

Big Data Smart City processes and tools

Distributed Data Intelligence and Technologies Lab
Distributed Systems and Internet Technologies Lab

Prof. Paolo Nesi

DISIT Lab

Dipartimento di Ingegneria dell'Informazione

Università degli Studi di Firenze

Via S. Marta 3, 50139, Firenze, Italia

tel: +39-055-2758515, fax: +39-055-2758570

<http://www.disit.dinfo.unifi.it>

paolo.nesi@unifi.it

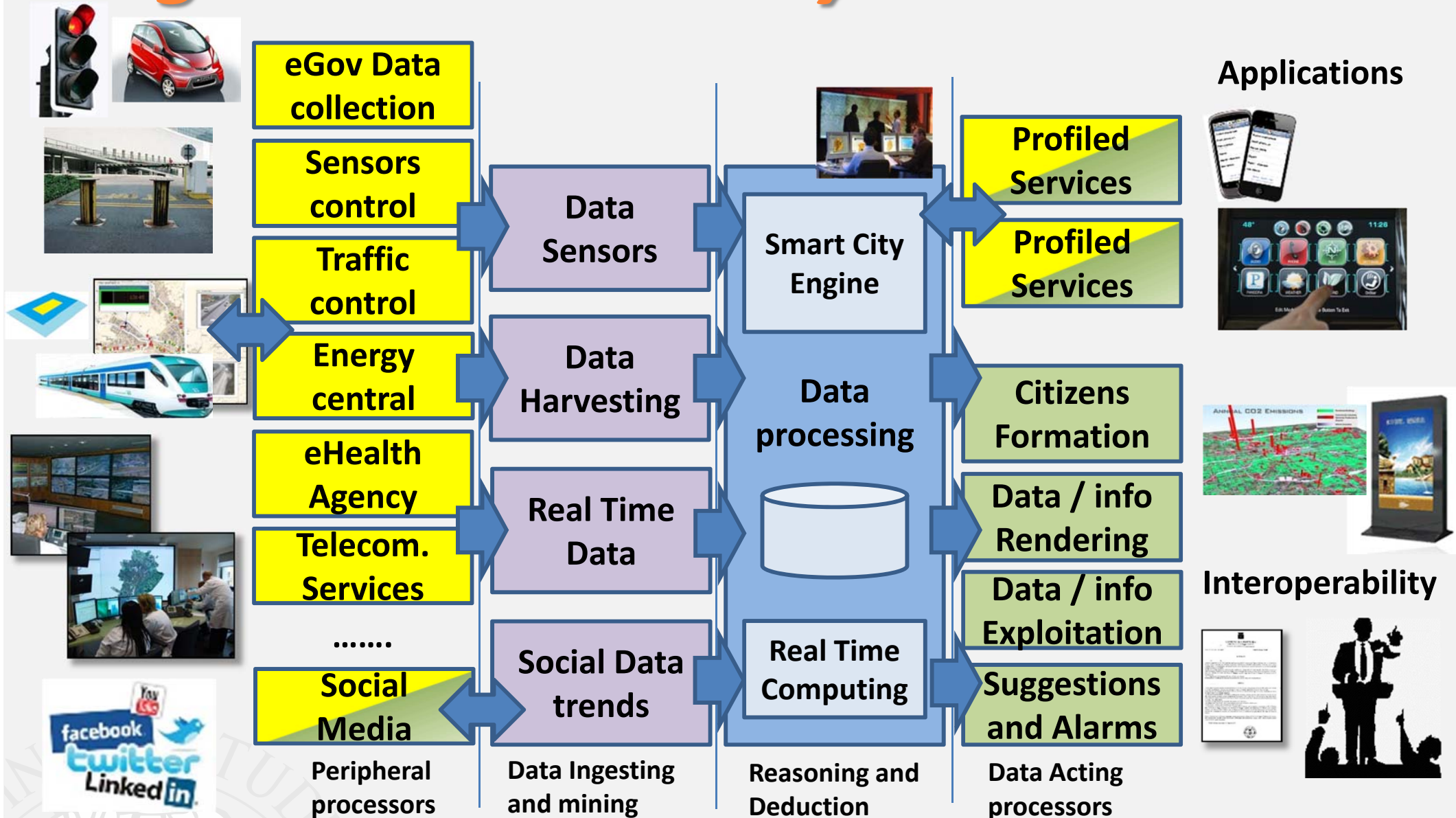


Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility**



Big Data Smart City Architecture



Big Data Smart-City service

- **Main Aim**

- Provide a platform able to ingest and take advantage a large number of data:
 - *Exploit data integration and reasoning*
 - *Deliver new services and applications to citizens*

- **Problems & Challenges**

- Data are provided in many different formats and protocols and from many different institutions, different convention and protocols, a different time, !
- Data are
 - not aligned (e.g., street names, dates, geolocations, tags, ...)
 - **not semantically interoperable**
- resulting a big data problem: volume, velocity, variability, variety,

Major topics addressed

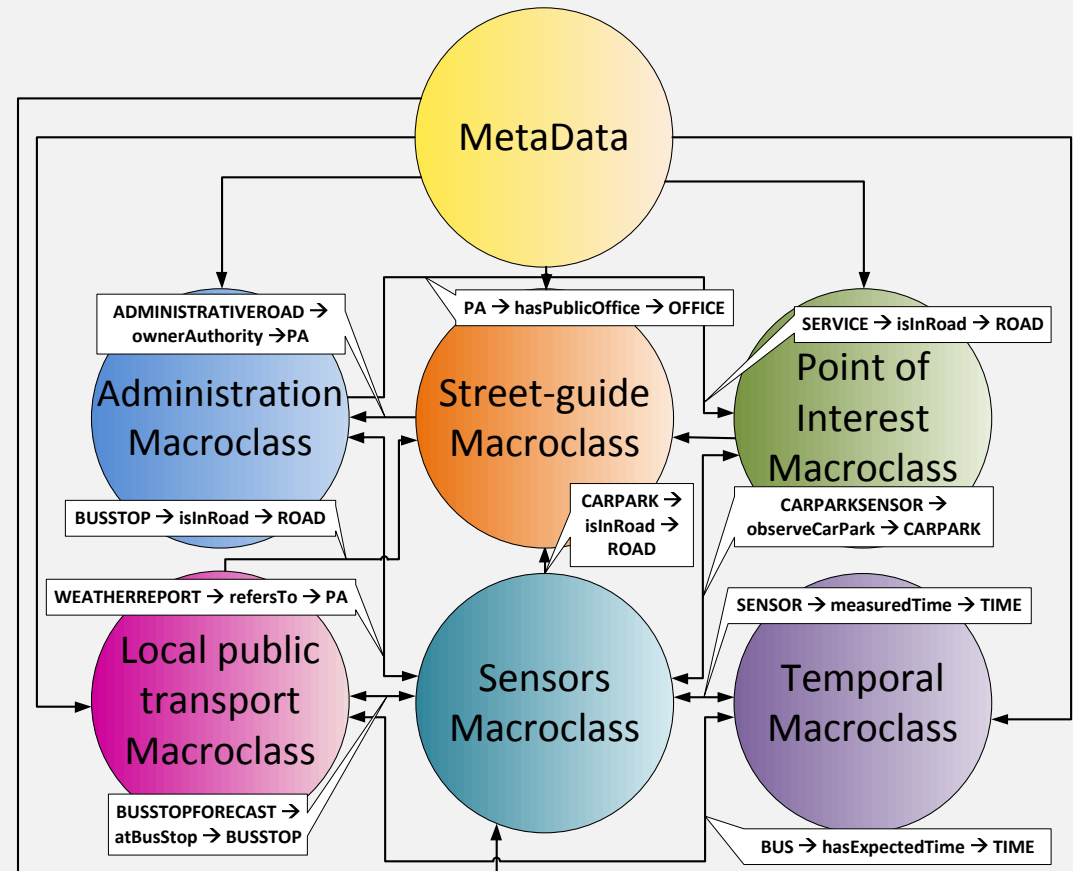
- **Big Data Smart City Architecture**
- **Smart-city Ontology** 
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility**

Smart-city Ontology objectives

- Create a unified knowledge base grounded on a common ontology that allows to combine all data coming from different sources making them semantically interoperable
- To.
 - Create coherent queries independently from the source, format, date, time, provider, etc.
 - Enrich the data, make it more complete, more reliable, more accessible
 - Enable to perform inference as triple materialization from some of the relations
 - to enable the implementation of new integrated services related to mobility
 - to provide repository access to SMEs to create new services

Smart-city Ontology

- The data model provided have been mapped into the ontology, it covers different aspects:
 - Administration
 - Street-guide
 - Points of interest
 - Local public transport
 - Sensors
 - Temporal aspects
 - Metadata on the data



Smart-city Ontology

- **Administration:** structure of the general public administrations (*Municipality*, *Province* and *Region*) also includes *Resolutions* (ordinance issued by administrations, may change the viability, infrastructural works, schedule for RTZ, etc.)
- **Street-guide:** formed by entities as *Road*, *Node*, *RoadElement*, *AdministrativeRoad*, *Milestone*, *StreetNumber*, *RoadLink*, *Junction*, *Entry*, *EntryRule*, *Maneuver*,... represents the entire road system of the region, including the permitted maneuvers and the rules of access to the limited traffic zones. Based on OTN (Ontology of Transportation Networks) vocabulary
- **Points of Interest:** includes all *Services*, activities, which may be useful to the citizen and who may have the need to search for and to arrive at, commercials, public administration, Cultural,

Smart-city Ontology

- **Local public transport:** includes the data related to major local public transport companies as **scheduled times**, the **rail graph**, and data relating to **real time passage at bus stops**, **real time position**, ...
- **Sensors:** data provided by sensors: currently, data are collected from various sensors (**parking status**, **meteo**, **pollution**) installed along some streets of Florence and surrounding areas, and from sensors installed into the main car parks of the region.
 - Plus: car sharing, bike sharing, AVM, RTZ, etc.
- **Temporal:** that puts **concepts related with time** (time intervals and instants) into the ontology, so that associate a timeline to the events recorded and is possible to make forecasts. It uses time ontologies such as OWL-Time.

Smart-city Ontology

- **Metadata:** modeling the additional information associated with:
 - **Descriptor** of Data sets that produced the triples: data set ID, title, description, purpose, location, administration, version, responsible, etc..
 - **Licensing** information
 - **Process** information: IDs of the processes adopted for ingestion, quality improvement, mapping, indexing,.. ; date and time of ingestion, update, review, ...;

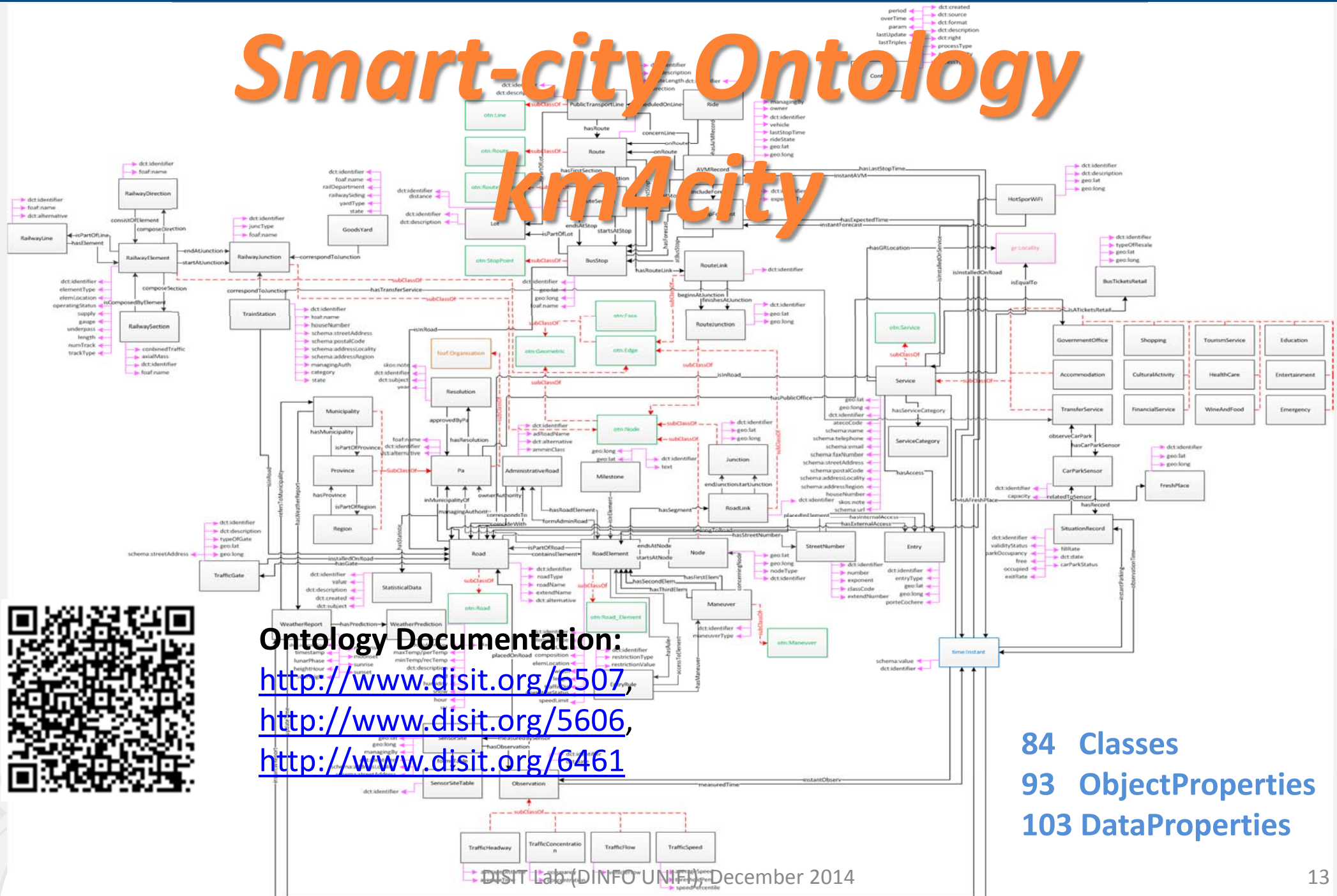
When a problem is detected, we have the information to understand when and how the problem has been included
- **Including basic ontologies as:**
 - *DC: Dublin core, standard metadata*
 - *OTN: Ontology for Transport Network*
 - *FOAF: for the description of the relations among people or groups*
 - *vCard: for a description of people and organizations*
 - *wgs84_pos: for latitude and longitude, GPS info*
 - *OWL-Time: reasoning on time, time intervals*
 - *GoodRelations: commercial activities models*

