3.1.1 Generating the identifier for the unit

The following expression is applied to all cells containing English language names of the various organisational units and generates an identifier for them.

```
"'/id/'+ toLowercase(replace(replace(value,
/(ministry|office|department|managing|directorate\\s(general|authority
)?).*/i, \"$1\"), /\\s+/, '')) + '/' + value.urlify()"
```

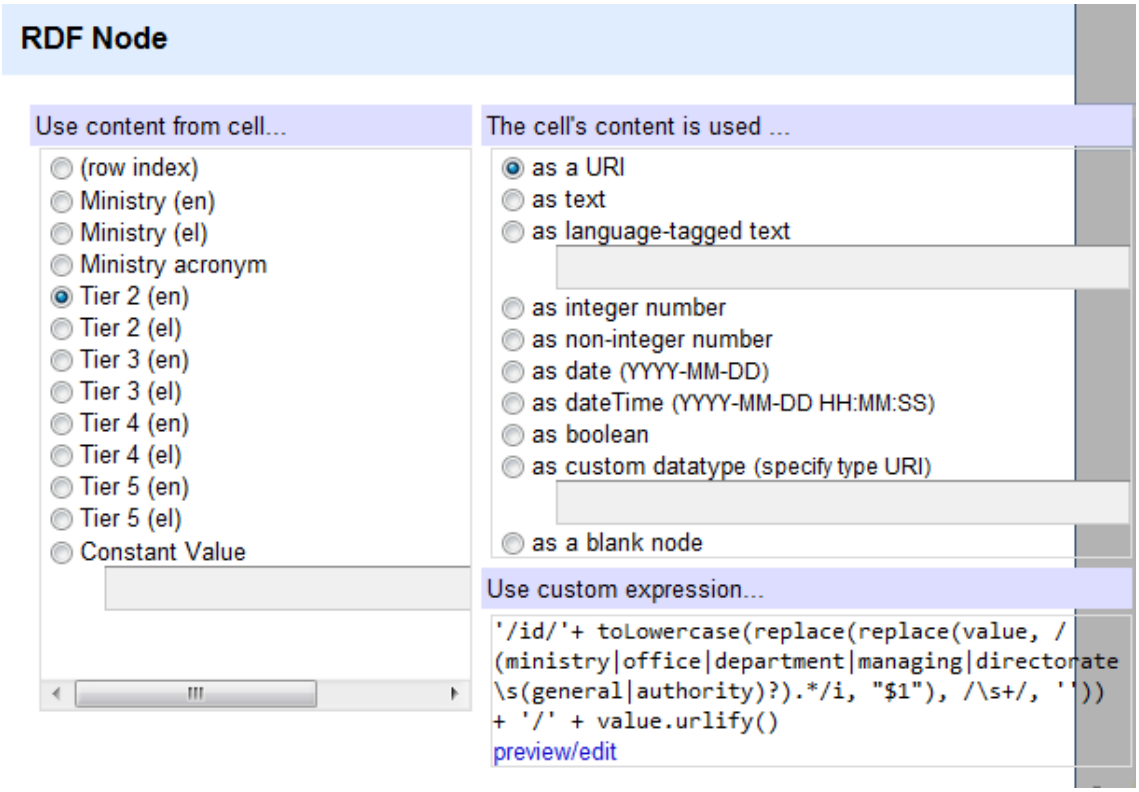

Figure 6 An example of a rule encoded in the RDF Skeleton

It works by taking the base URI of the data set, http://org.testproject.eu/mareg/, and appending it first with '/id/' and lastly with a 'url-ified' version of the string, that is, with spaces replaced by hyphens and any other excluded characters escaped. In between, a regular expression extracts the key words ministry, office, department, managing and directorate, optionally followed by general or authority. Given text that begins with one of the recognised organisational unit types, irrespective of case, the output is therefore one of:

```
ministry
```

```
office
department
managingauthority
directorategeneral
directorate
```

Given the string "Directorate General of Human Resources Management" the output is

```
http://org.testproject.eu/mareg/id/directorategeneral/directorate-general-of-human-resources-management
```

