Geolinked Open Data for the Municipality of Catania[†]

Sergio Consoli
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy
sergio.consoli@yahoo.it

Aldo Gangemi
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy
aldo.gangemi@lipn.univ-paris13

Andrea Giovanni Nuzzolese
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy
nuzzoles@cs.unibo.it

Silvio Peroni
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy
speroni@cs.unibo.it

Valentina Presutti
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy

Diego Reforgiato Recupero
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy

Daria Spampinato
Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies, National Research Council (CNR), Italy

ABSTRACT
Linked Open Data (LOD) is reaching significant adoption in Public Administrations (PAs), where it is often required to be connected to existing platforms, such as GIS-based data management. Bearing on previous experience with the pioneering data.cnr.it, through Semantic Scout, as well as Italian DigitPA agency recommendations for LOD in Italian PA, we are working on the extraction, publication, and exploitation of data from the Geographic Information System of the Municipality of Catania, referred to as SIT ("Sistema Informativo Territoriale"). This paper describes the results and lessons learnt from the first campaign, aiming at analysing, reengineering, linking, and formalizing the Shape-based geo-data from the SIT.

Categories and Subject Descriptors
J.1. [**Computer Applications**]: Administrative Data Processing – *Government*, I.2.4. [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods – *Semantic networks*, I.7.5. [**Document and Text Processing**]: Document Capture – *Document Analysis*.

General Terms
Languages, Management, Standardization, Experimentation.

† Authors are listed in alphabetical order as their contributions are equally distributed.

Keywords
Open Data applications, Linked Open Data extraction and publication, eGovernment, GIS-based data management.

1. INTRODUCTION
LOD has gained significant momentum over the past years as a best practice of promoting the sharing and publication of structured data on the semantic Web [6, 7]. LOD offers the possibility of using data across domains or organisations for purposes like statistics, analysis, maps and publications. By linking this knowledge, interrelations and correlations can be understood quickly, and new conclusions can be arising. The low-threshold addressability of keeping knowledge within the web, especially, helps scale back existing barriers. Through URI and RDF, slices of information and data can be arranged, shared, exported, and connected, and APIs, applications, and tools can be created. LOD is currently bootstrapping the Web of Data by converting into RDF and publishing existing datasets available to the general public under open licenses. Popular private and public stocks of the semantic web include, among others, DBpedia[1], a large reference dataset of structured information extracted from Wikipedia[2] providing encyclopaedic knowledge about a multitude of different domains.

These major changes in technology and society are involving also the way of doing politics, administration and the relationship between politicians, public servants and citizens. Transparency, participation and collaboration are the main issues of the integration of citizens in the paradigm of Open Government [13].

---

[1] Available at: http://dbpedia.org

[2] Available at: http://wikipedia.org

Because administrations have large amounts of data that could be made accessible for the purpose of the LOD movement, research on the opening process, data reengineering, linking, formalisation and consumption is of primary interest [4].

This paper describes the methodology used for the extraction, enrichment, publication and reuse of LOD for PA, often referred to as Linked eGovernment Data [14], through the application of semantic technologies and software components [20], supporting the external evaluation of the PA by the detection of the community trust. The Italian PA subject investigator of our work is the Municipality of Catania (MoC). The main motivation of the work consists of experimenting social eGovernment systems aimed at optimizing the performance of PA of the MoC for the provision of intelligent ICT services to citizens and businesses. The work falls within the spirit of the Smart Cities initiatives of the European Commission, which aims at bringing together cities, industry and citizens to improve urban life through more sustainable integrated solutions. One of the main development objectives consists in conceiving, designing and prototyping applications for the MoC related to certain areas of experimentation, such as online social services and health, traffic management and transport, and urban decor. Special emphasis will be given to electronic identification and trust services, following the new EU regulation for digital identification in PA on line services. Digital identification is indeed a crucial issue in the development of secure PA information management systems, to guarantee on line access only if identity is certified by either electronic identity or national service cards, or by means of a dedicated public digital identity management system (SPID).

With the aim of detecting and collecting the required data and processes for these applications, meetings with the Leadership of the Directorate of Information Systems Service of the MoC were carried out. A field of experimentation is special focused on the management of mobility, i.e. road traffic and public transport. Within this context, the scenario has identified the development of a prototype mobile application implementing a real-time system to inform on the state of roads in urban areas to support sustainable mobility. For the realization of the mobile prototype application, it is necessary to process the data and diagrams in the Geographic Information System of the MoC, referred to as SIT: "Sistema Informativo Territoriale" [16]. Therefore it was decided, by mutual agreement with the chief officers and experts of the city hall of Catania, to process the data in order to make them open, interoperable and compatible with the principles of LOD.

The paper is organized as follows: Section 2 provides a background on the state of the art on the use of Linked eGovernment Data. Section 3 introduces the techniques and tools to gather and to deal with LOD for the PA. Section 3.3 describes the resulting ontology of our use case, along with the methods adopted to publish and to query the data. Section 4 ends the paper with conclusions and future directions where we are headed.