P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Building vs Data Harvesting and Cleaning for Smart-city Services", *International Journal of Visual Language and Computing*, Elsevier,

<http://dx.doi.org/10.1016/j.jvlc.2014.10.023>

Smart-city Ontology

km4city



Ontology Documentation:


<http://www.disit.org/6507>,

<http://www.disit.org/5606>,

<http://www.disit.org/6461>

84 Classes
93 ObjectProperties
103 DataProperties

Major topics addressed

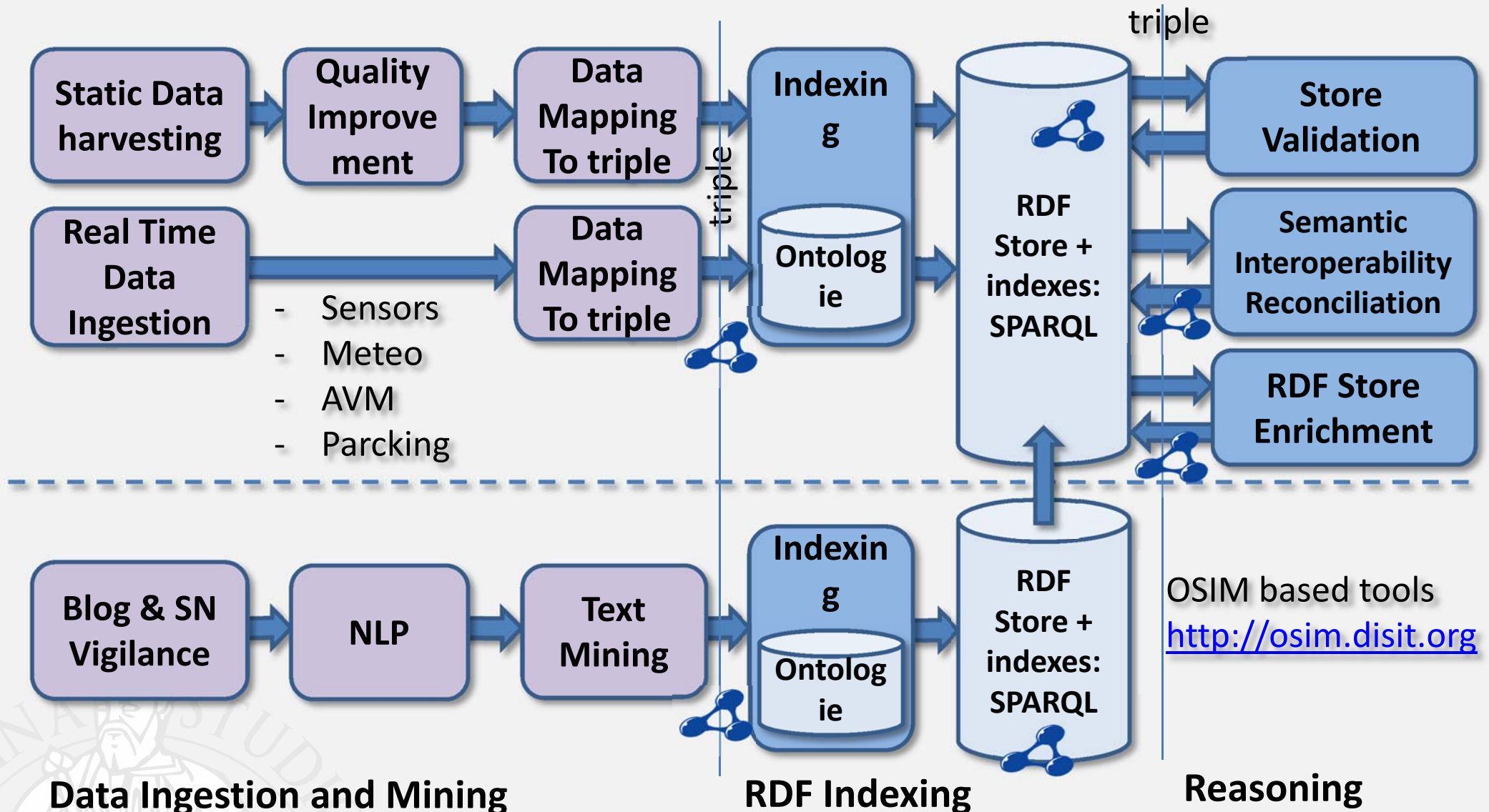
- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining** 
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility**

Data Ingestion and Mining

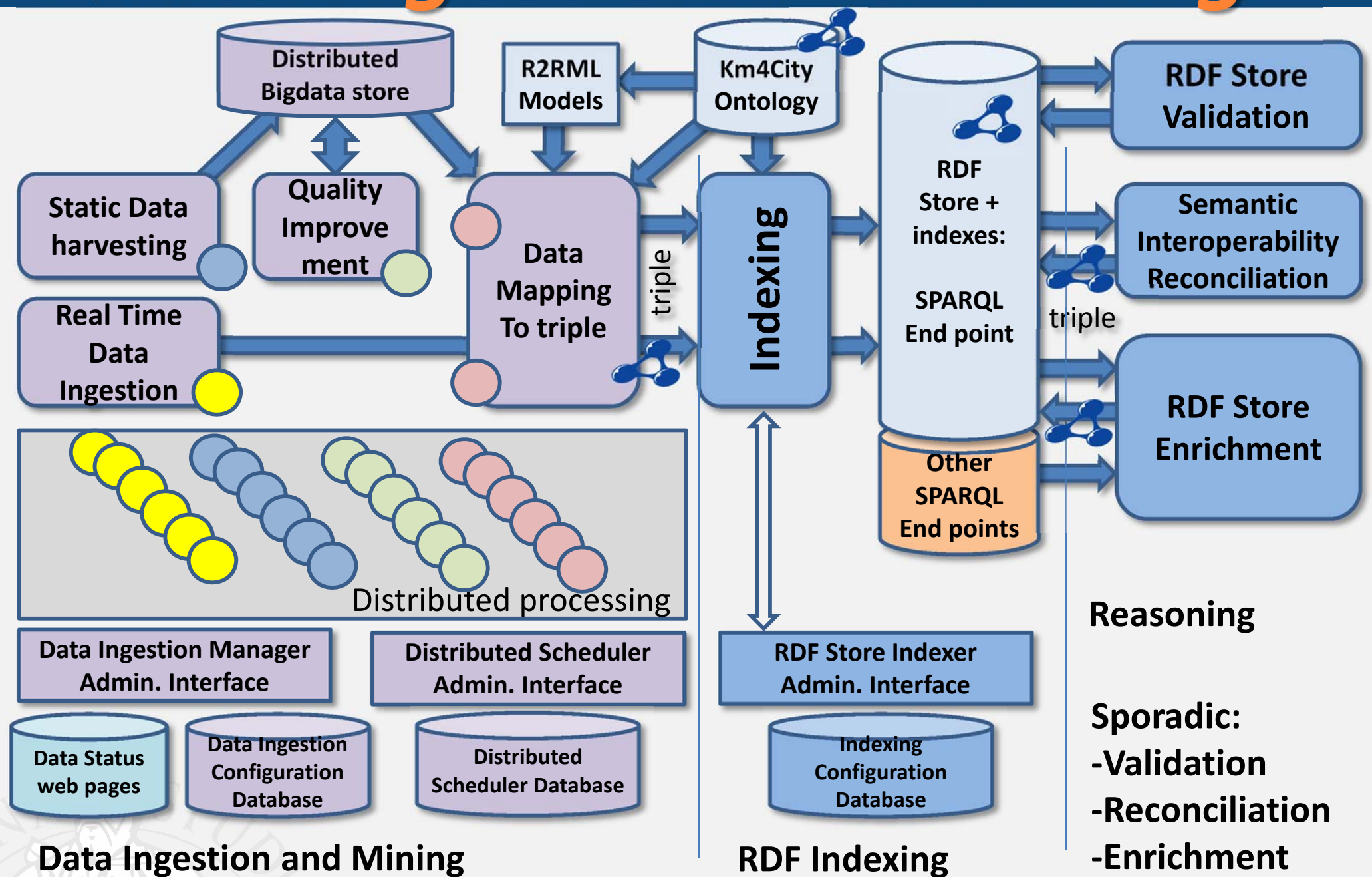
- **Data Ingestion Manager**
- **DataSets already integrated**
- **Static Data: harvesting**
- **Data Quality Improvement**
- **Data mapping to Triples**



Data Ingestion and Mining



Data Ingestion and Mining



Data Ingestion Manager

[illegible]

<http://192.168.0.100/processManager/>

Data Ingestion Manager

Show 10 entries
Schedule
Delete
Pause
Resume
Add row
Remove row

General informations												Ingestion
launch	Concatenate	process	Resource	Resource Class	Category	Format	Automaticity	Process_type	Access	Real Time	Source	I path
<input checked="" type="checkbox"/> I <input checked="" type="checkbox"/> Q <input checked="" type="checkbox"/> T <input checked="" type="checkbox"/> V <input checked="" type="checkbox"/> R	<input type="checkbox"/> <input type="checkbox"/> Q <input type="checkbox"/> T <input type="checkbox"/> V <input type="checkbox"/> R <input type="checkbox"/> Concatenate	Welfare_csv	Welfare	Servizi_csv1	Servizi	csv	automatic	ETL	HTTP	no	Opendata regione Toscana - http://dati.toscana.it	sh /home/ubuntu/programs/data-integration/kitchen.sh -file=/home/ubuntu/SiiMobility/Trasformazioni/TrasformazioneServizi/Main.kjb -level=Basic -param:processName=Welfare__csv
<input checked="" type="checkbox"/> I <input checked="" type="checkbox"/> Q <input checked="" type="checkbox"/> T <input checked="" type="checkbox"/> V <input checked="" type="checkbox"/> R	<input type="checkbox"/> <input type="checkbox"/> Q <input type="checkbox"/> T <input type="checkbox"/> V <input type="checkbox"/> R <input type="checkbox"/> Concatenate	Visite_guidate_csv	Visite guidate	Servizi_csv1	Servizi	csv	automatic	ETL	HTTP	no	Opendata regione Toscana - http://dati.toscana.it	sh /home/ubuntu/programs/data-integration/kitchen.sh -file=/home/ubuntu/SiiMobility/Trasformazioni/TrasformazioneServizi/Main.kjb -level=Basic -param:processName=Visite_guidate_csv
<input checked="" type="checkbox"/> I <input checked="" type="checkbox"/> Q <input checked="" type="checkbox"/> T <input checked="" type="checkbox"/> V <input checked="" type="checkbox"/> R	<input type="checkbox"/> <input type="checkbox"/> Q <input type="checkbox"/> T <input type="checkbox"/> V <input type="checkbox"/> R <input type="checkbox"/> Concatenate	Universita_e_conservatori_csv	Universita e conservatori	Servizi_csv1	Servizi	csv	automatic	ETL	HTTP	no	Opendata regione Toscana - http://dati.toscana.it	sh /home/ubuntu/programs/data-integration/kitchen.sh -file=/home/ubuntu/SiiMobility/Trasformazioni/TrasformazioneServizi/Main.kjb -level=Basic -param:processName=Universita_e_conservatori_csv

DataSet already integrated

- From MIIC web services (real time)
 - Parking payloadPublication (updated every h)
 - Traffic sensors payloadPublication (updated every 5-10min)
 - AVM client pull service (updated every 24h)
 - Street Graph
- From Municipality of Florence:
 - Tram lines: KMZ file that represents the path of tram in Florence
 - 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
 - Municipality of Florence resolutions
- From Tuscany Region:
 - 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
 - Weather forecast of the consortium Lamma (updated twice a day)

Data Ingestion Manager

- Numero di OD in tabella: 1169
- Numero di processi di OD trattati:
 - 38 diverse elaborazioni: static e real time
- Numero di processi schedulati: 353
 - solo real time: 286 previsioni, parcheggi, e sensori, AVM 2 line (un processo per linea)
- Sorgenti in Ingestion:
 - Comune,
 - Regione Toscana MIIC
- Altre possibili sorgenti in futuro:
 - car sharing, taxi, rifiuti, etc..

