



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013



## **D4.1.2 –Linked Open Data – Arricchimento e pubblicazione dei dati**


**PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation**
**Rapporto Scientifico**
**Ver. FINAL 27.01.2013**

## Informazioni deliverable

<b>Titolo del deliverable</b>	D4.1.2 – Linked Open Data – Arricchimento e pubblicazione dei dati
<b>Sigla di classificazione del documento</b>	D4.1.2
<b>Responsabile del Deliverable</b>	CNR
<b>Obiettivo Realizzativo</b>	<i>OR4: Linked Open Data e Community Trust nel Cloud PAL</i>
<b>Attività relativa</b>	<i>AR4.1 – Estrazione e arricchimento di dati di un campione di PAL secondo il paradigma dei Linked Open Data</i>
<b>Autori</b>	<i>CNR (Sergio Consoli, Aldo Gangemi, Silvio Peroni, Valentina Presutti, Diego Reforgiato, Daria Spampinato)</i>
<b>Approvazione</b>	
<b>Autorizzazione</b>	

### Abstract

Il presente Deliverable descrive le tecnologie e gli strumenti per l'estrazione e la pubblicazione dei dati di un caso d'uso della Pubblica Amministrazione Locale di riferimento e la produzione di un'ontologia secondo il paradigma dei Linked Open Data.

I metodi sono basati sugli standard del W3C, su buone pratiche internazionali, sulle linee-guida emanate dall'Agenzia per l'Italia Digitale e su esperienze vicine già avute dai partecipanti a questa attività, in particolare lo sviluppo di data.cnr.it e del Semantic Scout, nonché dei linked open data dell'Indice delle Pubbliche Amministrazioni.

È stata sviluppata un'ontologia a partire dai dati del Sistema Informativo Territoriale del Comune di Catania.



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

I contenuti del documento sono da intendersi “confidenziali” di proprietà esclusiva delle parti coinvolte nel progetto “PRISMA – PiattafoRme cloud Interoperabili per SMArt-government”.

Né il documento né le sue parti possono essere pubblicate, riprodotte, copiate o comunque divulgate senza autorizzazione scritta delle parti coinvolte nel progetto oltre quanto previsto dalla lista di distribuzione

## Lista di distribuzione

### Stato del deliverable

Ver.	Data	Autore modifica	dellaNote	Validazione



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## INDICE DEGLI ARGOMENTI

INDICE DELLE FIGURE .....	4
INDICE DELLE TABELLE .....	4
ACRONIMI.....	4
NOTAZIONE.....	4
EXECUTIVE SUMMARY .....	4
INTRODUZIONE .....	4
METODOLOGIE E STRUMENTI PER L'ESTRAZIONE, LA MODELLAZIONE E LA PUBBLICAZIONE DI LINKED OPEN DATA DELLA PA .....	4
Metodologia all'interoperabilità semantica tramite LOD .....	4
Analisi dello scenario di riferimento e dei requisiti.....	4
Strumenti per l'estrazione e tecnologie per la modellazione dei dati.....	4
DESCRIZIONE DELLA ONTOLOGIA PER LA TOPONOMASTICA DEL SIT .....	4



## Indice delle Figure

Figure 1. Due diagrammi UML che mostrano la codifica OWL del design pattern <i>time interval</i> .	15
Figure 2. L'idea di progettazione di ontologie basata su <i>pattern</i> .	16
Figure 3. ( a ) Il processo XD; ( b ) l'iterazione eseguita da coppie di progettisti.....	17
Figura 4. Lo stack di metodi usati per la realizzazione di data.cnr.it e del Semantic Scout. ....	20
Figure 5. Entità principali dell'ontologia Asili Nido. Le frecce spesse indicano relazioni di sottoclasse, quelle fini e nere o blu relazioni possibili fra individui di classi diverse e frecce rosse per relazioni di disgiunzione. Gli attributi all'interno dei rettangoli sono relazioni possibili con i relativi vincoli di cardinalità. ....	27
Figure 6. Un grafo dei dati relativi a un asilo nido specifico (profilo OWL per UML). Le relazioni fra i nodi rappresentano relazioni fra individui o di appartenenza a una classe. Gli attributi all'interno dei rettangoli sono relazioni fra individui e valori simbolici (ex. nomi, commenti, etc.). ....	28
Figure 7. Entità principali dell'ontologia Assistenza Domiciliare Integrata.....	29