3.1.2 Generating the type

A similar rule is applied to the same data cells to give the type according to the grOrgUnit mini ontology.

```
http://org.testproject.eu/mareg/def/orgunit/' +

replace(toTitlecase(replace(value,

/((ministry|office|department|managing|directorate)\\s(general|authority)?).*/i, \"$1\")), /\\s+/, '')"
```

Given a string like "Office of the Deputy Minister for Administrative Reform and e-governance", the RDF extension to Open Refine generates the following URI as the value of the `rdf:type` property:

```
http://org.testproject.eu/mareg/def/orgunit/Office
```

3.1.3 Other rules

Much simpler rules are defined to create the English and Greek language labels and to generate an `org:hasUnit` relationship between one organisational unit and the next one in the chain.

3.1.4 Portability

The full skeleton is published separately as a text file that can be copied and pasted into the RDF extension to Open Refine. Although the same rule is applied multiple times, there is no way to express that within the skeleton, i.e., you can't say "apply this rule to columns A, C and E" and so they are repeated throughout. Even so, by creating a minimal set of rules, the maintainability of the skeleton is greatly increased. If, for example, a new organisational unit were recognised in the grOrgUnit ontology, then the *same* adjustment could be made to *all*

instances of the rules highlighted above without having to work out specifically where the changes needed to be made.

Portability is however limited since the processing rules will only work for Greek ministries and the assumption that names of the organisational units are available in English and Greek is hard wired into the RDF skeleton.

3.2 THE OUTPUT DATA

RDF, particularly machine-generated RDF, is designed to be read by machines not people and so the following extract from the data has been adjusted in terms of presentation but not in terms of content.

```
<http://org.testproject.eu/mareg/id/ministry/ministry-of-
administrative-reform-and-e-governance>
  a grOrgUnit:Ministry ;
  skos:prefLabel "Ministry of Administrative Reform and
  e-Governance"@en ,
  "Υπουργείο Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής
  Διακυβέρνησης"@el ;
  skos:altLabel "MAREG" ;
  org:hasUnit <http://org.testproject.eu/mareg/id/office/office-of-
  the-deputy-minister-for-administrative-reform-and-e-governance> ,
  <http://org.testproject.eu/mareg/id/department/department-for-
  parliamentary-control> ,
  <http://org.testproject.eu/mareg/id/department/department-for-
  legislative-work> '
  <http://org.testproject.eu/mareg/id/department/department-of-
  internal-auditing> .
```

This states that the entity identified by

```
<http://org.testproject.eu/mareg/id/ministry/ministry-of-
administrative-reform-and-e-governance>
```

- is a Ministry of the Government of Greece;
- is called "Ministry of Administrative Reform and e-Governance" in English and "Υπουργείο Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης" in Greek;
- is alternatively known as MAREG;
- has 3 units identified by their URIs.

Similar data is generated for the units within the ministry.

The full dataset, unadjusted for presentation, is published separately.

The pilot, included all related documentation, e.g. raw and transformed data, the OrgUnitOntology and the RDF skeleton, can be accessed at http://org.tesproject.eu/mareg/

3.3 USE CASE: IDENTIFY THE RELEVANT/RESPONSIBLE GOVERNMENTAL UNIT FOR A PARTICULAR TOPIC.

The linked data infrastructure deployed for the purpose of this pilot illustrates how the organisational data of MAREG is made accessible in both human- and machine-readable formats using Linked Data technologies.

After visiting http://org.tesproject.eu/mareg/, the user lands on the following page, which illustrates the organisational data of MAREG in human-readable format.

Figure 7 Screenshot – MAREG linked organisational data in human readable format

By clicking on any of the links, the user can get additional information for this resource (through dereferencing). For example, clicking on 'Ministry of Administrative Reform and eGovernance', additional information about the ministry will be retrieved, as shown in Figure 8.