Present addressed Data Sets

- Tabella con tutti gli OD trattati (38 dataset)

Nome	formato	Nome2	Formato2
Arte e cultura	csv	Strutture ricettive	csv
Banche	csv	Strutture ricettive georeferenziate	csv
Corrieri espresso	csv	tempo libero	csv
Emergenze	csv	uffici vati	csv
Enogastronomia	csv	ubiversita' e conservatori	csv
Farmacie	csv	visite guidate	csv
Imprese del Commercio	csv	welfare	csv
Infrastrutture aeree	csv	Accessi sportello suolo pubblico e taxi	csv
Scuole dell'infanzia	csv	Delibere	csv
Scuola elementare	csv	Arrivi turistici	csv
Scuola media	csv	Ataf	csv
Scuola superiore	csv	Linee Tram	kmz
Corsi di Lingue e di formazione	csv	Sinistri per via	csv
Sport	csv	Veicoli circolanti	csv
Previsioni meteo	csv	Vetrina toscana - botteghe	csv
Salute e sanita'	csv	Strutture ricettive nuovo dataset	csv
Servizi epr il trasporto su strada	csv	musei	xml
Servizi vari	csv	POI dell'osservatorio dei trasporti	csv
Luoghi Freschi a Firenze	kmz	gate ZTL	kmz

Private Data from MIIC

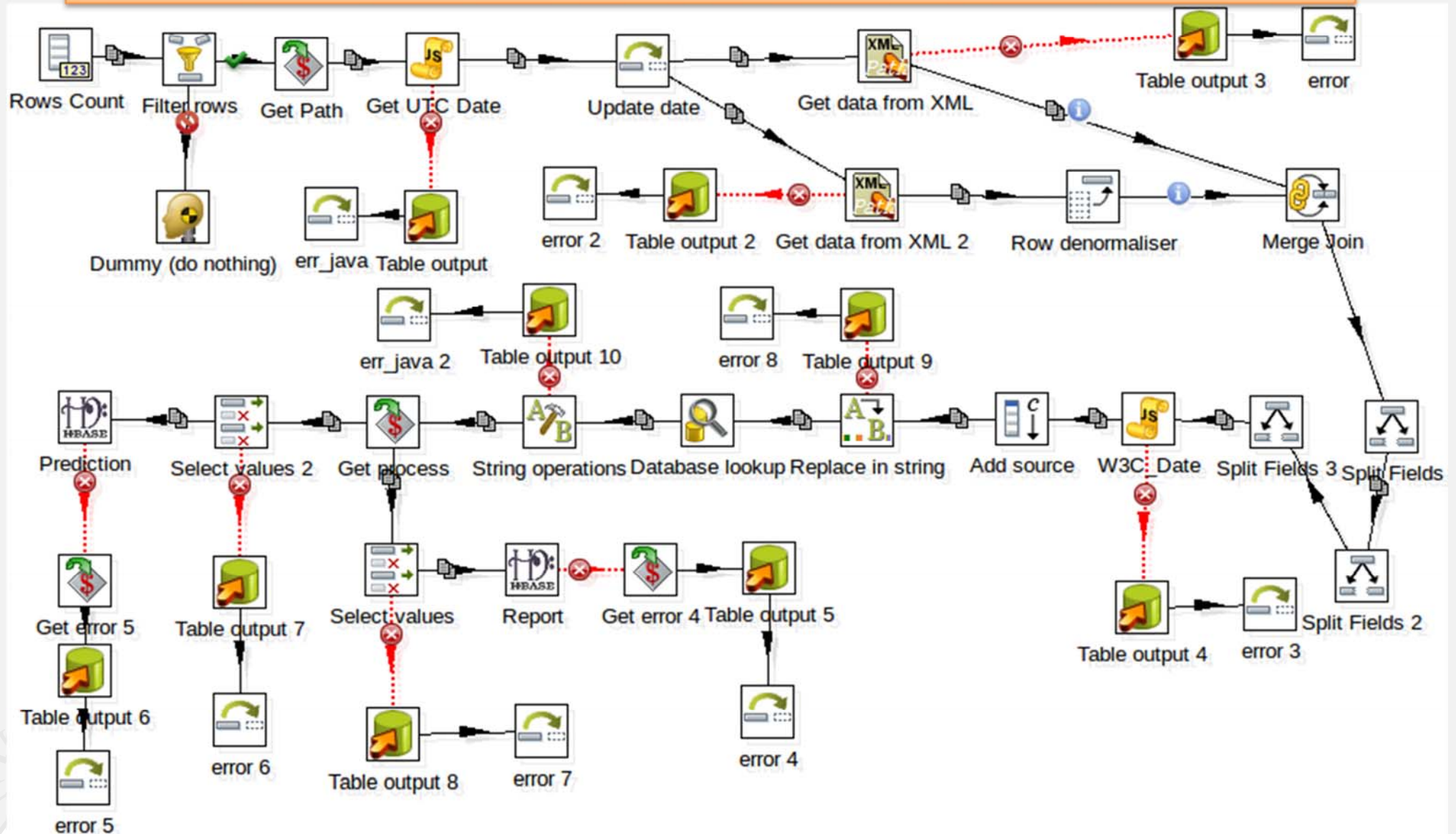
- Altri dati non Open (16 dataset)

Name	format	Name	Format
Grafo della provincia di Arezzo	dbf,shp,kml	Grafo della provincia di Massa e Carrara	dbf,shp,kml
Grafo della provincia di Firenze	dbf,shp,kml	Grafo della provincia di Arezzo	dbf,shp,kml
Grafo della provincia di Livorno	dbf,shp,kml	Grafo ferroviario della Toscana	dbf,shp,kml
Grafo della provincia di Pisa	dbf,shp,kml	RealTime data Parcheggi	xml
Grafo della provincia di Pistoia	dbf,shp,kml	RealTime data Sensori stradali	xml
Grafo della provincia di Prato	dbf,shp,kml	RealTime data AVM (solo 5 linee)	xml
Grafo della provincia di Grosseto	dbf,shp,kml	Fermate autobus di tutta la toscana	csv
Grafo della provincia di Siena	dbf,shp,kml	Intera rete Ataf (linee, percorsi, fermate)	txt

Static Data: harvesting

- **Ingesting a wide range of OD/PD:** public and private data, static, quasi static and/or dynamic real time data.
- For the case of Florence, we are addressing about **150 different data sources** of the 564 available, plus the regional, province, other municipalities,
- Using **Pentaho - Kettle** for data integration (Open source tool)
 - using specific ETL Kettle transformation processes (one or more for each data source)
 - data are stored in HBase (Bigdata NoSQL database)
- **Static and semi-static data** include: points of interests, geo-referenced services, maps, accidents statistics, etc.
 - files in several formats (SHP, KML, CVS, ZIP, XML, etc.)
- **Dynamic data** mainly data coming from sensors
 - parking, weather conditions, pollution measures, bus position, etc.
 - using Web Services.

Example of Ingestion process



Static Data: harvesting

- OD processati al momento: 38
- Altri dati gestiti in tabella ma non attivati in questa versione:

Nome	formato	Nome2	Formato2
Cento Luoghi – Assemblee 2012	kmz	Uffici Informazioni Turistiche	kmz
Distributori di carburante	kmz	Firenze Card	kmz
Consolati	kmz	Taxi	kmz
Percorsi piste ciclabili	kmz	Colonnine per la ricarica dei veicoli elettrici	kmz
Punto noleggio bici	kmz	Rastrelliere	kmz
Wifi	kmz	Cimiteri	kmz
Ospedali	kmz	Sedi elettorali	kmz

Data Quality Improvement

- **Problems kinds:**
 - Inconsistencies, incompleteness, duplications and redundancy, ..
- **Problems on:**
 - CAPs vs Locations
 - Street names (e.g., dividing names from numbers and localities, normalize when possible)
 - Dates and Time: normalizing
 - Telephone numbers: normalizing
 - Web links and emails: normalizing
- **Partial Usage of**
 - Certified and accepted tables and additional knowledge

Data Quality Improvement

Data quality's aspect:

- **Completeness:** presence of all information needed to describe an object, entity or event (e.g. Identifying).
- **Consistency:** data must not be contradictory. For example, the total balance and movements.
- **Accuracy:** data must be correct, i.e. conform to actual values. For example, an email address must not only be well-formed nome@dominio.it, but it must also be valid and working.

Data Quality Improvement

- **Reputation needed to avoid duplication:** in the case of multiple conflicting information for the same entity, the reputation of the sources would guide
- **Absence of duplication:** tables, records, fields should be stored only once, avoiding the presence of copies. Duplicate information involve double handling and can lead to problems of synchronization (consistency).
- **Integrity** is a concept related to relational databases, where there are tools to implement integrity constraints. Example a control on the types of data (contained in a column), or on combinations of identifiers (to prevent the presence of two equal rows).

Data Quality Improvement

Class	%QI	Total rows	Class	%QI	Total rows
Accoglienza	34,627	13256	Georeferenziati	38,754	2016
Agenzie delle Entrate	27,124	306	Materne	41,479	539
Arte e Cultura	37,716	3212	Medie	42,611	116
Visite Guidate	38,471	114	Mobilità Aerea	41,872	29
Commercio	42,105	323	Mobilita' Auto	38,338	196
Banche	41,427	1768	Prefetture	39,103	449
Corrieri	42,857	51	Sanità	42,350	1127
Elementari	42,004	335	Farmacie	42,676	2131
Emergenze	42,110	688	Università	42,857	43
Enogastronomia	42,078	5980	Sport	52,256	1184
Formazione	42,857	70	Superiori	42,467	183
Accoglienza	34,627	13256	Tempo Libero	25,659	564

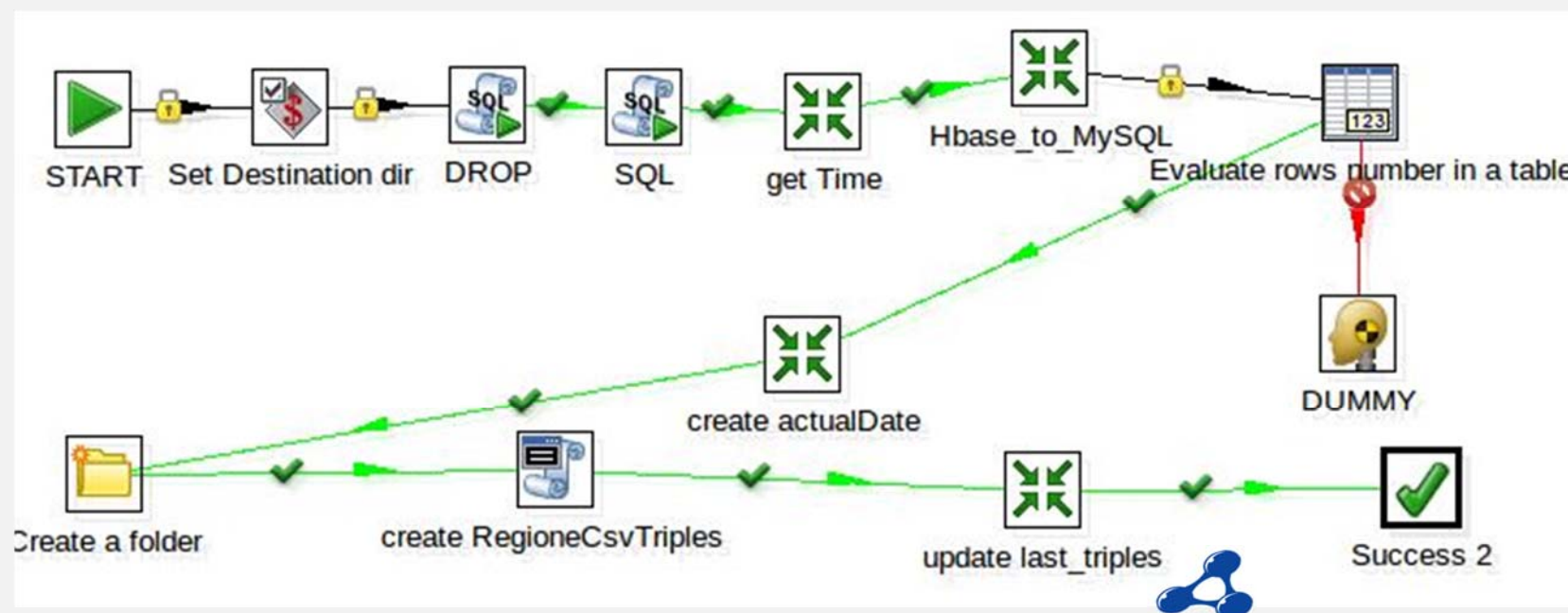
Service data from Tuscany region.

%QI = improved service data percentage after QI phase.

Data mapping to Triples



- Transforms the data from HBase to RDF triples
- Using **Karma Data Integration tool**, a mapping model from SQL to RDF on the basis of the ontology was created
 - Data to be mapped first temporarily passed from Hbase to MySQL and then mapped using Karma (in batch mode)
- The mapped data in triples have to be uploaded (and indexed) to the **RDF Store** (OpenRDF – sesame with OWLIM-SE)

RDF Triples generation



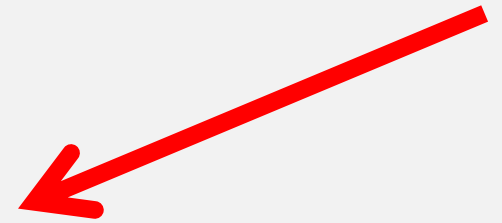
The triples are generated with the **km4city** ontology and then loaded on OWLIM RDF store.