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## **Indice delle Tabelle**

**Non è stata trovata alcuna voce dell'indice delle figure.**



## Rapporto Scientifico

Ver. FINAL 27.01.2013

## Acronimi

PRISMA PiattafoRme cloud Interoperabili per SMARt-government

SIT                      Sistema Informativo Territoriale



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## Notazione





PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## Executive Summary

Il presente Deliverable descrive le tecnologie e gli strumenti per l'estrazione e la pubblicazione dei dati di un caso d'uso della Pubblica Amministrazione Locale di riferimento e la produzione di un'ontologia secondo il paradigma dei Linked Open Data.

I metodi sono basati sugli standard del W3C, su buone pratiche internazionali, sulle linee-guida emanate dall'Agenzia per l'Italia Digitale e su esperienze vicine già avute dai partecipanti a questa attività, in particolare lo sviluppo di data.cnr.it e del Semantic Scout, nonché dei linked open data dell'Indice delle Pubbliche Amministrazioni.

È stata sviluppata un'ontologia a partire dai dati del Sistema Informativo Territoriale del Comune di Catania.



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## Introduzione

Questo documento descrive la metodologia utilizzata per l'estrazione, l'arricchimento, la pubblicazione e il riuso di *linked open data* dei dati della Pubblica Amministrazione Locale (PAL) attraverso l'applicazione di tecnologie semantiche e componenti software.

La Pubblica Amministrazione soggetto sperimentatore del progetto previsto dal paragrafo 4.10 della seconda parte del progetto esecutivo PRISMA è il Comune di Catania.

Uno dei principali obiettivi realizzativi del progetto consiste nell'ideazione, progettazione e realizzazione di prototipi applicativi riferiti ad alcuni ambiti di sperimentazione.

Al fine del recupero dei dati e dei processi sono state realizzate alcune riunioni conoscitive con la Dirigenza della Direzione del Servizio Sistemi Informativi del Comune di Catania.

L'azione dell'AR4 si è al momento concentrata su un ambito di sperimentazione: gestione del traffico e dei trasporti - mobilità. All'interno di questo contesto è stato individuato lo scenario per un prototipo applicativo ("app") denominato "Catania now", un sistema per informare in tempo reale sullo stato della viabilità in ambito urbano per il supporto alla mobilità sostenibile. Per la realizzazione del prototipo di app, che è uno degli obiettivi realizzativi del progetto e che verrà attuata durante la fase di sperimentazione, è necessario trattare i dati e gli schemi contenuti nel Sistema Informativo Territoriale del Comune di Catania. Si è deciso, di comune accordo con gli altri partner di progetto, di trattare i dati in modo da renderli aperti, interoperabili e compatibili con i principi dei Linked Open Data.



## 1. Metodologie e strumenti per l'estrazione, la modellazione e la pubblicazione di Linked Open Data della PA

I metodi sono basati sugli standard semantici del [W3C](#), su buone pratiche internazionali, sulle [linee-guida](#) emanate dall'[Agenzia per l'Italia Digitale](#) e su esperienze vicine già avute dai partecipanti a questa attività, in particolare lo sviluppo di [data.cnr.it](#) e del [Semantic Scout](#), nonché dei *linked open data* dell'[Indice delle Pubbliche Amministrazioni](#).

Gli standard semantici, le buone pratiche di design e le linee guida emanate dall'Agenzia per l'Italia Digitale sono state descritte nel Deliverable 4.1.1. Qui verranno descritte le varie fasi della metodologia all'interoperabilità semantica e gli strumenti usati per l'estrazione dei dati dal SIT e per la loro modellazione e pubblicazione tramite linked open data.