Figure 8 Screenshot – dereferencing 'Ministry of Administrative Reform and eGovernance'

The SPARQL endpoint can be accessed at the following location: http://org.testproject.eu/sparql

4. OPTIONS FOR LINKING TO THE ITALIAN DATA

The dataset created by Agenzia per l'Italia Digitale is comprehensive and covers more than 20,000 Italian public sector bodies. It makes use of the Org Ontology in much the same way as was used in the pilot for the Greek ministry, namely, by creating a mini ontology of the types of public body the exist in Italy. Each public body is then declared to be an instance of the relevant class.

As we noted in section 2, for example, the URI http://spcdata.digitpa.gov.it/Amministrazione identifies the class of Italian public administrations and this is defined as a sub class of `org:FormalOrganization` in exactly the same way as `grOrgUnit:Ministry`. Immediately though we see a difference: Agenzia delle Dogane is a ministry in the Italian government. Υπουργείο Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης is a ministry in the Greek government. At this point the only point of commonality is that both Greek and Italian ministries are defined as sub classes of org:FormalOrganization which tells us simply that they are *an Organization which is recognized in the world at large, in particular in legal jurisdictions, with associated rights and responsibilities. Examples include a corporation, charity, government or church* which is something but not a great deal. It certainly *doesn't* provide a useful 'interoperability hook.' For a start, what is a ministry?

Agenzia per l'Italia Digitale turned to the ESA95 manual on government deficit and debt [ESA95] for an answer to that question. This document provides a definition of central and local administrations that are useful for our discussion although the focus of the classification is on accountancy rules. It does not provide a detailed classification of all public bodies that we could apply to the Greek and Italian organograms. Therefore, Agenzia per l'Italia Digitale created their own SKOS concept scheme for their public authorities and simply used the ESA95 definition of a top level central administration as their top concept [EAP].

http://spcdata.digitpa.gov.it/CategoriaAmministrazione/30 identifies the concept of a ministry (specifically Presidenza del Consiglio dei Ministri e Ministeri or Presidency of the Council of Ministers and ministries). This equates roughly to the concept of a ministry in Greece and elsewhere and so we can classify the Greek Ministry of Administrative Reform and e-Governance in the same way. But we can't go any further than this and use the remainder of the Italian classification since it includes definitions and concepts that are unique to Italy.

The RDF skeleton produced in the pilot generates a triple stating that any ministry can be classified (org:classification) as http://spcdata.digitpa.gov.it/CategoriaAmministrazione/30 and the `grOrgUnit` mini ontology itself also classifies ministries in the same way.

It is certainly possible to provide deeper connections between different public authorities in different Member States. The most promising mechanism for doing this would be to link each organisational unit to the relevant classification in the UN's Classification of the Functions of Government [COFOG]. This is a detailed list of the functions of government and, if each public administration were to classify itself according to COFOG, hopping between different departments in different Member States that did the same task would be easy.

The problem is scale. The data itself must realistically come from the public administrations themselves. For an individual, or even a team, to work through a dataset the size of the Italian one and match up departments with their COFOG code would be an enormous task. It is for this reason that interoperability between the two datasets considered in this pilot is limited.

5. VALIDATING THE APPROACH

Having completed the initial stages of the pilot, the Greek ministry was contacted again and two requests were made:

- would it be possible to add COFOG classifications?
- was there any more data that could be added to the revised spreadsheet?

The answer to the first question was yes – but with difficulty. It is clear that the only people who would be able to provide detailed classification data for a particular department according to the COFOG terms would be someone with first hand knowledge of the department itself. For an 'outsider' – even someone within the same ministry – providing such data is difficult and, through no fault of any individual, error prone. This re-enforces the finding that provision of COFOG classification data is difficult.