RDF Triples generated from static data

Macro Class	 Static Triples	 Reconciliation Triples
Administration	2.431	0
Metadata of DataSets	416	0
Point of Interest (35.273 POIs in Tuscany)	471.657	34.392
Street-guide (in Tuscany)	68.985.026	0
Local Public Transport (<5 lines of FI)	644.405	2.385
Total	70.103.935	

Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility**



Distributed and real time processes

- **Distributed Scheduler**
- **Real Time Data Ingestion**
- **Blog Vigilance, NLP, Text Mining**
- **Parallel and distributed processing**



Distributed Scheduler

- Use of a scheduler to manage periodic execution of ingestion and triple generation processes.
 - This tool throws the processes with predefined interval determined in phase of configuration.
- **Static Data: as Sporadic processes:**
 - scheduled every months or week
- **Real Time data** (car parks, road sensors, etc.)
 - ingestion and triple generation processes should be performed periodically (no for **static data**).


<http://192.168.0.72>

Real Time Data Ingestion


























- Sensors on Traffic flow data: Florence, Empoli, Piombino, Arezzo
 - number of catalogs: 63
- Parkings status: Florence (14), Empoli (6), Arezzo (9), Grosseto (4), Livorno (5), Lucca (5), Massa-Carrara (3), Prato (1)
 - Total number: 48
- AVM busses 5 lines, in Florence:
 - 816 rides for Line 4, 1210 rides for Line 6
- Weather conditions and forecast for
 - Municipalities in Tuscany: about 285 cities
- **Attualmente sul grafo sono riconciliati solo i parcheggi di Firenze, Empoli, ..., ma tutti mandano dati.**

Example for Parkings

- To process the parking data for Sii-Mobility
 - real time data from Osservatorio Trasporti of Tuscany region (MIIC).



The screenshot shows the SIM Consultazione web interface. On the left, there is a navigation menu under 'Classi Dati' with options for 'Classi dati', 'TPL', 'Infrastrutture di trasporto', and 'Tempo reale'. The main content area is titled 'Consultazione Tempo reale' and contains a table with the following data:

Classe dati	Metadati	Interfaccia Input	Interfaccia Output	Validità	Consultazione	Mappe
Sensori				-		
Parcheggi				-		
Emergenze				-		
Rilievi AVM				-		
Meteo				-		

- 2 phases:
 - **INGESTION** phase;
 - **TRIPLES GENERATION** phase.

Example for AVM data


































- AVM data area obtained for CodeRace:
 - A CodeRace identifies a Bus race on a given line
- At the same time, on a single Line:
 - A number of Busses (with their code race) are running.
 - For each of them a forecast for each bus-stop is provided.
 - At every new forecast, all obsolete forecasts for next and passed predictions have to be deprecated,
- → thus filtering is needed to avoid growing of not useful data into the RDF Store




Smart Cloud Engine

DISIT - Distributed Systems and Internet Technology Lab

<http://192.168.0.72>

SCHEDULER NAME	ID	FIRE INSTANCE ID	DATE	JOB NAME	JOB GROUP	JOB DATA	STATUS
 SCE	297230	hadoopnode01d14183 077042351418307705 019	2014-12-15 15:25:33	sensori47_A	sensori47	#processParameter s= [{"processPath":"/ho mologhuntu/program	  RUNNING
 SCE	297229	hadoopnode06c14183 076279641418307629 359	2014-12-15 15:25:33	sensori44_A	sensori44	#processParameter s= [{"processPath":"/ho mologhuntu/program	  RUNNING
 SCE	297228	hadoopnode02141830 838738214183083917 58	2014-12-15 15:22:39	ZTL_notturna_shp_I	ZTL_notturna_shp	#processParameter s=null; #isNonConcurrent=t ruc	  SUCCESS
 SCE	297227	hadoopnode02141830 838738214183083917 57	2014-12-15 15:22:39	ZTL_notturna_kmz_I	ZTL_notturna_kmz	#processParameter s=null; #isNonConcurrent=t ruc	  SUCCESS
 SCE	297226	hadoopnode01c14183 085186101418308520 365	2014-12-15 15:21:49	sensori45_A	sensori45	#processParameter s= [{"processPath":"/ho mologhuntu/program	  RUNNING
 SCE	297225	hadoopnode06141830 832370214183083258 68	2014-12-15 15:21:49	sensori40_A	sensori40	#processParameter s= [{"processPath":"/ho mologhuntu/program	  RUNNING
 SCE	297224	hadoopnode01b14183 075646221418307566 749	2014-12-15 15:21:49	sensori46_A	sensori46	#processParameter s= [{"processPath":"/ho mologhuntu/program	  RUNNING
 SCE	297223	hadoopnode02141830 838738214183083917 56	2014-12-15 15:21:37	ZTL_notturna_kmz_I	ZTL_notturna_kmz	#processParameter s=null; #isNonConcurrent=t ruc	  SUCCESS
 SCE	297222	hadoopnode02141830 838738214183083917 55	2014-12-15 15:21:00	sensori31_C	sensori31	#processParameter s= [{"processPath":"/ho mologhuntu/program	  SUCCESS
 SCE	297221	hadoopnode06c14183 076279641418307629 358	2014-12-15 15:21:00	sensori30_C	sensori30	#processParameter s= [{"processPath":"/ho mologhuntu/program	  SUCCESS
 SCE	297220	hadoopnode02141830 838738214183083917 54	2014-12-15 15:18:58	ZTL_notturna_shp_I	ZTL_notturna_shp	#processParameter s=null; #isNonConcurrent=t ruc	  SUCCESS
 SCE		hadoopnode01c14183				#processParameter	

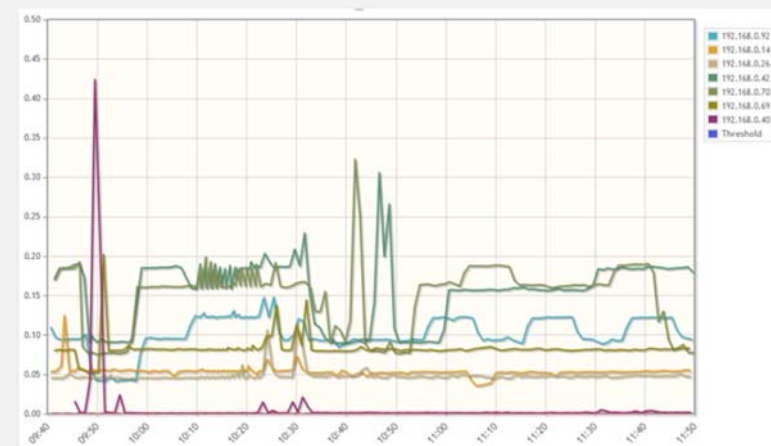
Distributed Scheduler


Smart Cloud Engine
 DISIT - Distributed Systems and Internet Technology Lab

192.168.0.14	192.168.0.26	192.168.0.40	192.168.0.42	192.168.0.69	192.168.0.70	192.168.0.92
<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:04 • SCHEDULER_INSTANCE_ID hadoopnode01d1418718662488 • CPU_LOAD 0.05322341999577256 • FREE_PHYSICAL_MEMORY 4686650584 • JOBS_EXECUTED 0 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 0 • RUNNING_SINCE 2014-12-16 09:31:02 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 8.877615726062143E-4 • SYSTEM_LOAD_AVERAGE 0.0 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3679342592 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12860071936 • PROCESS_CPU_TIME 39770000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 2321541775360 • UNALLOCATED_DISK_SPACE 1937102204928 • USABLE_DISK_SPACE 1819765923840 • PREV_FIRE_TIME 2014-12-15 23:09:17 • CPU Intel(R) Xeon(R) CPU X3470 @ 2.93GHz 	<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:04 • SCHEDULER_INSTANCE_ID hadoopnode06c1418719522697 • CPU_LOAD 0.04810651796803606 • FREE_PHYSICAL_MEMORY 10056519680 • JOBS_EXECUTED 0 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 0 • RUNNING_SINCE 2014-12-16 09:32:03 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 8.425309630128908E-4 • SYSTEM_LOAD_AVERAGE 0.13 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3679342592 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12633550848 • PROCESS_CPU_TIME 39770000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 2321541775360 • UNALLOCATED_DISK_SPACE 1938265976832 • USABLE_DISK_SPACE 1820919695744 • PREV_FIRE_TIME 2014-12-15 23:14:19 • CPU Intel(R) Xeon(R) CPU E5-4620 @ 2.20GHz 	<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:11 • SCHEDULER_INSTANCE_ID hadoopnode06d1418719522697 • CPU_LOAD 0.0013337223356812403 • FREE_PHYSICAL_MEMORY 10849054720 • JOBS_EXECUTED 20 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 14.91 • RUNNING_SINCE 2014-12-16 09:45:22 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 5.001458758804651E-4 • SYSTEM_LOAD_AVERAGE 0.0 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3687526400 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12881752064 • PROCESS_CPU_TIME 29620000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 21252098688 • UNALLOCATED_DISK_SPACE 195266711552 • USABLE_DISK_SPACE 185156763648 • PREV_FIRE_TIME 2014-12-16 09:53:47 • CPU Intel(R) Xeon(R) CPU X5690 @ 3.47GHz 	<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:35 • SCHEDULER_INSTANCE_ID hadoopnode06f1418718994664 • CPU_LOAD 0.16369819341126463 • FREE_PHYSICAL_MEMORY 1921798144 • JOBS_EXECUTED 0 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 0 • RUNNING_SINCE 2014-12-16 09:36:34 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 7.651759904778099E-4 • SYSTEM_LOAD_AVERAGE 1.04 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3679342592 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12159328256 • PROCESS_CPU_TIME 29620000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 2321541775360 • UNALLOCATED_DISK_SPACE 1937021210624 • USABLE_DISK_SPACE 1819684929536 • PREV_FIRE_TIME 2014-12-15 23:09:17 • CPU Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz 	<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:56 • SCHEDULER_INSTANCE_ID hadoopnode021418718882292 • CPU_LOAD 0.081939516810272 • FREE_PHYSICAL_MEMORY 5102755840 • JOBS_EXECUTED 0 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 0 • RUNNING_SINCE 2014-12-16 09:33:55 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 8.870865543023698E-4 • SYSTEM_LOAD_AVERAGE 0.6 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3683553280 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12881752064 • PROCESS_CPU_TIME 29770000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 2321541775360 • UNALLOCATED_DISK_SPACE 1937341145088 • USABLE_DISK_SPACE 1820004864000 • PREV_FIRE_TIME 2014-12-15 23:09:15 • CPU Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz 	<ul style="list-style-type: none"> • LAST_CHECK 2014-12-16 11:29:43 • SCHEDULER_INSTANCE_ID hadoopnode01c1418718921761 • CPU_LOAD 0.09430552637108637 • FREE_PHYSICAL_MEMORY 7336054784 • JOBS_EXECUTED 0 • SCHEDULER_NAME SCE • CURRENT_TIME 2014-12-16 11:29:58 • JOBS/h 0 • RUNNING_SINCE 2014-12-16 09:34:42 • CLUSTERED 1 • PERSISTENCE 1 • REMOTE_SCHEDULER 0 • CURRENTLY_EXECUTING_JOBS 0 • CPU_LOAD_JVM 8.870865543023698E-4 • SYSTEM_LOAD_AVERAGE 0.89 • OPERATING_SYSTEM_VERSION 3.13.0.24-generic • COMMITTED_VIRTUAL_MEMORY 3679342592 • OPERATING_SYSTEM_NAME Linux • FREE_SWAP_SPACE 12508909568 • PROCESS_CPU_TIME 35000000000 • TOTAL_PHYSICAL_MEMORY 1.2600922112E10 • NUMBER_OF_PROCESSORS 4 • OPERATING_SYSTEM_ARCHITECTURE amd64 • TOTAL_SWAP_SPACE 1.2881752064E10 • IS_SCHEDULER_STANDBY 0 • IS_SCHEDULER_SHUTDOWN 0 • IS_SCHEDULER_STARTED 1 • TOTAL_DISK_SPACE 2321541775360 • UNALLOCATED_DISK_SPACE 193826713344 • USABLE_DISK_SPACE 1820910432256 • PREV_FIRE_TIME 2014-12-15 23:09:15 • CPU Intel(R) Xeon(R) CPU X3470 @ 2.93GHz 	

CPU: 18.01 GHz
CPU Load: 1.48 GHz (8.19%)
Mem Tot: 82.15 GB
Mem Free: 45.47 GB
Cores: 28
Jobs/h: 14.91

