Smart City data via LOD/LOG Service

**P. Bellini, P. Nesi, N. Rauch**

Dipartimento di Ingegneria dell’Informazione, DINFO
Università degli studi di Firenze
Via S. Marta 3, 50139, Firenze, Italy
tel: +39-055-4796567, fax: +39-055-4796363

DISIT Lab

http://www.disit.dinfo.unifi.it/ alias http://www.disit.org
nadia.rauch@unifi.it

Slides for: LOD2014 event.
Research objectives

• **Why**: Create an ontology that allows to combine all data provided by the *city of Florence* and the *Tuscan region*.

• **Problems**: data have different formats, they must be reconciled in order to be effectively interconnected to each other, but sometimes information is incomplete.

• **Objective**: take advantage of the created repository and ontology to implement new integrated services related to mobility; to provide repository access to *SMEs* to create new services.
Analysis of Available Data

• 519 OpenData (Municipality of Florence)
• 145 OpenData (Tuscany Region)
• LPT Timetable and LPT Route
• Street Graph
• Points of Interest
• Real Time Data from traffic sensors
• Real Time Data from parking sensors
• Real Time Data from AVM systems
• Weather Forecast (consortium Lamma)
DataSet already integrated

• From MIIC web services (real time)
  o Parking payloadPublication (updated every 6h)
  o Traffic sensors payloadPublication (updated every 5-10min)
  o AVM client pull service (updated every 24h)
  o Street Graph

• From Municipality of Florence:
  o Tram line: KMZ file that represents the path of tram in Florence
  o Statistics on monthly access to the LTZ, tourist arrivals per year, annual sales of bus tickets, accidents per year for every street, number of vehicles per year
  o Municipality of Florence resolutions

• From Tuscany Region:
  o Museums, monuments, theaters, libraries, banks, courier services, police, firefighters, restaurants, pubs, bars, pharmacies, airports, schools, universities, sports facilities, hospitals, emergency rooms, doctors' offices, government offices, hotels and many other categories
  o Weather forecast of the consortium Lamma (updated twice a day)
Ontology’ Macroclasses

• Maps and Geographical information: formed by classes Road, Node, RoadElement, AdministrativeRoad, Milestone, StreetNumber, RoadLink, Junction, Entry, and EntryRule, Manoeuver, is used to represent the entire road system of Tuscany region.

• Point of Interest: economical services (public and privates), activities, which may be useful to the citizen and who may have the need to search for and to arrive at. Classification will be based on the division into categories planned at regional level.

• Weather: including status and forecasts from the consortium Lamma in Tuscany.
Ontology’ Macroclasses

- **Transport**: data coming from major LPT companies including scheduled times, the rail graph, data relating to real time passage at bus stops. *Classes*: bus line, Ride, Route, record, RouteSection, BusStopForecast, RouteLink.

- **Sensors**: concerning data coming from sensors; they may include information such as pressure, humidity, pollution, car flow, car velocity, number of passed cars and tracks, etc.

- **Administration**: includes information coming from public administrations such as resolutions issued by each administration, planned events, changes in the traffic arrangement, planned VIP visits, sports events, etc.
Maps Macroclass

- **RoadElement**: delimited by a start node and an end node (ObjectProperties "starts" e "ends");
- **Road**: composed by RoadElement and Node ("contains")
- **AdministrativeRoad**: connected to RoadElement ("isComposed" e "forming"), to Road ("coincideWith").
  
  \[ \text{Road : AdministrativeRoad} = N:M. \text{ Both in a 1:N relation with RoadElement; } \]

- **EntryRule**: connected to RoadElement ("hasRule", "accessTo ");
- **Maneouvre**: linked to EntryRule ("isDescribed").
  Described through "hasFirstElem", "hasSecondElem" and "hasThirdElem". "concerning" fastes a maneouvre to the concerned junction.
Maps Macroclass

- **Node**: georeferenced through geo:lat and geo:long.
- **Milestone**: associated with 1 AdministrativeRoad ("placedIn"), georeferenced through geo:lat and geo:long.
- **StreetNumber**: always related to at least 1 entry (internal or external). Connected to RoadElement and Road ("standsIn" and "belongsTo"); reverse:"hasStreetNumber".
- **Entry**: connected to StreetNumber through "hasInternalAccess" and "hasExternalAccess", with cardinality restrictions, subclass of geo:SpatialThing, maximum cardinality restriction 1 to geo:lat and geo:long.
- "ownerAuthority" and "managingAuthority": linked to PA macroclass.
Maps Macroclass