### 1.1 Metodologia all'interoperabilità semantica tramite LOD

Le linee guida per l'interoperabilità semantica attraverso i Linked Open Data, emanate dall'Agenzia per l'Italia Digitale a fine 2012, costituiscono un ottimo sistema di riferimento per lo sviluppo di *linked open data* nelle PAL. Le buone pratiche suggerite sono state recepite nel lavoro concreto con la PAL del Comune di Catania in riferimento ai dati e agli schemi del SIT.

La metodologia proposta si sviluppa in sette fasi successive:

1. individuazione e selezione dei dataset;
2. bonifica;
3. analisi e modellazione;
4. arricchimento;
5. linking esterno (interlinking);
6. validazione;
7. pubblicazione.



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## **1.2 Analisi dello scenario di riferimento e dei requisiti**

Durante la fase di individuazione e selezione dei dataset è stata svolta un'analisi del dominio di riferimento.

Per quel che concerne l'arricchimento e la pubblicazione dei dati delle Pubbliche Amministrazioni Locali (PAL), è stato deciso insieme al Comune di Catania di usare alcuni dati estratti dal Sistema Informativo Territoriale (SIT) (1) della provincia di Catania. Il SIT è un sistema composto da banche dati, hardware, software e tecnici, che gestisce, elabora e integra informazioni su una base spaziale geografica. I vari livelli territoriali (idrografia, orografia, edifici, infrastrutture, reti tecnologiche, limiti amministrativi e catastali, etc.) costituiscono una parte comune, quella geografica, in base alla quale possono essere messi in relazione tra loro.

Il SIT ha lo scopo di contenere tutti i dati disponibili dell'Amministrazione ai fini della conoscenza approfondita del territorio. Fondamentalmente il SIT contiene tre tipologie di dati: il db catastale, l'anagrafe e la toponomastica. Attraverso il sistema di consultazione via web è attualmente possibile visualizzare:

- la cartografia di base;
- le ortofoto;
- il grafo stradale;
- gli edifici con la suddivisione per corpo di fabbrica di alcune zone della città;
- le sezioni Censuarie;
- i dati del Censimento 1991 e 2001;
- l'ultimo Piano regolatore ma senza tutte le varianti;
- la localizzazione e le informazioni alfanumeriche di:



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

- Municipalità;
  - Ospedali;
  - Università;
  - Scuole;
  - Farmacie;
  - Uffici Postali;
  - Aree Emergenza;
  - Pubblica Sicurezza;
  - Vigili del Fuoco;
  - Verde Pubblico;
  - Sedi dei Centri Sociali;
  - Istituti per minori;
- La popolazione residente in aree selezionate (municipalità, intera via, area poligonale, area circolare); la popolazione complessiva, distribuita per arco stradale, numero civico, e altro; la suddivisione della popolazione per: municipalità, isolati, cittadinanza, sesso, componenti per famiglia, fasce d'età, stato civile,... ;
  - Estrazione e ricerche dei soggetti residenti, e loro localizzazione sugli archi stradali (chiaramente solo per operatori abilitati);
  - La gestione delle aree di competenza delle farmacie;
  - I lavori in esecuzione della rete gas.

Il SIT inoltre include le mappe del sottosuolo che contengono le informazioni geo-referenziate relative

a:



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

- sotto-servizi (condotte gas-acqua-elettriche);
- dati sosta;
- occupazione stalli Sostare;
- dati stalli disabili;
- dati dell' occupazione del suolo pubblico;
- la centrale AVL-AVM;
- la flotta AMT;
- dati incidentalità;
- segnaletica orizzontale e verticale;
- la manutenzione strade e marciapiedi;
- la gestione cantieri su sede stradale;
- dati polizia della municipale;
- dati delle linee e fermate AMT;
- la ragioneria del Comune.

In questa prima fase, il Comune di Catania ha fornito allo STLab-CNR una parte dei dati relativi alla toponomastica contenuti nel SIT.