<http://192.168.0.72>

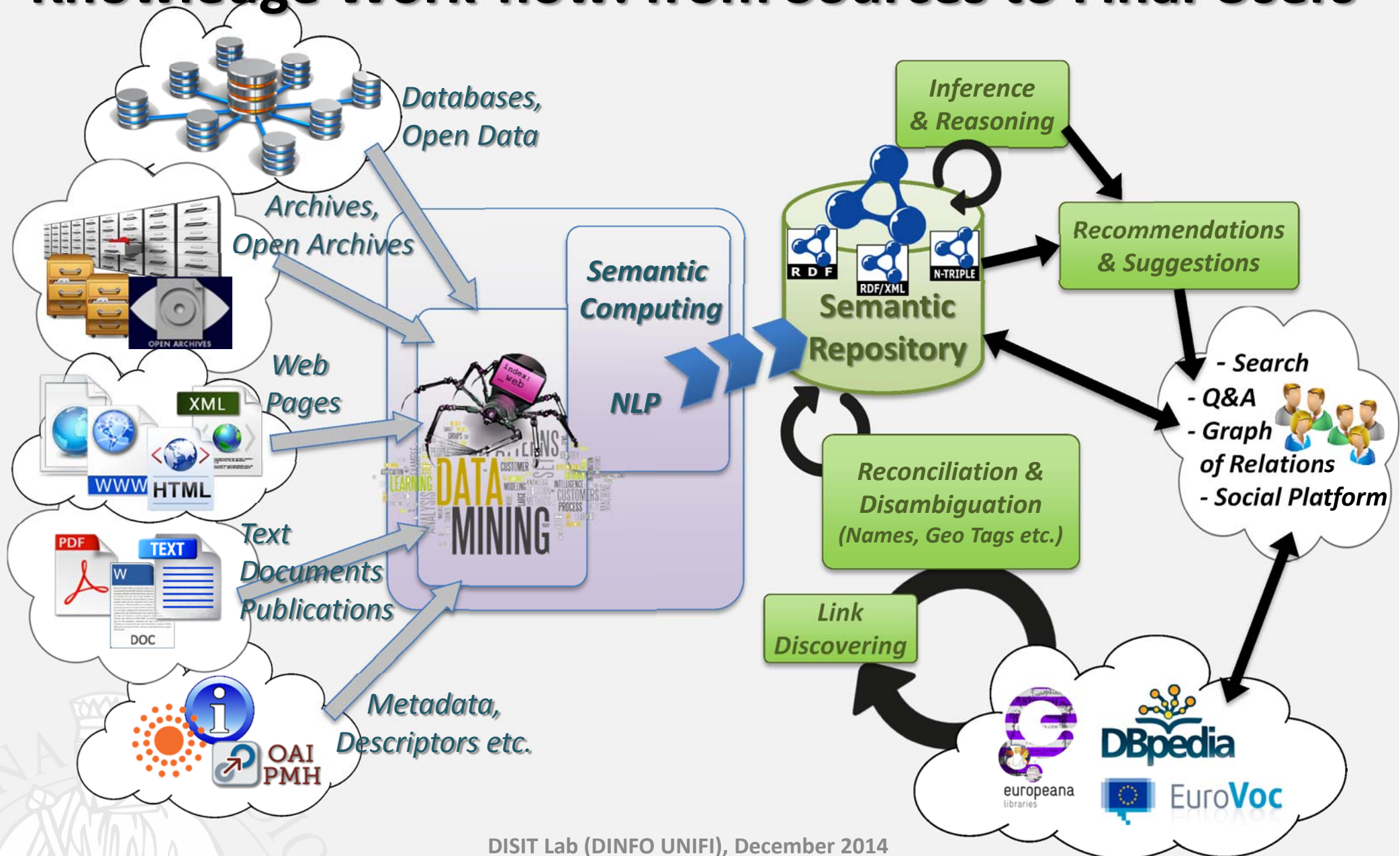


Macro Class	Static Triples	Reconciliation Triples	Real Time Triples Loaded	Total on 1.5 months
Administration	2.431	0	--	2.431
Metadata of DataSets	416	0	--	416
Point of Interest (35.273 POIs in Tuscany)	471.657	34.392	--	506.049
Street-guide (in Tuscany)	68.985.026	0	--	68.985.026
Local Public Transport (<5 lines of FI)	644.405	2.385	135.952 per line per day, to be filtered, read every 30 s, they respond in minutes	(static) 646.790
Sensors (<201 road sensors, 63 scheduled every two hours)	--	4.240	102 per sensor per read, every 2 hours, they are very slow in responding	51.111.078
Parking (<44 parkings, 12 scheduled every 30min)	--	1.240	7920 per park per day, 3 read per hour, they respond in seconds	
Meto (286 municipalities, all scheduled every 6 hours)	--	--	185 per location per update, 1-2 updates per day	
Temporal events, time stamp	--	--	6 for each event	1.715.105
Total	70.103.935	42.257		122.966.893

Blog Vigilance, NLP, Text Mining

- ***Monitoring the moods of the population vs city services.***
- Big Data Mining issues, Many different formats
- Ambiguities and inconsistencies of descriptors, metadata etc.
- Unstructured, decontextualized data does not allow to extract high level information
- Several different efforts of structure KBs, ontologies, taxonomies etc. in many fields of Knowledge
- There is necessity of:
 - Reconciliation and disambiguation of ingested data
 - Organize data into proper forms of structured knowledge
 - Standardize definitions, languages, vocabularies etc.
 - Link discovering among different knowledge bases
 - Make inference to produce additional knowledge
 - Detect unexpected correlations, produce suggestions and recommendations
 - Provide semantic interoperability among resources and applications.

Knowledge Work-flow: from Sources to Final Users




Parallel and distributed processing

- **A Cluster of 7** Ubuntu 64bit servers with Hbase for
 - Distributed Scheduler put in execution processes of Data Ingestion and Mining (Static and Real Time): DataIng, QualityImp., Triplification
- **A Cluster of 5** Ubuntu 64bit servers with Hadoop for
 - web crawling (Nutch)
 - RDF Enrichment via Natural Language Processing, GATE with DISIT Tools,
- **One Ubuntu Server (.72)** for
 - NFS, network file system
 - Hbase master del cluster Hbase
 - SCE per il controllo delle SLA (in sviluppo)
- **One Ubuntu server (.100)** for
 - Data Ingestion Manager
 - RDF store Index generator, RDF Indexer
 - performing Reconciliation via SILK processes
- **One Windows Server 2008 (.205)** for
 - RDF front end via SPARQL,
 - providing <http://ServiceMap.disit.org>
 - performing Validation processes
- **One Windows Server 2008** for: <http://LOG.DISIT.ORG>
- **A Total of 16** Virtual Machines on DISIT Cloud



Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing** 
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility**

RDF processing



- **RDF Store Indexing**
- **RDF Store Validation**
- **Semantic Interoperability, reconciliation**
- **RDF Store Enrichment, for link discovering**
- **RDF Store Enrichment, for service discovering via web crawling**



RDF Store Indexer, versioning

- RDF Store Indexer enables administrator to select - step by step - triples to be loaded from:
 - Ontologies: several of them as explained before
 - Triples produced from static data
 - Triples produced from real time data, taking into account historical windows of data (from-to)
 - Triples obtained by procedures of:
 - reconciliation
 - Enrichment
- Keeping under control incremental indexing for adding new static and historical data without re-indexing..
- An indexing process takes several hours or days
- The new indexing can be performed while a current index is in production with real time data...
- ***If you do not index, you cannot perform reconciliations, neither enrichments***

summary of an index generation

Choices summary

Selected ontologies

Name	URI Prefix	Path	Last file date	Today
doterms	http://purl.org/dc/terms/	/Ontologies/dcterms	2014-10-01 12:00:00	2014-10-01 12:00:00
dotype	http://purl.org/dc/dcmitype/	/Ontologies/dctype	2014-10-01 12:00:00	2014-10-01 12:00:00
foaf	http://xmlns.com/foaf/0.1/	/Ontologies/foaf	2014-10-01 12:00:00	2014-10-01 12:00:00
km4c	http://www.disit.org/km4city/schema#	/Ontologies/km4c	2014-12-04 10:00:00	2014-12-04 10:00:00

Showing 1 to 4 of 4 entries (filtered from 11 total entries)

First Previous 1 Next Last

Selected static data

Process Name	Resource	Description	Triples path	Last source update	Today
Arrivi_turistici_csv	Arrivi turistici	comune - Il dataset contiene il numero di arrivi alle strutture ricettive del comune di Firenze, per anno e nazionalità di origine, per anno. Periodo 2005-2010.	/Triples/Statistiche/Arrivi_turistici_csv	2014-10-30T10:59:00.592	2014-10-30 11:26:10
Arte_e_cultura_csv	Arte e cultura	regione - Musei, gallerie, monumenti, Teatri, biblioteche, ...	/Triples/Servizi/Arte_e_cultura_csv	Thu, 23 Jan 2014 15:50:08 GMT	2014-10-31 10:34:43
Ataf_csv	Ataf	comune - Il dataset contiene che il numero di biglietti venduti, abbonamenti venduti, numero di passeggeri e lunghezza della rete in km inerenti al servizio pubblico di trasporto ATAF, per anno. Periodo 2001-2010.	/Triples/Statistiche/Ataf_csv	2014-10-30T11:05:18.533	2014-10-30 11:25:11
Banche_csv	Banche	regione - Le filiali delle Banche nel Territorio regionale	/Triples/Servizi/Banche_csv	Thu, 23 Jan 2014 15:55:08 GMT	2014-10-31 12:00:45
Corrieri_espresso_csv	Corrieri espresso	regione - Uffici dei Corrieri espresso	/Triples/Servizi/Corrieri_espresso_csv	Thu, 23 Jan 2014 16:00:58 GMT	2014-10-31 12:08:03

Showing 1 to 5 of 5 entries (filtered from 57 total entries)

First Previous 1 Next Last

Selected real time data

Process Name	Resource	Description	Triples path	Last source update	Today
Previ_meteo_Abetone_xml	Previsioni meteo Abetone	regione - Previsioni Meteo generate ed aggiornate dal LaMMA 2 volte al giorno (mattina entro le 10; pomeriggio entro le 16)	/Triples/Previsioni/Previ_meteo_Abetone_xml	Mon, 15 Dec 2014 12:39:19 GMT	2014-12-01 0:00:00 - 2014-12-15 0:00:00
Previ_meteo_Aglana_xml	Previsioni meteo Aglana	regione - Previsioni Meteo generate ed aggiornate dal LaMMA 2 volte al giorno (mattina entro le 10; pomeriggio entro le 16)	/Triples/Previsioni/Previ_meteo_Aglana_xml	Mon, 15 Dec 2014 12:39:21 GMT	2014-12-01 0:00:00 - 2014-12-14 0:00:00
sensori1	sensori1		/Triples/Sensori/sensori1		2014-12-08 0:00:00 - 2014-12-14 0:00:00

Showing 1 to 3 of 3 entries (filtered from 453 total entries)

First Previous 1 Next Last

Selected reconciliations

Name	Macroclasses	Triples	Description	Path	Last file date	Today
No matching records found						
Name	Macroclasses	Triples	Description	Path	Last file date	Today

Showing 0 to 0 of 0 entries (filtered from 1 total entries)

Back Confirm

First Previous 1 Next Last

<http://192.168.0.100>

RDF Store Validation

- Some of the produced and addressed triples in indexing could not be loaded and indexed since the code can be wrong or for the presence of noise and process failure
- A set of queries applied to verify the consistency and completeness, after new re-indexing and new data integration
 - I.e.: the KB regression testing !!!!!
 - Success rate is presently of the 99,999%!
- Non loaded: 0,00000638314%

Semantic Interoperability, reconciliation

- After the loading and indexing into the RDF store a dataset may be connected with the others **if entities refer to the same triples**
 - **Missed connections** strongly limit the usage of the knowledge base,
 - e.g. the services are not connected with the road graph.
 - E.g., To associate each **Service** with a **Road** and an **Entity** on the basis of the street name, number and locality
- Data are coming from different sources, created in different dates, from different archives, by using different standards and codes.
- In the case of conflicting information the reputation of sources has to guide in the data fusion.

Semantic Interoperability, reconciliation

- **Examples:**

- Typos;
- Missing street number, or replaced with "0" or "SNC";
- Municipalities with no official name (e.g. Vicchio/Vicchio del Mugello);
- Street names and street numbers with strange characters (-, /, ° ? , Ang., ,);
- Road name with words in a different order (e.g. Via Petrarca Francesco, exchange of name and surname);
- Red street numbers (for shops);
- Presence/absence of proper names in road name (e.g. via Camillo Benso di Cavour / via Cavour);
- Number wrongly written (e.g. 34/AB, 403D, 36INT.1);
- Roman numerals in the road name (e.g., via XXVII Aprile).