2. BACKGROUND

2.1 Linked eGovernment Data

The Digital Administration Code incorporates many of the international experiences on the publication of LOD for PAs. The publication, reuse and interoperability of data of PA in a simple and open as possible has become a priority in recent years in countries where the government and politics are attentive to features such as transparency, traceability and accessibility of data [4]. There is a wide range of best practices in the usage of Linked eGovernment Data. They can be mainly summarized in the following types [14]:

- portals for the supply of the Linked eGovernment Data sets;
- portals providing raw data sets of LOD for PAs along with technical tools or developer kits for understanding, interpreting, or processing the provided data;
- already existing portals, acting as showrooms for best practices for Linked eGovernment Data;
- mobile apps for smartphones using LOD for PAs.

The main thrust on the publication of LOD for PA is coming from big initiatives in the United States (data.gov) [8, 9] and the United Kingdom (data.gov.uk) [18], both providing thousands of raw sets of LOD within their portals, but there are also some other experiences and notable initiatives that are in line with the international state of the art. In Germany, one of the first examples for a LOD portal is the one from the state of Baden-Wurttemberg (opendata.service-bw.de), divided into three main parts: LOD, applications, and tools. In addition to their potentials, Linked eGovernment Data can provide great benefits in the matter of accountability, as shown in the LOD portal example of Kenya (opendata.go.ke).

The subject of LOD may be very different, depending on the needs of the distinct actors in various countries, and relying also on the aims of the providers and the culture of the final users. A forecast for the usage of Linked eGovernment Data cannot be estimated with a determined accuracy yet. However the research community in this field is unanimous in believing that, by following the examples of these initiatives, it is also possible to encourage, in the medium-long term, policies and strategies to publish LOD also for other PAs worldwide.

One of such initiatives for the Italian PA is "data.cnr.it" [5, 12], the open data project of the National Research Council (CNR), designed and maintained by the Semantic Technology Laboratory of ISTC-CNR, and shared with the unit Information Systems Office of CNR. In addition, other LOD have been published in Italy by the city hall of Florence[3], Agency for Digital Italy[4], from the Piedmont region[5], the Chamber of Deputies[6]. In this paper we present the methods and the requirements used for the construction of the ontologies, and for the extraction and publication of LOD, for the MoC.

2.2 Spatial Data in the Semantic Web

Geospatial data or geographic information is the data that identifies a geographic location of natural or constructed features and boundaries on the Earth (e.g. oceans, buildings, countries, rivers, etc). Geographical knowledge bases are among the largest in existence and have high importance in a variety of everyday applications. The data can be mapped and often manipulated with Geographic Information Systems (GIS), however the integration of external data sets into these systems is time-consuming and complex. Many real-life information integration and aggregation tasks are impossible without comprehensive background knowledge related to spatial features of the ways, structures and landscapes surrounding us. Currently, there are three major sources of open geospatial data in the Web: Spatial Data Infrastructures (SDI), open data catalogues, and crowd sourced initiatives. Crowd sourced geospatial data are emerging as a potentially valuable source of geospatial knowledge. Among

---

[3] Available at: http://opendata.comune.fi.it/ linked_data.html

[4] Available at: http://www.digitpa.gov.it

[5] Available at: http://www.dati.piemonte.it/rdf.html

[6] Available at: http://dati.camera.it

various efforts we highlight OpenStreetMap[7], GeoNames[8], and Wikipedia as the most significant. Recently, GeoSPARQL [17] has emerged as a promising standard from W3C for geospatial RDF, with the aim of standardizing geospatial RDF data modelling and querying. Integrating Semantic Web with geospatial data management requires the scientific community to address two challenges: (i) the definition of proper standards and vocabularies that describe geospatial information according to RDF and SPARQL protocols, that also conform to the principles of established geospatial standards, (e.g. OGC), (ii) the development of technologies for efficient storage, robust indexing, and native processing of semantically organized geospatial data.

One of the central projects in this area is LinkedGeoData[9] [19], in which data from OpenStreetMap is lifted into the Semantic Web infrastructure and converted to RDF. Furthermore, interlinking to other knowledge bases, such as DBpedia, GeoNames, and other datasets, is performed, as well as icons and multilingual class labels from various sources are integrated. LinkedGeoData is also striving for the establishment of an OWL vocabulary with the purpose of simplifying exchange and reuse of geographic data [19].

Geoknow[10] [1] is another project that addresses the key issues of integrating geographically related information on the Web, scalable reasoning over billions of geographic features within the Linked Data Web, as well as efficient crowd-sourcing and collaborative authoring of geographic information. In particular, GeoKnow applies the RDF model and the GeoSPARQL standard as the basis for representing and querying geospatial data, and contributes to efficient geospatial RDF querying; fusion and aggregation of geospatial RDF data; visualization and authoring tools; to provide public or private geospatial data stocks [1].

3. METHODOLOGY FOR LOD PRODUCTION

This section describes the methodologies and tools for the extraction, modelling and publishing LOD for the PA. The methods are based on the standards of the W3C[11], on good international practices, on the guidelines issued by the Agency for Digital Italy [2] and those by the Italian Index of Public Administration[12], as well as on the in-depth experience of the research participants on this field, in particular related to the development of data.cnr.it [5], the LOD portal for the Italian CNR.