The second question had a more positive answer. The ministry was able to provide an updated and extended version of the spreadsheet, using the columns as revised in Figure 5. Loading this and the RDF skeleton into Open Refine produced a satisfying data set as output – the rules worked. However, the revised data includes a new type of organisational unit. The Managing Authority of the Operational Programme “Administrative Reform 2007-2013” has 4 Units within it (known as Units A – D). The original rules for generating the identifiers and types don't recognise the concept of 'a unit' and therefore generated things like:

`http://org.tesptproject.eu/mareg/id/unit-a-planning-and-evaluation`

This is acceptable but not ideal since the `{concept}` element is missing. What we're looking for is

`http://org.tesptproject.eu/mareg/id/unit/unit-a-planning-and-evaluation`

i.e. with `/unit/` after `/id/`. Achieving this required two simple steps:

1. A new class was added to the Organisational Units mini ontology, namely `grOrgUnit:Unit`.
2. The rules that generate the identifiers and types were extended by performing a simple search and replace on the RDF skeleton, replacing all instances of

   `ministry|office|department|managing|directorate`

   with

   `ministry|office|department|managing|directorate|unit`

   i.e. the regular expression was extended to include units.

Having made those two changes, the final data was generated. This 'second round' of data from the Greek ministry validated the approach as the amount of work required to include the new data was trivial and the results consistent.

6. CONCLUSIONS

The data from the Greek ministry was delivered as an Excel spreadsheet and requires knowledge of the ministry, not of any technical issues. The RDF Skeleton can be copied and pasted into the DERI extension for Open Refine and it will then generate the RDF - so the pilot shows that this is a process that is **manageable** and **scalable**. This meets the transparency objective in that the data can be published in a standard format from which visualisations could be created to show the organisational structure of ministries. Such visualisations are likely to depend on country-specific details but would not be any less useful for that.

Using the Organization Ontology and a common approach for modelling the public sector in different Member States offers **modest interoperability**. Deeper interoperability between Member States can only be achieved through the curation of data that classifies public administrations according to their function using a common schema, almost certainly COFOG. Such data is not readily available in Italy or Greece and it would require input from each public administration to create. Were it available, however, the pilot shows how it could readily be processed.

The pilot, including all related documentation, can be accessed at

- http://org.tesproject.eu/mareg/ and
- https://joinup.ec.europa.eu/asset/core_business/document/organization-ontology-pilot-linking-public-sectors-organisational-data

REFERENCES

CBV — Core Business Vocabulary, ISA programme Core Vocabularies Working group. https://joinup.ec.europa.eu/asset/core_business/description

COFOG — (Classification of the Functions of Government, United nations Statistics Division, http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=4. SKOS version available as download only (i.e. with non-dereferencable URIs at http://unstats.un.org/unsd/cr/registry/regdnld.asp

[EAP] — Elenco Amministrazioni Pubbliche per tipologia, Istituto Nazionalse de Statistica, http://en.istat.it/strumenti/definizioni/elencoPA/elencoS13.pdf

EPI — An organization ontology, Dave Reynolds, Epimorphics. http://www.epimorphics.com/public/vocabulary/org.html

ESA95 — ESA95 manual on government deficit and debt, Publications office of the European Communities, 2002. http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/naga_a_esms_an1.pdf

OR — Open Refine, formerly Google Refine, a power tool for working with messy data. See https://github.com/OpenRefine

ORG — An organization ontology, W3C Recommendation Track document. http://www.w3.org/TR/vocab-org/

RDFR — RDF Refine - a Google Refine extension for exporting RDF, Faadi Maali and Richard Cyganiak, DERI. http://refine.deri.ie/

RegOrg — Registered Organization Vocabulary, Phil Archer, Agisilaos Papantoniou http://www.w3.org/TR/vocab-regorg/

RFC 3987 — Internationalized Resource Identifiers (IRIs), M Duerst, M. Suignard. IETF 2005 http://www.ietf.org/rfc/rfc3987.txt

UKURI — Designing URI Sets for the UK Public Sector. UK Chief Technology Officer Council October 2009. http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

ACKNOWLEDGEMENTS

Specific acknowledgement is due to:

Name	Organisation
Thodoris Papadopoulos	MAREG
Vassilis Michalitsis	MAREG
Nancy Routzouni	MAREG
Giorgia Lodi	Agenzia per l'Italia Digitale
Antonio Maccioni	Agenzia per l'Italia Digitale
Francesco Tortorelli	Agenzia per l'Italia Digitale
Vassilios Peristeras	EC ISA Programme