- **Steps:**

1. *SPARQL Exact match* – match the strings as they are
2. *SPARQL Enhanced Exact Match* – make some substitutions (Via S. Marta → Via Santa Marta, ...)
3. *Last Word Search* – use only the last word of street name
4. Use Google GeoCoding API
5. Remove 'strange chars' (-, /, °, ? , Ang., ,) from Street number
6. Remove 'strange chars' from Street name
7. Rewrite wrong municipality names

Comparing different reconciliation approaches based on

- SILK link discovering language
- SPARQL based reconciliation described above

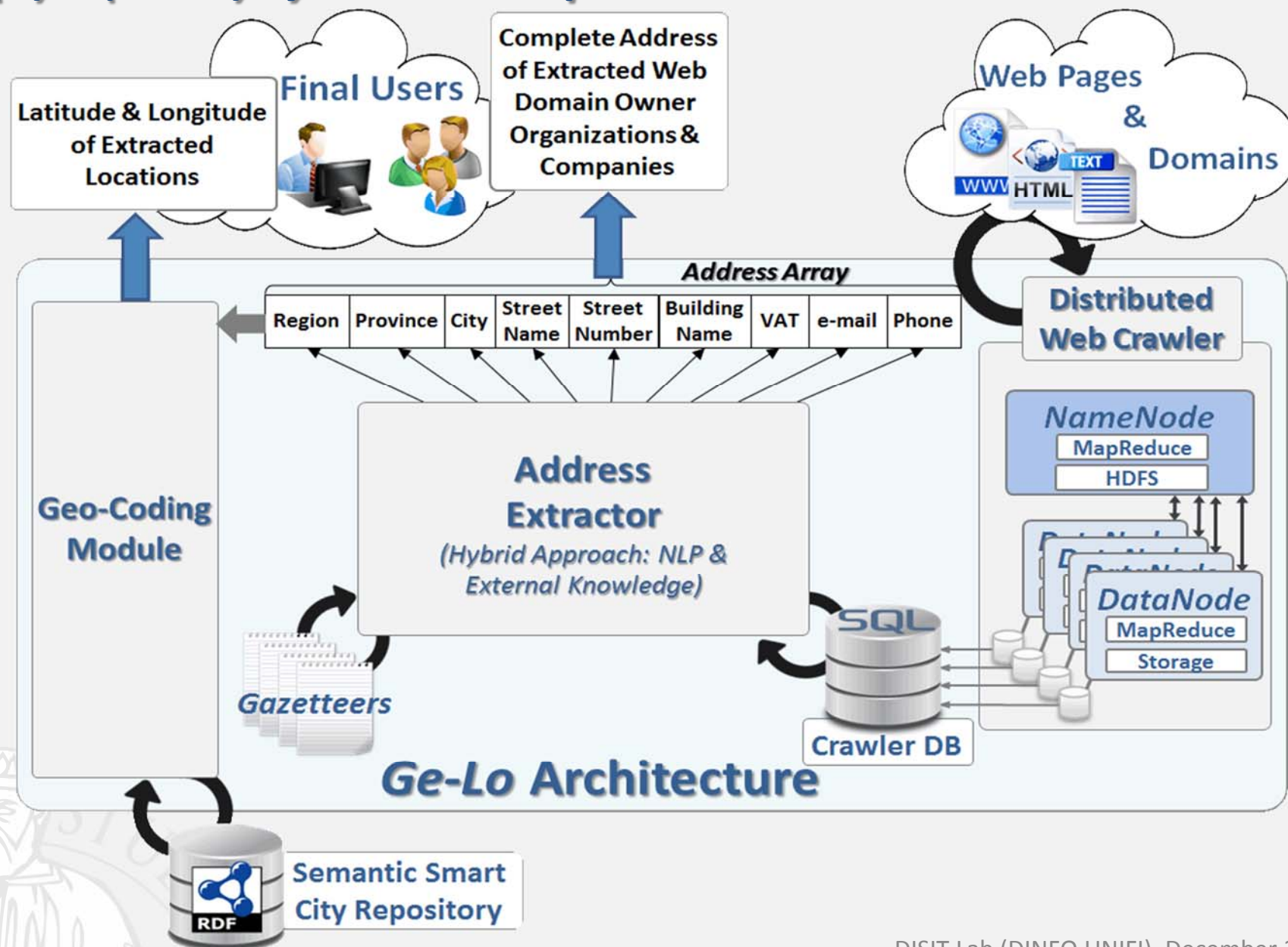
Method	Precision	Recall	F1
SPARQL –based reconciliation	1,00	0,69	0,820
SPARQL -based reconciliation + additional manual review	0,985	0,722	0,833
Link discovering - Leveisthein	0,927	0,508	0,656
Link discovering - Dice	0,968	0,674	0,794
Link discovering - Jaccard	1,000	0,472	0,642
Link discovering + heuristics based on data knowledge + Leveisthein	0,925	0,714	0,806

Thus automation of reconciliation is possible and produces acceptable results!!

Localization via web crawling

- Using the **Ge(o)Lo(cator)** framework:
 - Mining, retrieving and geolocating web-domains associated to companies in Tuscany (thanks to a **Distribute Web Crawler** based on Apache Nutch + Hadoop)
 - Extraction of geographical information based on a hybrid approach (thanks to Open Source **GATE** Framework + using external gazetteers)
 - Validation in 2 steps: Evaluation of Complete Address Array Extraction, Evaluation of Geographic Coordinate Extraction
- New services found, can be transformed into RDF triples and added to the repository!

Ge(o)Lo(cator) System Description – Architecture



RDF Store Enrichment, for service

Localization via web crawling

TABLE I. COMPARISON TABLE REPORTING EVALUATION DETAILS FOR EVALUATION TASK 1: ADDRESS ARRAY EXTRACTION, AND TASK 2: GEOGRAPHIC COORDINATES EXTRACTION (EMPLOYING BOTH THE SMART CITY REPOSITORY AND THE GOOGLE GEOCODING API).

Evaluation Tasks	TP	FP	FN	TN	Precision	Recall	F-Measure
1) Address Array Extraction	74.5%	7.8%	5.9%	11.8%	90.5%	92.7%	91.6%
2a) Geographic Coordinates Extraction (Smart City Semantic Repository)	57.8%	4.7%	29.5%	8.0%	92.5%	66.2%	77.1%
2b) Geographic Coordinates Extraction (Google Geocoding)	48.9%	31.1%	11.1%	8.9%	61.1%	81.5%	69.8%

- **Precision** rate for geographic coordinates extraction (employing the Smart City Semantic Repository) has increased, with respect to the value obtained in the evaluation of address array extraction.
- Slightly decreasing **TN** rate for Test (2a) with respect to Test (1): exploiting the extraction of high level features (such as building names) allows the system to obtain correct coordinates even for domains with incomplete Address Array.
- **Recall** rate for Test (2a) significantly decrease with respect to Test (1). This is due mainly to the noise generated by the supplementary logic and the extended semantic queries required to obtain the geographical coordinates.
- Higher **Recall** rate achieved when using the Google Geocoding APIs: Google Repository is by far larger than DISIT Smart City RDF datastore, so that it is able to index a huge amount of resources, even if this can affect the precision rate.

VIP names identification

- Searching RDF and/or MySQL stores looking for VIP names (citations) into strings:
 - *Via Leonardo Da Vinci*
 - *Piazza Lorenzo il Magnifico*
 - *Palazzo Medici Riccardi*
 - *Etc.*
- The idea is to link those entities with LD/LOD information as dbPedia information and

VIP names identification

Synonyms

Enter here a new name

a b c **D** e f g h i j k l m n o p q r s t u v w x y z X

'D' Pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

☒ All
☐ Dbpedia

☐ EclapU
☐ Both

☒ Norm

STATISTICHE

SETTINGS

Daria Guerrini	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Daria Marušića	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Daria Menichetti	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Daria Panettieri	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Darimonde Odette	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Abela	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Aggioli	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Antiseri	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Argento	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Benedetti	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Cincillà	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Di Donato	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Ferrara	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Fo	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Fo	<input type="radio"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Dario Fo	<input type="radio"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Fo Dario	<input type="radio"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Dario Gessati	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Giannozzi	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dario Giannini	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>

140544(20 mostrate) occorrenze di: Dario Fo

Axoid: urn:axmedis:00000:obj:e75313b5-90ab-4176-baab-aab19375eafc

Field: Group

Value: [Dario Fo](#) & Franca Rame Archive, CTFR, Milano, Italia

Link: <http://www.eclap.eu/urn:axmedis:00000:obj:e75313b5-90ab-4176-baab-aab19375eafc>

Axoid: urn:axmedis:00000:obj:e3edb41e-cb17-4ed5-a973-d516b000749b


Field: Group

Value: [Dario Fo](#) & Franca Rame Archive, CTFR, Milano, Italia

Link: <http://www.eclap.eu/urn:axmedis:00000:obj:e3edb41e-cb17-4ed5->

1 info di: Dario Fo

☒ Dbpedia
☒ Eclap

 http://dbpedia.org/resource/Dario_Fo

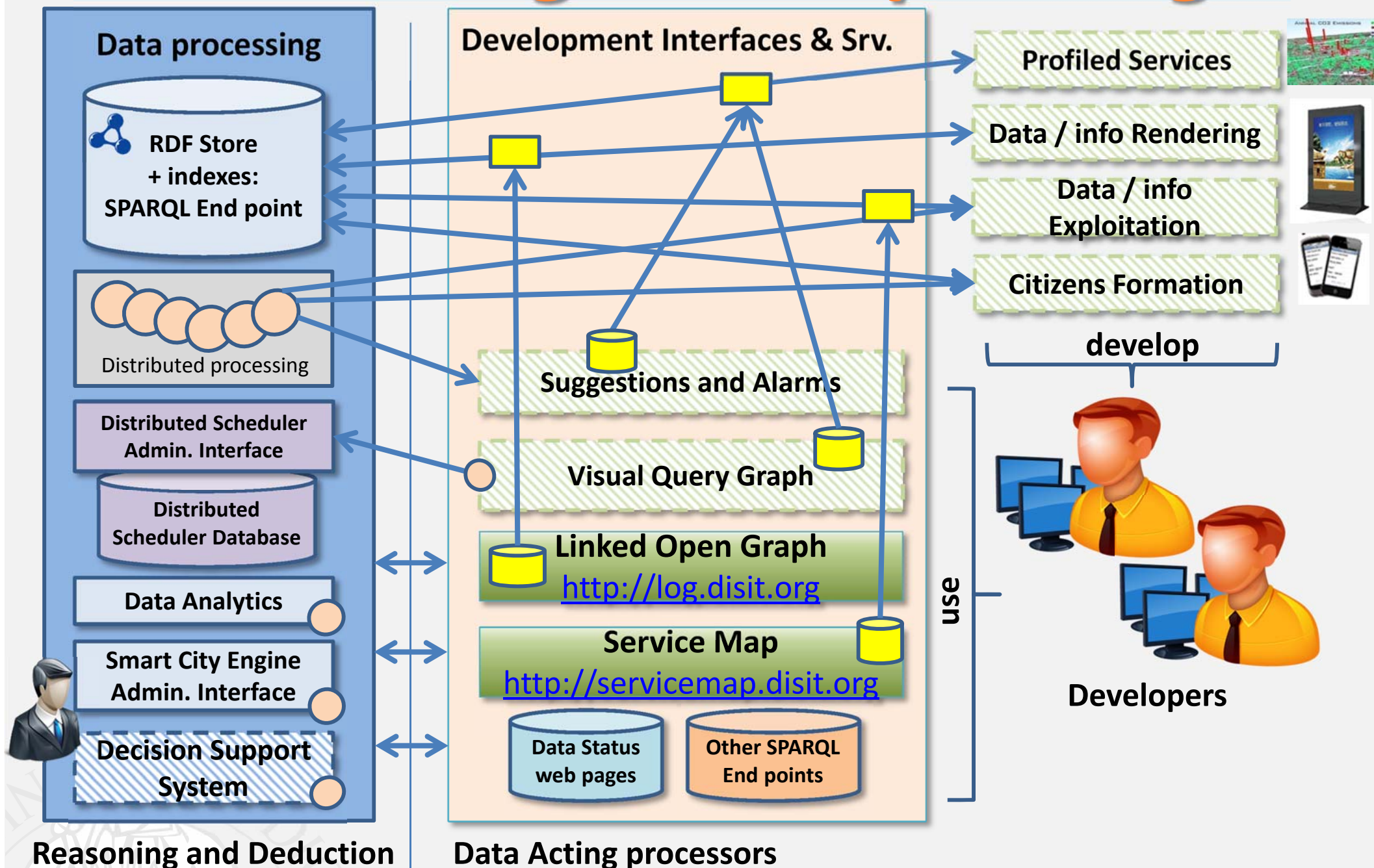
Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine** 
- **Development Interfaces**
- **Sii-Mobility**

Smart City Engine (Reasoning and Exploiting)

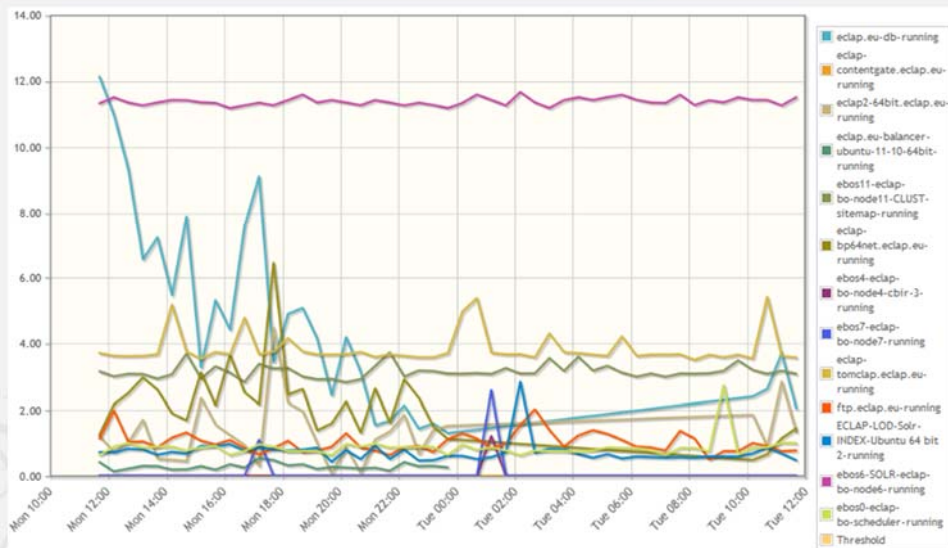
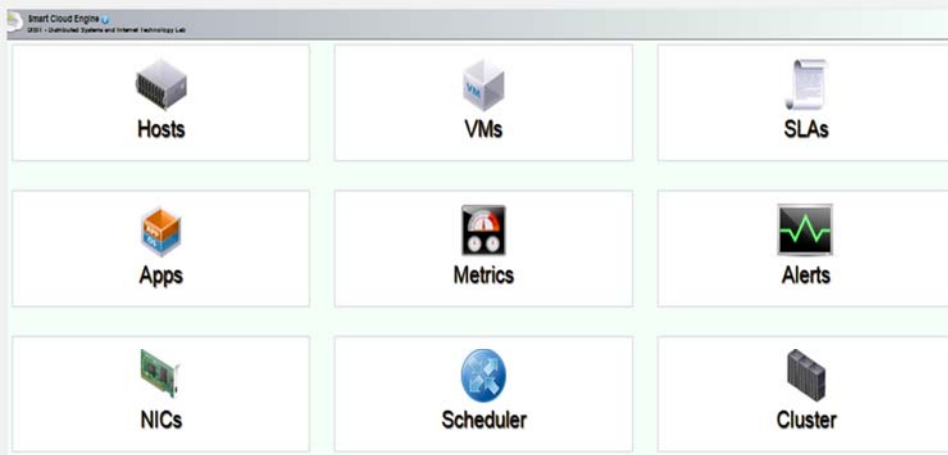
- **Service Level Agreements**
- **Distributed SPARQL queries**
- **Decision Support System Processes**
 - **Model processes**
 - **Computes processes on the basis of data coming from OD, PD, SD, RTD, SMD and soundages**
 - **under construction**