Tali dati riguardano 32 tabelle (mostrate in Figura 1) fornite sotto forma di shapefiles. Ogni tabella è rappresentata da files con le seguenti estensioni (.dbf, .shp, shx, .sbn, .sbx, .shp.xml).



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

ArchiStradali	ContornoDensità	PubblicaSicurezza	ServiziSocialiSedi
AreeEmergenza	Farmacie	ReteFibraOttica	SezioniCensimento1991
Carceri	Isolati	ReteGas	SezioniCensimento2001
CaseRiposo	Municipalità	ScuoleAree	UfficiComunali
CentralineSmog	NumeriCivici	ScuolePunti	Università
Chiese	Ospedali	Semafori	UtentiWAN
Circoscrizioni	Poste	SerbatoilIdrici	VerdeAree
ConfiniComunali	PRG	ServiziSocialiAree	WWFF

Figura 1 – Tabelle del SIT messe a disposizione dal Comune di Catania

### 1.3 Strumenti per l'estrazione e tecnologie per la modellazione dei dati

In questa fase della metodologia di lavoro si possono utilizzare strumenti che facilitano la reingegnerizzazione del dataset secondo il nuovo modello concettuale dei dati.

Alcuni strumenti sono:

TOPBRAID COMPOSER (2), contiene vari strumenti per l'importazione di dati da formati diversi su RDF; supporta anche OWL.

D2R (Database to RDF) (3), è un framework per accedere a database relazionali e produrre un "RDF dump" secondo determinate specifiche.

TABELS (4), uno strumento software di ricerca sviluppato dalla fondazione CTIC<sup>1</sup>, che, usando le librerie GeoTools (5), è capace di trasformare le informazioni codificate negli shape files in rappresentazioni RDF.

<sup>1</sup> <http://www.fundacionctic.org/en>



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

A partire dagli shape file di ogni tabella (in particolare i file con estensioni .dbf e .shp) è stato usato Tabela per codificare gli shape file in triple RDF parziali conformi all'ontologia di bozza descritta nella sezione successiva. Da un lato vengono immagazzinate le caratteristiche della tabella come rappresentazione RDF; dall'altro, le geometrie spaziali vengono modellate secondo la rappresentazione standard KML. Le coordinate geometriche in KML sono espresse secondo il sistema geodetico di riferimento Gauss-Boaga (o Roma 40). Tramite strumenti di conversione, come per esempio (5), si riescono a produrre le coordinate di latitudine, longitudine e i metri di quota nel sistema geodetico WGS84 (6).

In particolare, l'applicazione di Tabela a ciascuna coppia di file .dbf e .shp delle 32 tabelle riesce a produrre un insieme di triple RDF immagazzinate in una repository con altre risorse geometriche immagazzinate in un server pubblico.

Per esempio, a partire da informazioni memorizzate nel database del SIT che rappresentano le farmacie, come si può vedere in figura 2, vengono prodotte le triple RDF (figura 3) e il file con le coordinate geometriche KML di figura 4.





PON04a2\_A -PON04a2\_A / F– Settore mart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013



**Figura 2 – Rappresentazione grafica su Google earth dei punti indicanti le farmacia del SIT del Comune di Catania**

```

<http://www.essepuntato.it/2013/10/prisma/resource/farmacia/finocchiario-giuseppa>
a
<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>
"FINOCCHIARIO 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>
"FINOCCHIARIO 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> .