The W3C is the reference point for the Semantic Web in general, and especially for LOD. In this respect, several working groups were established by the W3C. The results of the groups on RDF, on the Semantic Web Deployment, and on the Government Linked Data are considered a reference in the context of PA for opening interoperable public data. The W3C and the European Commission have in fact built the infrastructure and the culture of the Semantic Web since 1999, through several working groups, standards, and several funded projects. Besides the W3C, other working groups on RDF and the Semantic Web deserve to be

mentioned, such as the working groups on OWL and RDF, which created the conditions for make automated reasoning schemas and data on the Web; the working group on SPARQL, that has allowed to realize a very flexible and rich query language for RDF; and the group on SKOS, which has defined a simple ontology for exporting structures metadata (classifications, thesaurus, etc.) into RDF; the group on RDF that defined formats to encapsulate semantic information directly into HTML, thus enabling de facto semantic annotations of interoperable text with RDF. Based on these recommendations, some specific initiatives were then implemented in order to define and implement a class of "repositories of semantic assets" on the JoinUp[13] web portal by the EU.

The semantic standards, best practices and design guidelines for semantic interoperability through the LOD issued by the Agency for Digital Italy [2, 3] provide an excellent reference frame for the development of LOD for PA. We used the good practice suggestions into our case study with the PA of the MoC, with respect to the data and schema of their Geographic Information System, SIT.

In the following we describe the various stages for semantic interoperability methodology and tools used for the extraction of data from the SIT, and for their modelling and publishing through LOD. The whole methodology can be summarised in the following main steps:

1. Analysis of the baseline scenario and requirements (which consists in identifying and selecting the dataset, and in the re-organisation of the business data model);
2. Geo-referenced data mining tools and modelling technologies (which includes analysis and modelling tasks);
3. Resulting ontology obtained for the toponymy of SIT (i.e. enrichment of the ontology and interlinking);
4. Conversion from the geo-data to the final ontology (validation task);
5. SPARQL endpoint and content negotiation of data (i.e. publication and consumption of data).

Note that the information contained in the SIT are in Italian language, therefore the produced LOD will be in Italian too (although the all generation process that will be described is completely language-independent).

3.1 Analysis of the baseline scenario and requirements

During the phase of selection of the source data, a thorough analysis of the reference domain was made. Thanks to the constant, close interaction with the PA experts of the MoC, the Geographic Information System, SIT [16], was identified as the source dataset for the enrichment and publication of data.

The SIT is a data warehouse used for reporting and data analysis, and consisting of databases, hardware, software, and technicians, which manages, develops and integrates information of the province of Catania based on a geographical space [16]. The various territorial levels (hydrography, topography, buildings, infrastructure, technological networks, administrative boundaries and land) form the geo-localised common part of the information flow of the MoC, according to which all the constituent parts are related to each other.

The SIT is designed to contain all the available data of the PA in Catania for the purpose of in-depth knowledge of the local area. Basically the SIT contains three types of data: the register base,

---

[7] Available at: http://www.openstreetmap.org

[8] Available at: http://www.geonames.org

[9] Available at: http://linkedgeodata.org

[10] Available at: http://geoknow.eu

[11] Available at: http://www.w3.org/standards/semanticweb/

[12] Available at: http://spcdata.digitpa.gov.it/data.html

[13] Available at: http://joinup.ec.europa.eu

the registry office, and the toponymy. Through the consultation platform on the web it is possible to display the following information:
- basic cartography;
- ortho-photos;
- the road graph;
- buildings with a breakdown by main body of some areas of the city;
- cadastral sections;
- data from the 1991 and 2001 census of the population;
- the last Master Plan but without all the variants;
- gas network works in current execution;
- the resident population in selected areas (municipalities, entire street, polygonal, circular area);
- the total population, distributed into bow street, house number, etc.;
- breakdown of the population by municipality, blocks, nationality, gender, family components, age, marital status, etc.;
- extraction and search of resident persons, and their location on the bow streets (only for authorised operators);
- management of the competence areas of pharmacies;
- location and alphanumeric information of: municipality, hospitals, universities, schools, pharmacies, post offices, areas or emergency, public safety, fire departments, public green areas, public community centres, institutions for minors and orphanages.

The SIT also includes maps containing geo-referenced information related to:
- sub-services (electricity-gas-water pipes);
- data on stoppage areas;
- occupation stalls;
- stalls for disadvantaged people;
- occupation of public land;
- public transport fleet;
- management and working state of the fleet (called AVLAVM system);
- data on lines and stops of the public transport system;
- accident traffic data;
- road signs and markings;
- maintenance state of roads and sidewalks;
- management of roadway construction;
- data of the municipal police;
- the accounting of the Municipality.

To show the methodology of enrichment and publication of data of the PA of the MoC, we use only a sample of data extracted from the SIT. This portion of data is related to 32 tables of the toponymy contained in the SIT (see Table 1), provided in the form of Shape-based files [15] for each data record, i.e. files with extensions: .dbf, .shp, .shx, .sbn, .sbx, .xml.