Reasoning and Exploiting



- **SCE Monitors** city at level of
 - Active Objects/Query on the base of City Services
 - Entity Centric Search Analysis of correlations and causalities, based on hypothesis driven
- **SCE Reads from RDF Store the City Service SLAs under control,** for example:
 - AVM have to provide information for the 80% of time, with maximum delay of 5%
 - Sensors have to provide service for the 80% of time
 - *Parking should have not be always full ...*
 - *Forecasts have to be available in the XX % of the cases*
 - *Taxi availability has to be higher than*
 - *Queue on Service XXXX has to be shorter than 20 minutes in the peak hours..*
- **For each City Service SLA** an Active Process including:
 - Computing SPARQL Query
 - One or more actions to be executed: reshape, reschedule, change of service, ...
 - Send alerts to the Service responsible and to administrators
 - Graphs production with respect to reference triggering values

- Status: in progress
- SLA derived from Cloud SLA
- Model similar to Smart Cloud Engine of ICARO



Smart Cloud Engine 				
DISIT - Distributed Systems and Internet Technology Lab				
Id	Slas	Timestamp	Alarms	Tot
394443	urn:cloudicaro:ServiceLevelAgreement:mobmed	2014-12-16 11:54:58	4	15
394434	urn:cloudicaro:ServiceLevelAgreement:log	2014-12-16 11:53:57	1	9
394430	urn:cloudicaro:ServiceLevelAgreement:inea-lab	2014-12-16 11:52:56	2	4
394407	urn:cloudicaro:ServiceLevelAgreement:icaro-disit	2014-12-16 11:51:59	2	23
394400	urn:cloudicaro:ServiceLevelAgreement:disit-lab	2014-12-16 11:51:57	1	7
394354	urn:cloudicaro:ServiceLevelAgreement:icaro-dev	2014-12-16 11:51:01	4	46

[Http://192.168.0.45](http://192.168.0.45)

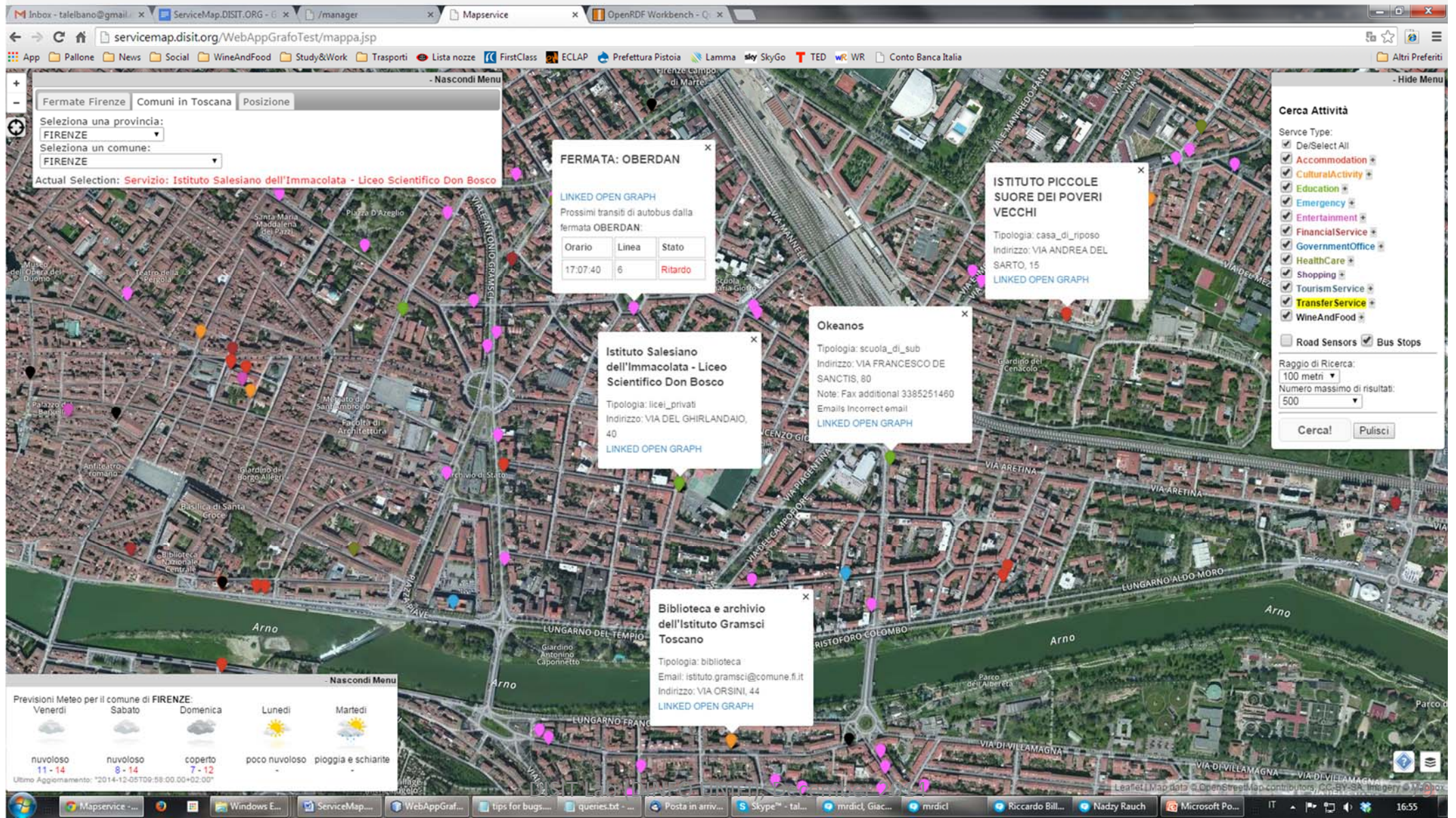
Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces** 
- **Sii-Mobility**

Development Interfaces

- **Service map:** <http://servicemap.disit.org>
 - service based on OpenStreetMaps that allows to search services available in a preset range from the selected bus stop.
- **Linked Open Graph:** <http://log.disit.org>
 - 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.
- **Ontology Documentation:** <http://www.disit.org/6507>,
 - <http://www.disit.org/5606>, <http://www.disit.org/6461>
- **Data Status Web pages:** active
- **Visual Query Graph:** under development

<http://servicemap.disit.org>



The screenshot displays the Service Map web application interface. The main map shows Florence, Italy, with various service points marked by colored pins. Several information panels are visible:

- Top Left:** Navigation and search controls. Includes a dropdown for "Fermate Firenze" and "Comuni in Toscana". Below it, a search bar for "Seleziona una provincia:" and "Seleziona un comune:". The "Actual Selection" shows "Servizio: Istituto Salesiano dell'Immacolata - Liceo Scientifico Don Bosco".
- Top Right:** "Cerca Attività" (Search Activities) panel. It lists various service types with checkboxes: De/Select All, Accommodation, CulturalActivity, Education, Emergency, Entertainment, FinancialService, GovernmentOffice, HealthCare, Shopping, TourismService, TransferService, and WineAndFood. It also includes a "Raggio di Ricerca:" (Search Radius) set to 100 metri and a "Numero massimo di risultati:" (Maximum number of results) set to 500. Buttons for "Cerca!" and "Pulisci" are at the bottom.
- Center:** A pop-up window for "FERMATA: OBERDAN" showing a table of bus schedules.

Orario	Linea	Stato
17.07.40	6	Ritardo
- Bottom Left:** A weather forecast panel for "FIRENZE" showing conditions for Venerdì, Sabato, Domenica, Lunedì, and Martedì.

Venerdì	Sabato	Domenica	Lunedì	Martedì
nuvoloso	nuvoloso	coperto	poco nuvoloso	pioggia e schiarite
11 - 14	8 - 14	7 - 12		
- Bottom Center:** A pop-up window for "Istituto Salesiano dell'Immacolata - Liceo Scientifico Don Bosco" showing its location, address, and contact information.

Tipologia: licei_privati
Indirizzo: VIA DEL GHIRLANDAIO, 40
LINKED OPEN GRAPH
- Bottom Right:** A pop-up window for "Biblioteca e archivio dell'Istituto Gramsci Toscano" showing its location, address, and contact information.

Tipologia: biblioteca
Email: istituto.gramsci@comune.fi.it
Indirizzo: VIA ORSINI, 44
LINKED OPEN GRAPH

Line 4

FERMATA : GUA STI

LINKED OPEN GRAPH

Prossimi transiti di autobus dalla
fermata GUA STI:

Orario	Linea	Stato	Ride
09:08:02	4	Ritardo	4854072
09:08:26	4	Ritardo	4854042
09:14:16	4	Ritardo	4854112
09:15:50	4	In orario	4854077
09:25:18	4		
09:25:48	4		

FERMATA : STATUTO 04

LINKED OPEN GRAPH

Prossimi transiti di autobus dalla
fermata STATUTO 04:

Orario	Linea	Stato	Ride
09:09:04	4	Ritardo	4854072
09:09:28	4	Ritardo	4854042
09:15:06	4	Ritardo	4854112
09:16:52	4	In orario	4854077
09:26:20	4	In orario	4854020
09:26:50	4	In orario	4853998

FERMATA : FABBRONI

LINKED OPEN GRAPH

Prossimi transiti di autobus dalla
fermata FABBRONI:

Orario	Linea	Stato	Ride
09:09:24	4	In orario	4854020
09:10:42	4	In orario	4853998
09:19:24	4	Anticipo	4854133
09:20:48	4	In orario	4854126

FERMATA : GIOIA

LINKED OPEN GRAPH

Prossimi transiti di autobus dalla
fermata GIOIA:

Orario	Linea	Stato	Ride
09:10:22	4	In orario	4854020
09:11:40	4	In orario	4853998
09:20:22	4	Anticipo	4854133
09:21:46	4	In orario	4854126

Linea 6

The screenshot shows a web browser window displaying a map application. The browser's address bar shows the URL: `servicemap.disit.org/WebAppGrafoTest/mappa.jsp`. The application interface includes a sidebar on the left with a search bar and a list of service categories. The main area displays a satellite map of Florence, Italy, with several bus stop locations marked by pink dots. Four pop-up windows are visible, each showing the next bus arrivals for a specific stop. The stops are LEOPARDI, OBERDAN, FRA' PAOLO SARPI, and ALBERTI. Each window lists the time, line number, status, and ride ID. The interface also includes a weather forecast for Florence at the bottom left.

Search Bar:

Seleziona una linea: Line 6

Seleziona una fermata: TUTTE LE FERMATE

Actual Selection: Fermata Bus: LEOPARDI

Prossimi transiti di autobus dalla fermata LEOPARDI:

Orario	Linea	Stato	Ride
08:54:06	6	In orario	4868247
09:01:08	6	In orario	4867793

Prossimi transiti di autobus dalla fermata OBERDAN:

Orario	Linea	Stato	Ride
08:56:34	6	In orario	4868247
09:03:14	6	In orario	4867793

Prossimi transiti di autobus dalla fermata FRA' PAOLO SARPI:

Orario	Linea	Stato	Ride
08:57:54	6	In orario	4868247
09:04:38	6	In orario	4867793

Prossimi transiti di autobus dalla fermata ALBERTI:

Orario	Linea	Stato	Ride
08:59:48	6	In orario	4868247
09:07:00	6	In orario	4867793

Cerca Attività

Service Type:

- ☐ De/Select All
- ☐ Accommodation
- ☐ CulturalActivity
- ☐ Education
- ☐ Emergency
- ☐ Entertainment
- ☐ FinancialService
- ☐ GovernmentOffice
- ☐ HealthCare
- ☐ Shopping
- ☐ TourismService
- ☒ TransferService
- ☐ WineAndFood

☐ Road Sensors ☐ Bus Stops

Raggio di Ricerca: 100 metri

Numero massimo di risultati: 100

Previsioni Meteo per il comune di FIRENZE:

Martedì	Mercoledì	Giovedì	Venerdì	Sabato
poggia moderata-forte	nuvoloso	poco nuvoloso	nuvoloso	nuvoloso
9-13	6-14	3-13		

Ultimo Aggiornamento: "2014-12-15T15:21:00.00+02:00"

<http://servicemap.disit.org>

The screenshot displays the ServiceMap application interface, which is a web-based map of Florence, Italy. The interface includes a search bar, a filter panel on the right, and several information popups for specific locations.