```

**Figura 3 – Triple RDF per una farmacia del Comune di Catania**



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

```
<?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>
```

**Figura 4 – Coordinate geometriche KML di una farmacia del Comune di Catania**

Per quanto riguarda il funzionamento di Tabela, esso importa formati comuni come XLS o CSV. Successivamente esso riesce a generare un programma di trasformazione in automatico dai dati di input. In programma generato riesce a trasformare ogni riga dei dati in ingresso in una nuova istanza RDF di una classe ad-hoc; inoltre ogni valore di colonna delle tabelle in input è convertito in una nuova tripla dove il soggetto è l'istanza menzionata, il predicato è una proprietà ad-hoc basata sul nome dell'header della colonna e l'oggetto è il valore della colonna come un *rdfs:Literal*. E' da sottolineare che il programma di trasformazione generato in automatico da Tabela è uno script SPARQL completamente customizzabile dall'utente. Ovvero è possibile modificare i nomi associati alle classi e proprietà, annotarle appropriatamente, etc. Una volta che il programma di trasformazione è definito, l'esecuzione di Tabela genera l'RDF di output, che poi abbiamo reso disponibile nello SPARQL Endpoint dedicato (cf. Deliverable D4.2.1).



**PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation**

**Rapporto Scientifico**

**Ver. FINAL 27.01.2013**



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## 2. Descrizione dell'ontologia per la toponomastica del SIT

URI (indirizzo dell'ontologia):

<http://www.essepuntato.it/2013/prisma/>

Namespace (indirizzo di default delle entità definite nell'ontologia):

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

Live OWL Documentation Environment (documentazione dell'ontologia):

<http://www.ontologydesignpatterns.org/dati/prisma/>

L'ontologia fornisce classi e proprietà per rappresentare le entità del database del SIT.

A partire dalla definizione delle 32 tabelle della toponomastica è stata sviluppata una prima versione dell'ontologia OWL. Il processo di creazione di questa ontologia è stato suddiviso in due fasi e ha seguito le buone pratica di rappresentazione formale, denominazione e assunzioni semantiche in uso nel dominio del Web Semantico e dei Linked Open Data, descritti già in buona parte nel deliverable D4.1.1.

In primo luogo, l'intera struttura delle tabelle è stata convertite in una bozza di ontologia OWL, dove ogni tabella (ovvero ogni tipo di entità descritto dalla base di dati fornita) è rappresentata da una classe e ogni campo di tabella è stato convertito in data property. Questa traduzione è stata svolta in modo del tutto automatico a partire dai sorgenti in formato XML forniti (estensione .shp.xml) mediante l'utilizzo di una trasformazione XSLT. Notare che a campi aventi lo stesso nome ma appartenenti a tabelle diverse sono state corrisposte proprietà distinte. Ad esempio, i campi "NOME" delle tabelle "CaseRiposo" e "Farmacie" sono stati tradotti con due diverse data property, rispettivamente "NOME-



**PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation**

**Rapporto Scientifico**

**Ver. FINAL 27.01.2013**

of-CATANIA.SDO\_CaseRiposo” e “NOME-of-CATANIA.SDO\_CaseFarmacie”. Un estratto (in RDF/XML) di questa traduzione preliminare è mostrato in Figura 5.





PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

```

- <owl:Class rdf:about="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori">
  <rdfs:label>CATANIA.SDO_semafori</rdfs:label>
</owl:Class>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/OBJECTID-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>OBJECTID</rdfs:label>
  <rdfs:range rdf:resource="http://www.essepuntato.it/2013/10/prisma/oid"/>
  <rdfs:comment>Internal feature number.</rdfs:comment>
  <rdfs:comment>ESRI</rdfs:comment>
- <rdfs:comment>
  Sequential unique whole numbers that are automatically generated.
</rdfs:comment>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/Shape-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>Shape</rdfs:label>
  <rdfs:range rdf:resource="http://www.essepuntato.it/2013/10/prisma/geometry"/>
  <rdfs:comment>Feature geometry.</rdfs:comment>
  <rdfs:comment>ESRI</rdfs:comment>
  <rdfs:comment>Coordinates defining the features.</rdfs:comment>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/Id-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>Id</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double"/>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/NOME-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>NOME</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/sde_SDE_sc-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>sde_SDE_sc</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/SE_ANNO_CAD_DATA-of-CATANIA.SDO_semafori">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_semafori"/>
  <rdfs:label>SE_ANNO_CAD_DATA</rdfs:label>
  <rdfs:range rdf:resource="http://www.essepuntato.it/2013/10/prisma/blob"/>
</owl:DatatypeProperty>
- <owl:Class rdf:about="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_Serbatoildrici">
  <rdfs:label>CATANIA.SDO_Serbatoildrici</rdfs:label>
</owl:Class>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/OBJECTID_1-of-CATANIA.SDO_Serbatoildrici">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_Serbatoildrici"/>
  <rdfs:label>OBJECTID_1</rdfs:label>
  <rdfs:range rdf:resource="http://www.essepuntato.it/2013/10/prisma/oid"/>
  <rdfs:comment>Internal feature number.</rdfs:comment>
  <rdfs:comment>ESRI</rdfs:comment>
- <rdfs:comment>
  Sequential unique whole numbers that are automatically generated.
</rdfs:comment>
</owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:about="http://www.essepuntato.it/2013/10/prisma/Shape-of-CATANIA.SDO_Serbatoildrici">
  <rdfs:domain rdf:resource="http://www.essepuntato.it/2013/10/prisma/CATANIA.SDO_Serbatoildrici"/>
  <rdfs:label>Shape</rdfs:label>
  <rdfs:range rdf:resource="http://www.essepuntato.it/2013/10/prisma/geometry"/>
  <rdfs:comment>Feature geometry.</rdfs:comment>
  <rdfs:comment>ESRI</rdfs:comment>
  <rdfs:comment>Coordinates defining the features.</rdfs:comment>
</owl:DatatypeProperty>

```



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

## Figura 5 – Estratto della traduzione delle 23 tabelle della base di dati nell'ontologia intermedia di bozza.

A partire da questa bozza intermedia e dai dati a disposizione, è stata realizzata una prima versione dell'ontologia in OWL. Sono stati seguiti i seguenti principi:

- Il nome di tutte le classi è stato portato al singolare (es. da “Farmacie” a “Farmacia”);
- Il nome delle data property è stato uniformato quando i dati mostravano chiaramente la medesima semantica. Ad esempio le proprietà “NOME-of-CATANIA.SDO\_CaseRiposo” e “NOME-of-CATANIA.SDO\_CaseFarmacie” sono confluite nella proprietà “nome”, avente come dominio o individui della classe “CasaRiposo” o individui della classe “Farmacia”;
- Le data property (probabilmente aventi funzioni di *foreign key* sulla base di dati) che sembravano fare riferimento a individui di altre classi sono state trasformate in object property. Ad esempio, la proprietà “MUNI-of-CATANIA.SDO\_CaseRiposo” è diventata “municipalita” in modo da collegare individui della classe “CasaRiposo” con individui della classe “Municipalita”;
- Le data property aventi chiaramente come valore delle vere e proprie risorse sono state trasformate in object property e i loro valori sono stati “reificati” come individui propri di classi create appositamente.

## Dai dati RDF conformi alla bozza all'ontologia finale

Tutte le modifiche apportate all'ontologia intermedia di bozza per la realizzazione della prima versione dell'ontologia finale sono state documentate sotto forma di SPARQL CONSTRUCT. Questo ci ha



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

permesso di realizzare un semplice script per la conversione dei dati estratti mediante Tabela in modo da renderli totalmente conformi con l'ontologia finale prodotta.

### **Esempio di trasformazione da shape files a RDF**

Prendiamo come riferimento una delle 32 tabelle di Figura 1, la tabella Semafori. Lo schema SQL di questa tabella include i seguenti campi:

- OBJECTID – numero univoco incrementato sequenziale;
- Shape – tipo geometria che rappresenta le coordinate che definiscono le caratteristiche geometriche;
- Id – di tipo double;
- Nome – di tipo Stringa;
- Sde\_SDE\_se – di tipo intero;
- Se\_ANNO\_CAD\_DATA – di tipo blob.

Passando i file .shp e .dbf a Tabela, quest' ultimo genera il programma di trasformazione, ovvero lo script in SPARQL usato l'importazione dei dati (vedere Figura 6). Come già menzionato, è possibile editare tale script in base alle proprie esigenze.





PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

```

1 PREFIX project: <http://idi.fundacionctic.org/tabela/project/Semafori/>
2 PREFIX my: <http://idi.fundacionctic.org/tabela/project/Semafori/resource/>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX dc: <http://www.w3.org/ns/dc#>
6 PREFIX dct: <http://purl.org/dc/terms/>
7 PREFIX foaf: <http://xmlns.com/foaf/0.1/#>
8
9 FOR ?rowId IN rows FILTER get-row(?rowId)
10 LET ?resource = resource(?rowId,<http://idi.fundacionctic.org/tabela/project/Semafori/resource/>)
11 MATCH [?sUBJECTID,?ID,?sNAME,?sDESDISE,?geometry,?kml] IN horizontal
12 CONSTRUCT {
13   my:TabelaDataCatalog a dc:Catalog .
14   my:TabelaDataCatalog dct:title "Tabela AutoGenerated Catalog" .
15   my:TabelaDataCatalog dct:description "Tabela AutoGenerated Catalog" .
16   my:TabelaDataCatalog dct:publisher my:TabelaAutoGenerator .
17   my:TabelaDataCatalog dc:dataset my:DataSet
18 }
19
20 CONSTRUCT {
21   my:TabelaAutoGenerator a foaf:Project .
22   my:TabelaAutoGenerator rdfs:label "Tabela autogenerator program" .
23   my:TabelaAutoGenerator foaf:homepage project:index
24 }
25
26 CONSTRUCT {
27   my:DataSet a dc:Dataset .
28   my:DataSet dct:title "Semafori data sets" .
29   my:DataSet dct:description "Default description for autogeneratedSemaforidata sets" .
30   my:DataSet dct:identifier "Semafori" .
31   my:DataSet dc:keyword "tabela" .
32   my:DataSet dc:distribution my:DataSetRDF .
33   my:DataSet dc:distribution my:DataSetTurtle .
34   my:DataSet dc:distribution my:DataSetN3 .
35   my:DataSet dct:publisher my:TabelaAutoGenerator .
36   my:DataSet dc:landingPage project:index
37 }
38
39 CONSTRUCT {
40   my:DataSetRDF a dc:Distribution .
41   my:DataSetRDF dct:title "SemaforiRDF/XML Distribution" .
42   my:DataSetRDF dct:description "Default description for autogeneratedSemaforiRDF/XML Distribution" .
43   my:DataSetRDF dc:accessURL project:spqrql .
44   my:DataSetRDF dc:downloadURL project:data .
45   my:DataSetRDF dc:format "RDF" .
46   _:B0 a dct:INT .
47   _:B0 rdf:value "application/rdf+xml" .
48   _:B0 rdfs:label "RDF+XML"
49 }
50
51 CONSTRUCT {
52   my:DataSetTurtle a dc:Distribution .
53   my:DataSetTurtle dct:title "SemaforiTurtle Distribution" .
54   my:DataSetTurtle dct:description "Default description for autogeneratedSemaforiTurtle Distribution" .
55   my:DataSetTurtle dc:accessURL project:spqrql .
56   my:DataSetTurtle dc:downloadURL project:data?format=ttl .
57   my:DataSetTurtle dc:format "TTL" .
58   _:B1 a dct:INT .
59   _:B1 rdf:value "text/turtle" .
60   _:B1 rdfs:label "TURTLE"
61 }
62
63 CONSTRUCT {
64   my:DataSetN3 a dc:Distribution .
65   my:DataSetN3 dct:title "SemaforiN3 Distribution" .
66   my:DataSetN3 dct:description "Default description for autogeneratedSemaforiN3 Distribution" .
67   my:DataSetN3 dc:accessURL project:spqrql .
68   my:DataSetN3 dc:downloadURL project:data?format=text .
69   my:DataSetN3 dc:format "N3" .
70   _:B2 a dct:INT .
71   _:B2 rdf:value "text/n3" .
72   _:B2 rdfs:label "N3"
73 }
74
75 CONSTRUCT {
76   ?resource a my:CATANIA.SEO_semafori .
77   ?resource my:OBJECTID ?sUBJECTID .
78   ?resource my:Id ?ID .
79   ?resource my:NOME ?sNAME .
80   ?resource my:sde_SDE as ?sDESDISE .
81   ?resource my:Shape ?geometry .
82   ?resource my:kml ?kml
83 }
84
85 CONSTRUCT {
86   my:CATANIA.SEO_semafori a rdfs:Class
87 }
88
89 CONSTRUCT {
90   my:OBJECTID a rdf:Property .
91   my:Id a rdf:Property .
92   my:NOME a rdf:Property .
93   my:sde_SDE as a rdf:Property .
94   my:Shape a rdf:Property .
95   my:kml a rdf:Property
96 }
97