Table 1: Processed sample data from the SIT's toponymy

Original Italian name	English translation
Archi Stradali	Road Arches
Contorno Densità	Density Contour
Pubblica Sicurezza	Public Safety
Sedi Servizi Sociali	Locations of Social Services
Aree di Emergenza	Areas of Emergency
Farmacie	Pharmacies
Rete Fibra Ottica	Fiber Optic Network
Sezioni Censimento 1991	Sections 1991 Census
Sezioni Censimento 2001	Sections 2001 Census
Carceri	Prisons
Isolati	Blocks
Rete Gas	Network of Gas Pipes
Case Riposo	Nursing Homes
Municipalità	Municipality
Scuole Aree	Schools Areas
Uffici Comunali	Municipal Offices
Centraline Smog	Pollution Control Units
Numeri Civici	Civic Numbers
Scuole	Schools
Università	Universities
Chiese	Churches
Ospedali	Hospitals
Semafori	Traffic Lights
Utenti WAN	WAN Users
Circoscrizioni	Jurisdictions
Poste	Post Offices
Serbatoi Idrici	Water Tanks
Aree Verde	Green Areas
Confini Comunali	Municipal Boundaries
Piano Regolatore Generale	General Plan
Aree Servizi Sociali	Social Services Areas
Vigili del Fuoco	Firefighters
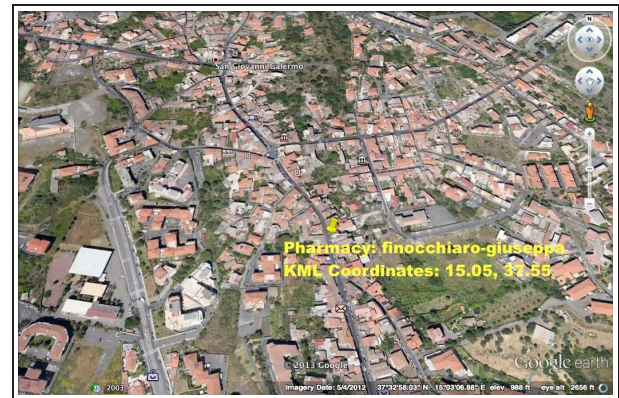
Figure 1: Example of a geo-localised entity of "pharmacies".

3.2 Geo-referenced data mining tools and modeling technologies

At this stage it is possible to use the following tools to facilitate the re-engineering of the dataset according to the new conceptual data model:
- TopBraid Composer[14] − it contains various tools for importing data from different RDF formats, including OWL;
- D2R (Database to RDF)[15] − it is a framework for accessing relational databases and to produce "RDF dumps" according to certain specifications;
- Tabels[16] − it is a software tool developed by the research foundation CTIC, which, using the GeoTools libraries[17], is

---

[14] Available at: http://www.topbraidcomposer.com

[15] Available at: http://d2rq.org/d2r-server

[16] Available at: http://idi.fundacionctic.org/tabels/

capable of transforming the information encoded in the shape files into RDF representations.

As we were not allowed by MoC authorities to install any driver directly into the SIT data warehouse, the first two options needed to be discarded. We therefore processed geo-data records coming from SIT using Tabels. In particular, from the shape files supplied for each data record (in particular, the files with

```
<http://www.essepuntato.it/2013/10/prisma/resource/farmacia/finocchiaro-giuseppa>
a <http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_Farmacie> ;
    <http://www.essepuntato.it/2013/10/prisma/CODICE-of-CATANIA.SDO_Farmacie>
      "10625" ;
    <http://www.essepuntato.it/2013/10/prisma/MUNI-of-CATANIA.SDO_Farmacie>
      "5" ;
    <http://www.essepuntato.it/2013/10/prisma/Municipali-of-CATANIA.SDO_Farmacie>
      5 ;
    <http://www.essepuntato.it/2013/10/prisma/NOME-of-CATANIA.SDO_Farmacie>
      "FINOCCHIARO GIUSEPPA" ;
    <http://www.essepuntato.it/2013/10/prisma/NUMERO-of-CATANIA.SDO_Farmacie>
      "60" ;
    <http://www.essepuntato.it/2013/10/prisma/OBJECTID-of-CATANIA.SDO_Farmacie>
      1 ;
    <http://www.essepuntato.it/2013/10/prisma/OBJECTID_1-of-CATANIA.SDO_Farmacie>
      1 ;
    <http://www.essepuntato.it/2013/10/prisma/PROPRIETA-of-CATANIA.SDO_Farmacie>
      "FINOCCHIARO GIUSEPPA" ;
    <http://www.essepuntato.it/2013/10/prisma/RECAPITO-of-CATANIA.SDO_Farmacie>
      "VIA SAN GIOVANNI BATTISTA 74" ;
    <http://www.essepuntato.it/2013/10/prisma/Shape-of-CATANIA.SDO_Farmacie>
      "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
    <http://www.essepuntato.it/2013/10/prisma/kml-of-CATANIA.SDO_Farmacie>
      <http://www.essepuntato.it/2013/10/prisma/farmacia/kml/Farmacie.1.kml> .
```

Figure 2: RDF triples produced by Tabels for the example of entity in "pharmacies".

extensions .dbf and .shp), Tabels was used for encoding the shape files into RDF triples related to the designed ontology, that it will be described in more detail in Section 3.3. On the one hand the characteristics of the table are stored as RDF representation, and, on the other hand, the spatial geometry is modelled on the standard KML representation [10]. At this stage we are mapping

```
<?xml version="1.0" encoding="UTF-8"?>
<kml:kml xmlns:kml="http://earth.google.com/kml/2.1">
    <kml:Document id="featureCollection">
        <kml:Placemark id="Farmacie.1">
            <kml:Point>
                <kml:coordinates>15.0520808419018,37.5490041443454</kml:coordinates>
            </kml:Point>
        </kml:Placemark>
    </kml:Document>
</kml:kml>
```