- otn:Geometric
- otn:Edge
- otn:Node

Road
- isPartOf: RoadElement
- contains: Node

RoadElement
- starts
- ends

Node
- hasRule
- accessTo: Maneuver

Milestone
- situated

Junction
- ending
- starting

RoadLink
- hasSegment
- placedIn: StreetNumber

AdministrativeRoad
- isComposed
- forming

Entry
- hasInternalAccess
- hasExternalAccess

EntryRule
- isDescribed

StreetNumber
- hasStreetNumber

Maneuver
- hasFirstElem
- hasSecondElem
- hasThirdElem

hasRule
- hasRule
Reused Vocabulary

- **OTN**: an ontology of traffic networks that is more or less a direct encoding of GDF (Geographic Data Files) in OWL;

- **dcterms**: set of properties and classes maintained by the Dublin Core Metadata Initiative;

- **foaf**: dedicated to the description of the relations between people or groups;

- **vCard**: for a description of people and organizations;

- **wgs84_pos**: vocabulary representing latitude and longitude, with the WGS84 Datum, of geo-objects.
Macroclasses’ Connections

- Administration Macroclass
- Street-guide Macroclass
- Point of Interest Macroclass
- Local public transport Macroclass
- Sensors Macroclass
- Temporal Macroclass
From Open Data to Triples

• **Phase 1**: collect data from different sources (MIIC Web Service, Osservatorio dei Trasporti e della Mobilita’ portal, Municipality of Florence and Tuscany Region Web Sites).

• **Phase 2**: first processing means ETL tool and NoSQL database storage.

• **Phase 3**: second transformation using ETL tools and RDF triples creation.

• **Phase 4**: Saving triple in RDF store.
Helpful Tools

- ETL Transformation

- To realize the R2RML model

- RDF Store
Architecture

• To automate the different phases, we have created an **architecture** that includes a **process scheduler**.

• The process scheduler implementation was necessary to **repeat the 4 phases**, from ingestion to transformation in triple.

• We storing data in Hbase according to a programmed rate, which is closely linked to the type of data (static/real time):
  - Real-time data: every 10min;
  - Other data: 2 - 15 times a day;
  - Static data: once a month or more.
Architecture’ Block Diagram
Data Validation & Reconciliation

• Major problems with the data:
  o inconsistent data (different municipality to the same service, city names that are not a municipality)
  o missing data (street number)
  o incorrect data (spelling errors)

• Need to validate the data, but above all to reconcile them to be able to connect with each other:
  o Service – Street Name Reconciliation
  o Service – Coordinate Reconciliation
Reconciliation Numbers

- **Services**: ~ 30,100 (all over Tuscan region) of which:
  - Geolocalized Services: ~ 12,400
  - Services located at street level: ~ 8,300
- **Remaining Services**: ~ 9,000 of which:
  - Non-unique results to locate the service at street level
  - Street Number missing
  - Unusual letters in municipality names or street names
  - Address does not exist on Street Graph: ~ 2,200 (next step: use the Google geocoding API)
Real Time Data Numbers

- **Weather**: 286 files uploaded twice a day → 270,000 Hbase rows/month → ~4 million triples/month;

- **Sensors**: 126 active sensors → 18,000 Hbase rows/day, 50 supervised parking → ~10GB/month;

- **Street Graph**: 68M triples.

- **For an amount of ~ 80MTriples** on repository
App Examples

• **Linked Open Graph (LOG):** a tool developed to allow exploring semantic graph of the relation among the entities. It can be used to access to many different LOD repository. ([http://log.disit.org/](http://log.disit.org/))

• **Maps:** service based on OpenStreetMaps that allows to search services available in a preset range from the selected bus stop. ([http://servicemap.sii-mobility.org/](http://servicemap.sii-mobility.org/))
http://servicemap.sii-mobility.org
Future Works

• Integration of rail graph into the ontology;
• Insertion of other static datasets from the municipality of Florence and other Tuscany PA;
• Using Google Geocoding API to finish services reconciliation;
• Improvement of services’ list and their geolocation;
• Creation of other apps that suggest to SME and PA how to use data.