```



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

***Figura 6 – Programma di trasformazione dagli shapes files a RDF per la tabella Semafori***

Una volta conclusa l'eventuale modifica del programma di trasformazione, è possibile cliccare sul pulsante “*Save program*” che genera le triple RDF dai dati della tabella in input.

In Figura 4 viene mostrata la parte iniziale del file “*Semafori.ttl*”, che corrisponde alle triple in formato RDF/Turtle prodotte da Tabela usando la metodologia appena descritta per la tabella di riferimento “*Semafori*”.



PON04a2\_A -PON04a2\_A / F- Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

```
@prefix project: <http://idi.fundacionctic.org/tabels/project/Semafori/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix my: <http://idi.fundacionctic.org/tabels/project/Semafori/resource/> .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/28>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 27 ;
  my:NOME "Giuffrida - Brancati" ;
  my:OBJECTID 28 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.28.kml" ;
  my:sde_SDE_se 1 .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/14>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 12 ;
  my:NOME "Monserrato-Imbriani-D Annunzio" ;
  my:OBJECTID 14 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.14.kml" ;
  my:sde_SDE_se 1 .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/34>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 30 ;
  my:NOME "Sicilia - Puccini" ;
  my:OBJECTID 34 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.34.kml" ;
  my:sde_SDE_se 0 .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/5>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 5 ;
  my:NOME "Etnea-R.Margherita" ;
  my:OBJECTID 5 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.5.kml" ;
  my:sde_SDE_se 1 .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/18>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 13 ;
  my:NOME "Etnea - Borgo - Monserrato" ;
  my:OBJECTID 18 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.18.kml" ;
  my:sde_SDE_se 1 .

<http://idi.fundacionctic.org/tabels/project/Semafori/resource/24>
  a <http://idi.fundacionctic.org/tabels/project/Semafori/resource/CATANIA.SDO_semafori> ;
  my:Id 21 ;
  my:NOME "Beato Angelico" ;
  my:OBJECTID 24 ;
  my:Shape "http://www.w3.org/2003/01/geo/wgs84_pos#Point" ;
  my:kml "http://idi.fundacionctic.org/tabels/project/Semafori/kml/Semafori.24.kml" ;
  my:sde_SDE_se 3 .
```



PON04a2\_A -PON04a2\_A / F– Settore smart Cities and Communities and Social Innovation

Rapporto Scientifico

Ver. FINAL 27.01.2013

***Figura 4 – Triple in RDF/Turtle prodotte dal programma di trasformazione di Tabela per la tabella “Semafori”***

## Referenze

- [1] <http://www.sitr.provincia.catania.it:81/il-sit>
- [2] <http://www.topbraidcomposer.com>
- [3] <http://d2rq.org/d2r-server>
- [4] <http://idi.fundacionctic.org/tabela/>
- [5] [http://www.ultrasoft3d.net/Conversione\\_Coordinate.aspx](http://www.ultrasoft3d.net/Conversione_Coordinate.aspx)
- [6] ["Manuale d'implementazione del WGS 84", EUROCONTROL - IfEN, v. 2.4, 1998.](#)