Figure 3: KML coordinates produced by Tabels for the example of entity in "pharmacies".

to existing vocabularies, in particular NeoGeo[18], suitable for GeoData. The geometric coordinates in KML are expressed according to the geodetic reference system GaussBoaga (or Rome 40). By means of different conversion tools available on-line (e.g. http://www.ultrasoft3d.net/Conversione_Coordinate.aspx), it is possible to produce the coordinates of latitude, longitude and altitude in meters using the geodetic system WGS84 [11]. In particular, the application of Tabels to each pair of files, .dbf and .shp, of the data tables is able to produce a set of RDF triples stored in a repository with other geometric resources contained in a public server. For example, from the information stored in the database of the SIT representing an entity of "pharmacies" (Figure 1), Tabels produces the related RDF triples, shown in Figure 2, and the file with the geometric KML coordinates (Figure 3).

Tabels works is able to import common file formats, such as XLS or CSV, including shape files. Afterwards it generates automatically a transformation program from the input data files. The generated program is able to transform each row of the input data into a new instance of a RDF class ad-hoc. In addition, each value in the column of the input tables is converted into a new triple where the subject is the instance mentioned, the predicate is a property based on the name of the column header, and the object is the value of the column as a rdfs:Literal. It is worth noting that the transformation program automatically generated, is a SPARQL-based script completely customisable by the user. Thus it is possible to change classes, names and associated properties, and then to annotate them appropriately. Once the transformation program is defined, the execution of Tabels generates the corresponding RDF in output, which we make publicly available online through a dedicated SPARQL endpoint. In addition, information regarding each resource object of the ontology data can be obtained through negotiation mechanisms of the content (content negotiation) based on HTTP REST that make them accessible, for example, through a browser or as REST web service. Data consumption is described in more detail in Section 3.5.

3.3 Resulting ontology obtained for the toponymy of SIT

The ontology resulting from the applied strategy provides classes and properties that represent the database entities of the SIT. Starting from the definition of the tables of the toponymy, a first version of OWL ontology has been developed. This is publicly available at the following URI:

http://ontologydesignpatterns.org/ont/prisma/ontology.owl

having the namespace (i.e. the default address of the entities in the ontology):

http://www.ontologydesignpatterns.org/ont/prisma/.

The creation process of this ontology was divided into two main phases and has followed the good practice of formal representation, naming, and semantic assumptions in use in the domain of the Semantic Web and LOD [2, 3]. In the first phase, the entire structure of the tables has been converted into a draft OWL ontology, where each table (i.e. each entity type described by the supplied data) is represented by a class and each field of the table has been converted into a data property. This translation was carried out in a fully automatic way from the sources provided in XML format (extension .shp.xml) by means of the use of an XSLT transformation. Note that fields with the same name but belonging to different tables have been provided with distinct properties.

For example, the fields "Name" of the tables "Nursing Homes" ("Case Riposo") and "Pharmacies" ("Farmacie") have been translated with two different data properties, respectively "Name-of-CATANIA.SDO_NursingHomes" and "Name-of-CATANIA.SDO_Pharmacies".

From this interim draft ontology and from the available data, a first version of the ontology in OWL was produced. At this stage we have followed the suggestions of the W3C Organization Ontology[19], a set of guidelines for generating, publishing and consuming LOD for organizational structures. In this respect we have named the graph nodes as URIs and pursued the following principles:

---

[17] Available at: http://geotools.org

[18] Available at: http://geovocab.org/doc/neogeo.html

[19] Available at: http://www.w3.org/TR/2014/ REC-vocab-org-20140116/

```
PREFIX project: <http://www.essepuntato.it/2013/10/prisma/semaforo/>
PREFIX my: <http://www.essepuntato.it/2013/10/prisma/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/#>

FOR ?rowId IN rows FILTER get-row(?rowId)
    MATCH [?oBJECTID,?iD,?nOME,?sDESDESE,?geometry,?kml] IN horizontal
    LET ?resource = resource(replace(replace(replace(lower-case(?nOME)," +","-"),"--+","-"),",|\.",""),
                            <http://www.essepuntato.it/2013/10/prisma/resource/semaforo/>)

CONSTRUCT {
    my:TabelsDataCatalog a dcat:Catalog .
    my:TabelsDataCatalog dct:title "Tabels AutoGenerated Catalog" .
    my:TabelsDataCatalog dct:description "Tabels AutoGenerated Catalog" .
    my:TabelsDataCatalog dct:publisher my:TabelsAutoGenerator .
    my:TabelsDataCatalog dcat:dataset my:DataSet
}

CONSTRUCT {
    ?resource a my:CATANIA.SDO_semafori .
    ?resource my:OBJECTID-of-CATANIA.SDO_semafori ?oBJECTID .
    ?resource my:Id-of-CATANIA.SDO_semafori ?iD .
    ?resource my:NOME-of-CATANIA.SDO_semafori ?nOME .
    ?resource my:sde_SDE_se-of-CATANIA.SDO_semafori ?sDESDESE .
    ?resource my:Shape-of-CATANIA.SDO_semafori ?geometry .
    ?resource my:kml-of-CATANIA.SDO_semafori ?kml
}

CONSTRUCT {
    my:CATANIA.SDO_semafori a rdfs:Class
}

CONSTRUCT {
    my:OBJECTID-of-CATANIA.SDO_semafori a rdf:Property .
    my:Id-of-CATANIA.SDO_semafori a rdf:Property .
    my:NOME-of-CATANIA.SDO_semafori a rdf:Property .
    my:sde_SDE_se-of-CATANIA.SDO_semafori a rdf:Property .
    my:Shape-of-CATANIA.SDO_semafori a rdf:Property .
    my:kml-of-CATANIA.SDO_semafori a rdf:Property
}
```

Figure 4: A view on the transformation program used by Tabels to convert the shape files to RDF for the table "Traffic Lights" ("Semafori").

- The name of all the classes was taken to the singular (e.g., from "Pharmacies" to "Pharmacy");
- The names of the data properties were aligned when they were clearly showing the same semantics. For example, the properties
- "Name-of-CATANIA.SDO_NursingHomes" and "Name-of-CATANIA.SDO_Pharmacies" ended in the same property "name", assigned to "NursingHome" and "Pharmacy" as domain or entity class;
- The data properties that seemed to refer to individuals of other classes, probably having foreign key functions on the database, were transformed into object properties. For example, the property
- "MUNI-of-CATANIA.SDO_NursingHomes" became "municipality" in order to connect individuals of class "Nursing Home' with individuals of class "Municipality";
- The data properties having values clearly assigned to some resources were transformed into object properties and their values were reified as individuals of specially created classes.

All changes made to the intermediate draft ontology for the implementation of the first version of the ontology have been documented in the form of SPARQL CONSTRUCT. This allowed us to create a simple script to convert the data extracted by Tabels in order to make them fully compliant with the final expected ontology, produced as output in RDF format.

3.4 Example of conversion from the geo-data to the final ontology

In this section we want to focus on the phase of transformation from shape files to the final RDF ontology by reporting an example. Consider as reference the data record "Traffic Lights" ("Semafori"). The SQL schema of this table includes the fields:

- **ObjectID:** unique number incremented sequentially;
- **Shape type:** Geometry that represents the coordinates defining the geometric characteristics of the entity;
- **Id:** Identification number of type Double;
- **Name:** String type name of the entity;
- **Sde SDE se:** Integer number;
- **Se ANNO CAD DATA:** Blob representing the date.

Passing the .shp and .dbf files to Tabels, this generates the transformation program, that is the SPARQL-based script used to import the data (see Figure 4). As already mentioned, it is possible to edit the script to suit custom requirements.

Once any change in the transformation program is completed, it is possible to save and run it, which generates the RDF triples from the table data given as input. Figure 5(a) shows the RDF/Turtle produced by Tabels by using the methodology already described for a single 'Traffic Light" entity as example. Figure 5(b) shows the corresponding final ontology of this entity obtained by conversion through SPARQL CONSTRUCT of the related data extracted by Tabels, in order to fully match the designed ontology.

This example further shows the ability and simplicity of the proposed methodology to gather the complex structure of a non-structured database, allowing a rapid analysis, retrieval, and conversion of the data into a structured RDF format, and the publication in the form of LOD. As already mentioned, these LOD are publicly accessible through a browser or as REST web service by means of content negotiation based on HTTP REST, as described in the next section.

3.5 SPARQL endpoint and content negotiation of data

The produced data consist of 854,221 triples and can be publicly queried by selecting the RDF graph called *<prisma>* on the dedicated SPARQL endpoint accessible at

```
@prefix my: <http://www.essepuntato.it/2013/10/prisma/> .
@prefix : <http://www.essepuntato.it/2013/10/prisma/resource/semaforo/> .

:cso-italia-cso-provincie-vle-ionio a my:CATANIA.SDO_semafori ;
    my:Id-of-CATANIA.SDO_semafori 8 ;
    my:NOME-of-CATANIA.SDO_semafori "C.so Italia-C.so Provincie-V.le Ionio" ;
    my:OBJECTID-of-CATANIA.SDO_semafori 10 ;
    my:Shape-of-CATANIA.SDO_semafori "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
    my:kml-of-CATANIA.SDO_semafori
        <http://www.essepuntato.it/2013/10/prisma/semaforo/kml/Semafori.10.kml> ;
    my:sde_SDE_se-of-CATANIA.SDO_semafori 1 .
```

```
@prefix ont: <http://ontologydesignpatterns.org/ont/prisma/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .

<http://ontologydesignpatterns.org/ont/prisma/semaforo/cso-italia-cso-provincie-vle-ionio>
    a ont:Semaforo ;
    ont:forma geo:Point ;
    ont:identificativoOggetto 10 ;
    ont:nome "C.so Italia-C.so Provincie-V.le Ionio" ;
    ont:sde 1 .
```

Figure 5: Top panel (a): RDF/Turtle for a single entity of the table "Traffic Lights" ("Semafori").

Bottom panel (b): Corresponding final RDF/Turtle ontology obtained through SPARQL CONSTRUCT conversion to fully match the designed ontology.

http://wit.istc.cnr.it:8894/sparql.

Queries can be made by editing the text area available into the interface for the SPARQL query language. SPARQL is the standard language reference and a W3C recommendation for querying RDF data. The SPARQL endpoint is also accessible as a REST web service, whose synopsis is the following:

- URL: http://wit.istc.cnr.it:8894/sparql
- Method: GET
- Parameters: query (mandatory)
- MIME type supported output: *text/html; text/rdf+n3; application/xml; application/json; application/rdf+xml.*

Data are also accessible through content negotiation. The reference namespace for the ontology (*ontology.owl*) is:

http://www.ontologydesignpatterns.org/ont/prisma/

which is identified by the prefix *prisma-ont*. The namespace associated with the data is, instead:

http://www.ontologydesignpatterns.org/data/prisma/

which is identified by the prefix *prisma*. These two namespaces allow content negotiation related to the ontology and the associated data. For example, to access data of the "Barriera-Canalicchio" suburb ("municipalità"), you have to browse: http://www.ontologydesignpatterns.org/data/prisma/municipalita/ barriera-canalicchio; or else to access geo-data of the road arch ("arco stradale") identified by the "1001", you have to browse: http://www.ontologydesignpatterns.org/data/prisma/arcostradale/1 001. The negotiation can be done either via a web browser (in this case the MIME type of the output is always text/html), or by making HTTP REST requests to one of the two namespaces. The synopsis of the REST requests to the web service associated with the namespace identified by the prefix *prisma-ont* is the following:

• URL: http://www.ontologydesignpatterns.org/ont/prisma/
• Method: GET
• Parameters: ID of the ontology object (mandatory the PATH parameter)
• MIME type supported output: *text/html;text/rdf+n3; text/turtle; text/owl-functional; text/owl-manchester; application/owl+xml; application/rdf+xml; application /rdf+json.*

Instead, the synopsis of the REST requests to the web service associated with the namespace identified by the prefix prisma is the following:

• URL: http://www.ontologydesignpatterns.org/ data/prisma/
• Method: GET
• Parameters: ID of the ontology object (mandatory the PATH parameter)
• MIME type supported output: *text/html;text/rdf+n3; text/turtle; text/owl-functional; text/owl-manchester; application/owl+xml; application/rdf+xml; application /rdf+json.*

4. CONCLUSIONS AND REMARKS

This paper presents an application of LOD for PA. The used methodology was implemented by following the standards of the W3C, the good international practices, the guidelines issued by the Agency for Digital Italy and the Italian Index of Public Administration, as well as by the in-depth experience of the research participants in the field. The method was applied to the case study of the PA of the MoC, in particular from their data stored in the Geographic Information System, SIT, available in Italian language. By using the tools and technologies for the extraction and publication of data of the used approach it was possible to produce an ontology of the SIT according to the paradigm of LOD. The data are publicly accessible to users through queries to a dedicated SPARQL endpoint, or alternatively through calls to dedicate REST web services. Mayor of Catania has acknowledged that the work described in this paper will be widely used by the PA of the MoC and foretells that more data from the SIT can be converted and used similarly as LOD.

The proposed methodology has the advantage to rely on robust and user-friendly open-source tools, fully customisable and that avoid cumbersome calculations. The approach can be of great helping especially when the data are extremely large and divergent and integration of them is needed. In addition, the method is fully generalizable and, in principle, can be applied to

any type of non-structured database where semantics promote ease of understanding and efficiency of selection. With the shown methodology, scientists and practitioners are able to gather and convert complex, non-structured database into ontologies, and to publish them to the general public in the form of LOD. We have recently joined the W3C Location and Address Community Group[20], to share and maximise the diffusion of our experience.

In future work we plan to perform a critical analysis and comparison with other Geo Linked datasets produced by other initiatives from other countries, in order to characterise carefully the main features, and to quantify precisely the improvements, obtained with our approach. In addition we also plan to include an evaluation of specific tools designed for geo-referenced data like GeoSPARQL [17]. In particular we plan align the vocabulary to GeoSPARQL ontology, which will also allow us to use queries integrated with the geometry to gather information about polygons, lines, points, etc., since, in fact, GeoSPARQL implementations support reasoning with RCC8, the algebra of region connection calculus (RCC) with 8 basic possible relations among two geometrical regions, used for qualitative spatial representation and reasoning [17].

As the data has been published just recently, prototype applications based on them and related to services supporting transport, public health, urban decor, and social services, are currently under development. In particular in currently on-going work a mobile application based on this LOD and related to sustainable mobility and emergency vehicle routing is under implementation. This will support the real-time management of road traffic and public transport, informing citizens on the state of roads in urban areas, in particular during urban emergencies, from small accidents to more serious disasters, and redirecting the road traffic by providing best alternatives routes to find way outs, the nearest hospitals or other locations of interest. The idea is to give users the possibility to contribute traffic and other road data, sharing road reports on accidents, advising on unexpected obstacles or inaccessible zones, or any other hazards along the way, helping to give other users in the area real-time information about what is currently happening.

A feedback from users will be crucial to acknowledge the whole process, which is finally aimed to promote transparency and involving citizens and businesses into politics and PA of the MoC, as well as to improve urban life through more sustainable integrated solutions. Soon, when the first ICT services based on these LOD will be launched, user-centric tests and an experiments will be object of investigation, aimed at inspecting the external evaluation of the PA by the detection of the community trust. In particular in the near future a recommendation system will be developed to allow citizens to provide warnings and opinions to the PA, and local curators to manage and analyse the produced information flow. The scientific challenges and applications of the recommendation system will consist in:

• motivating citizens;
• analyzing the contents of the communication;
• aggregating content analysis with information on sources;
• providing feedback mechanisms that promote trust and cooperation;
• analyzing the sentiment of the community.

Our work is a concrete step supporting the MoC to move into the paradigm of Open Government, boosting the metropolis towards the route of a modern Smart City.

---

[20] http://www.w3.org/community/locadd/wiki/Main_Page

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Athanasiou, S., Hladky, D., Giannopoulos, G., Garcia Rojas, A., and Lehmann, J. 2014. GeoKnow: Making the Web an Exploratory Place for Geospatial Knowledge, ERCIM News 96, 12-13.

[2] Agency for a Digital Italy 2011. *Linee guida per i siti web delle PA*. Art. 4 della Direttiva n. 8/2009 del Ministro per la pubblica amministrazione e l'innovazione [Online] http://www.digitpa.gov.it/sites/default/files/linee_guida_siti_web_delle_pa_2011.pdf

[3] Agency for a Digital Italy 2012. *Linee guida per l'interoperabilit`a semantica attraverso Linked Open Data*. Commissione di coordinamento SPC [Online] http://www.digitpa.gov.it/sites/default/files/allegati_tec/CdC-SPC-GdL6-InteroperabilitaSemOpenData_v2.0_0.pdf

[4] Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., and Tullo, C. 2007. Unlocking the potential of public sector information with Semantic Web technology. In *Proceedings of the 6th International Semantic Web Conference (ISWC 07)*, vol. 4825 of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Berlin Heidelberg, 708–721.

[5] Baldassarre, C., Daga, E., Gangemi, A., Gliozzo, A., Salvati, A., and Troiani, G. 2010. Semantic scout: Making sense of organizational knowledge. In *Knowledge Engineering and Management by the Masses*, P. Cimiano and H. Pinto, Eds., vol. 6317 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, 272–286.

[6] Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., and Sheets, D. Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd International Semantic Web User Interaction Workshop (2006)*, SWUI 2006, Athens, USA.

[7] Bizer, C., Heath, T., and Berners-Lee, T. 2009. Linked Data The Story So Far. *International Journal on Semantic Web and Information Systems* 5, 3, 1–22.

[8] Ding, L., Difranzo, D., Graves, A., Michaelis, J., Li, X., McGuinness, D., and Hendler, J. 2010. Data-gov Wiki: Towards Linking Government Data. In *Proceedings of the AAAI 2010 Spring Symposium on Linked Data Meets Artificial Intelligence*, Palo Alto, CA, SS-10-07. AAAI Press, 38–43.

[9] Ding, L., Lebo, T., Erickson, J. S., DiFranzo, D., Williams, G. T., Li, X., Michaelis, J., Graves, A., Zheng, J. G., Shangguan, Z., Flores, J., McGuinness, D. L., and Hendler, J. 2011. A. TWC LOGD: A portal for linked open government data ecosystems. *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 3, 325 – 333.

[10] Dodsworth, E., and Nicholson, A. 2012. Academic uses of Google Earth and Google Maps in a library setting. *Information Technology and Libraries* 31, 2, 81–85.

[11] EUROCONTROL 1998. *WGS 84 implementation manual*. Institute of Geodesy and Navigation (IfEN), University FAF Munich, Germany.

[12] Gangemi, A., Daga, E., Salvati, A., Troiani, G., and Baldassarre, C. 2011. Linked Open Data for the Italian PA: the CNR Experience. *Informatica e Diritto* 1, 2.

[13] Geiger, C. P., and von Lucke, J. 2011. Open Government Data. In *CeDEM11. Conference for E-Democracy and Open Government*, P. Parycek, J. M. . Kripp, and N. Edelmann, Eds., vol. 6317 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, 183–194.

[14] Geiger, C. P., and von Lucke, J. 2012. Open Government and (Linked) (Open) (Government) (Data). *JeDEM eJournal of eDemocracy and Open Government* 4, 2.

[15] Lamb, A., and Johnson, L. 2010. Virtual expeditions: Google Earth, GIS, and geovisualization technologies in teaching and learning. *Teacher Librarian* 37, 3, 81–85.

[16] Municipality of Catania 2014 (Jan). *Il Sistema Informativo Territoriale*. [Online] http://www.sitr.provincia.catania.it:81/il-sit

[17] Perry, M., and Herring, J. 2012. OGC GeoSPARQL standard - A geographic query language for RDF data. Open Geospatial Consortium Inc, v.1.0, Available at: https://portal.opengeospatial.org/files/?artifact_id=47664

[18] Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., and Schraefel, M. 2012. Linked Open Government Data: Lessons from data.gov.uk. *Intelligent Systems, IEEE* 27, 3, 16–24.

[19] Stadler, C., Lehmann, J., Höffner, K., and Auer, S. 2012. LinkedGeoData: A Core for a Web of Spatial Open Data. Semantic Web Journal 3, 4, 333-354.

[20] Tan, P. N., Steinbach, M., and Kumar, V. 2006. *Introduction to Data Mining*. Addison-Wesley, Boston, 2006.