Search and Filter Panel (Left):

- Selezione una provincia: FIRENZE
- Selezione un comune: FIRENZE
- Actual Selection: Servizio: Istituto Salesiano dell'Immacolata - Liceo Scientifico Don Bosco

Filter Panel (Right):

Cerca Attività

Service Type:

- ☒ De/Select All
- ☒ Accommodation
- ☒ CulturalActivity
- ☒ Education
- ☒ Emergency
- ☒ Entertainment
- ☒ FinancialService
- ☒ GovernmentOffice
- ☒ HealthCare
- ☒ Shopping
- ☒ TourismService
- ☒ TransferService
- ☒ WineAndFood

☐ Road Sensors ☒ Bus Stops

Raggio di Ricerca: 100 metri

Numero massimo di risultati: 500

Cerca! Pulisci

Information Popups:

- FERMATA: OBERDAN**
 - LINKED OPEN GRAPH
 - Prossimi transiti di autobus dalla fermata OBERDAN:

Orario	Linea	Stato
17.07.40	6	Ritardo
- ISTITUTO PICCOLE SUORE DEI POVERI VECCHI**
 - Tipologia: casa_di_riposo
 - Indirizzo: VIA ANDREA DEL SARTO, 15
 - LINKED OPEN GRAPH
- Okeanos**
 - Tipologia: scuola_di_sub
 - Indirizzo: VIA FRANCESCO DE SANCTIS, 80
 - Note: Fax additional 3385251460
 - Emails Incorrect email
 - LINKED OPEN GRAPH
- Istituto Salesiano dell'Immacolata - Liceo Scientifico Don Bosco**
 - Tipologia: licei_privati
 - Indirizzo: VIA DEL GHIRLANDAIO, 40
 - LINKED OPEN GRAPH
- Biblioteca e archivio dell'Istituto Gramsci Toscano**
 - Tipologia: biblioteca
 - Email: istituto.gramsci@comune.fi.it
 - Indirizzo: VIA ORSINI, 44
 - LINKED OPEN GRAPH

Bottom Panel:

Previsioni Meteo per il comune di FIRENZE:

Venerdì	Sabato	Domenica	Lunedì	Martedì
nuvoloso	nuvoloso	coperto	poco nuvoloso	pioggia e schiarite
11 - 14	8 - 14	7 - 12		

Ultimo Aggiornamento: "2014-12-05T09:58:00.00+02:00"

DISIT Lab (DINFO-UNIFI) December 2014

<http://servicemap.disit.org>

The screenshot displays the ServiceMap web application interface. The main map shows the area around Empoli, Italy, with various streets and landmarks. A popup window for sensor EM0100201 is open, displaying the following data:

EM0100201
 Tipologia: sensore
 Indirizzo: VIA DELLA REPUBBLICA
 LINKED OPEN GRAPH
 Dati del sensore: EM0100201:
 Ultimo Aggiornamento: "2014-12-01T19:30:00.000+01:00"

Avg Distance(m)	Avg Time (sec)	Occupancy (%)	Concentration (car/km)
69.23333	5.285	0.06324805	13

Vehicle Flow (car/h)	Avg Speed (Km/h)	Threshold Perc (%)	Speed Perc (%)
100	43.33	0.0	N.A.

On the right side, there is a 'Cerca Attività' (Search Activities) panel with a list of service types and a search radius of 100 metri. The bottom left corner shows weather forecasts for Empoli.

DISIT Lab (DINFO) December 2014

<http://servicemap.disit.org>

- Position on the map by:
 - selected point, current position, bus-stop, municipality, selected PIN
- Selecting services by kind, numbers, distances from the position, as:
 - Static data as: restaurants, museums, bus stops, etc.
 - real time data as: sensors, parking, meteo, AVM,
- Connections to data model via Linked Open Graph for understanding the RDF Storage data, ontology and SPARQL queries, save and share models, etc.
- Change on map: kind, zoom, pan, etc. etc.

<http://log.disit.org>

Linked Open Graph

log.disit.org/service/?graph=df5b46701702380TO...562768e...

Linked Open Graph

SiiMobility (by DISIT)

Examples:

- VIA GIACOMO MATTEOTTI
- Bagno a ripoli
- Florence

Choose a class:

Search for keyword

keyword:

uri: http://...

Request

Your data

sparql endpoint: (optional)

http://...

uri: http://...

Request

Status

Requests:

http://www.disit.dinfo.unifi.it/SiiMobility/MUSE...

Remove

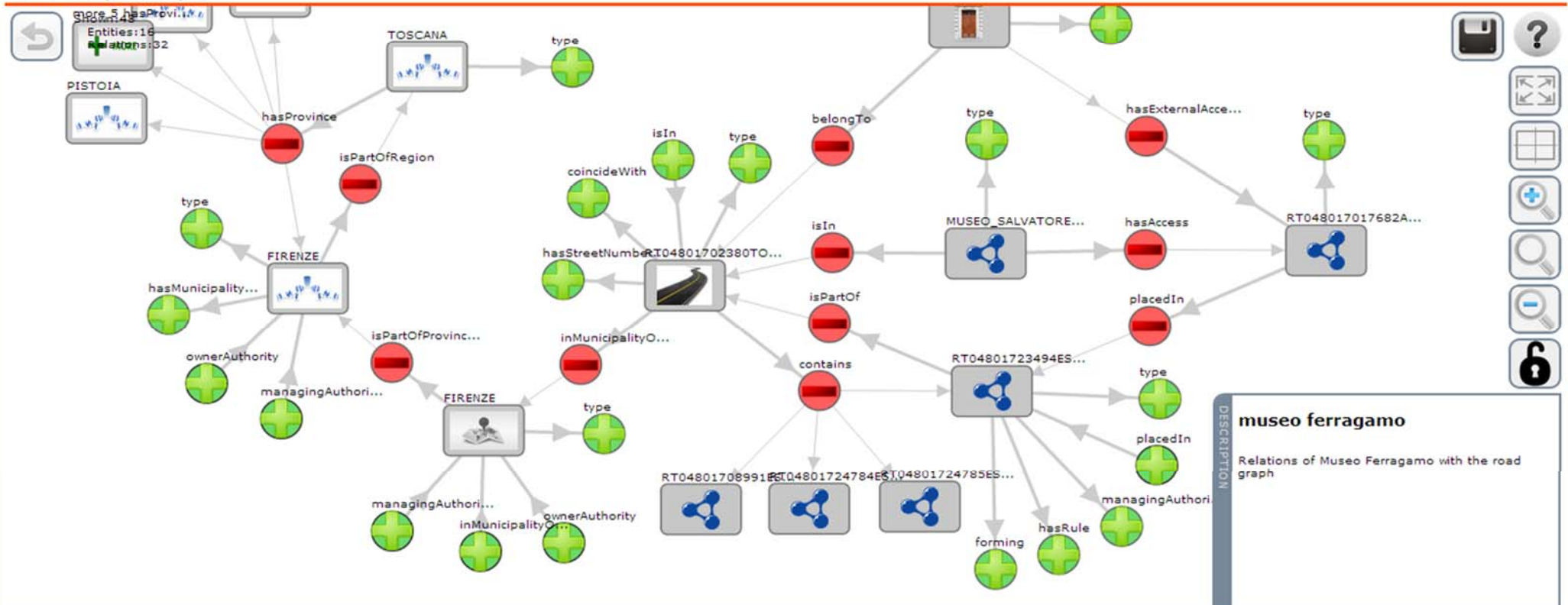
Clear

Type of relations

Select all Deselect all Invert Hide all inverse

- | | |
|-------------------------------------------------------|------------------------------------------------------|
| <input checked="" type="checkbox"/> belongTo | <input checked="" type="checkbox"/> coincideWith |
| <input checked="" type="checkbox"/> contains | <input type="checkbox"/> depiction |
| <input type="checkbox"/> ends | <input checked="" type="checkbox"/> forming |
| <input type="checkbox"/> has | <input checked="" type="checkbox"/> hasAccess |
| <input checked="" type="checkbox"/> hasExternalAccess | <input checked="" type="checkbox"/> hasMunicipality |
| <input checked="" type="checkbox"/> hasProvince | <input checked="" type="checkbox"/> hasRule |
| <input checked="" type="checkbox"/> hasStreetNumber | <input checked="" type="checkbox"/> inMunicipalityOf |
| <input checked="" type="checkbox"/> isIn | <input checked="" type="checkbox"/> isPartOf |
| <input checked="" type="checkbox"/> isPartOfProvince | <input checked="" type="checkbox"/> isPartOfRegion |
| <input checked="" type="checkbox"/> managingAuthority | <input checked="" type="checkbox"/> ownerAuthority |
| <input checked="" type="checkbox"/> placedIn | <input type="checkbox"/> sameAs |
| <input checked="" type="checkbox"/> seeAlso | <input type="checkbox"/> starts |

Linked Open Graph

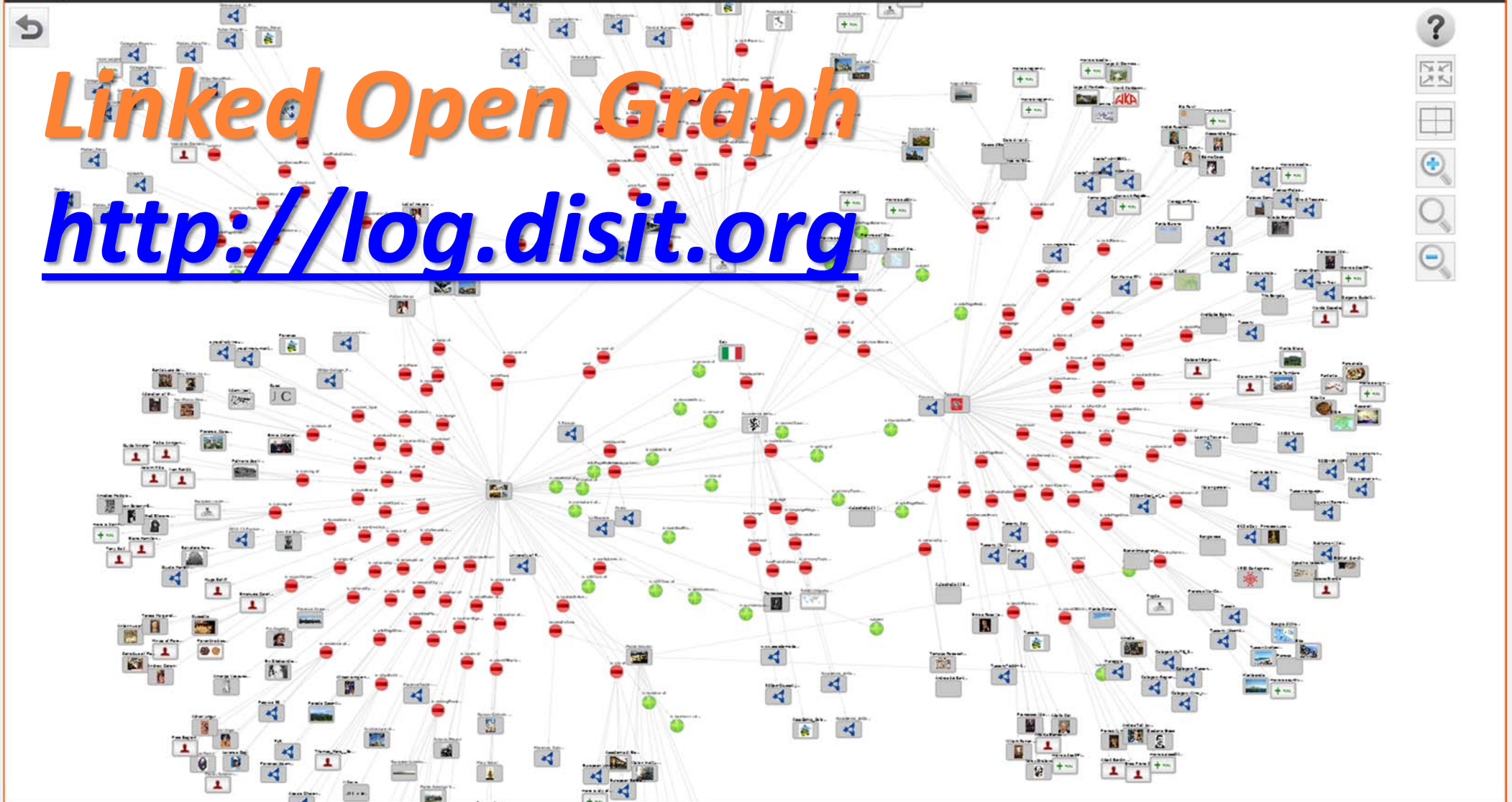


Data Graph

Close

Linked Open Graph

<http://log.disit.org>



Type of relations

Select all Deselect all Invert

- | | | | | | | | | | | |
|--------------------------------------------------------|----------------------------------------------------|-----------------------------------------------|--------------------------------------------------|----------------------------------------------------------|---------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------|------------------------------------------------------|------------------------------------------------------|----------------------------------------------------------|
| <input type="checkbox"/> sameAs | <input type="checkbox"/> depiction | <input checked="" type="checkbox"/> seeAlso | <input type="checkbox"/> is province of | <input type="checkbox"/> is region of | <input type="checkbox"/> country | <input type="checkbox"/> mayorParty | <input checked="" type="checkbox"/> saint | <input checked="" type="checkbox"/> mayor | <input type="checkbox"/> region | <input type="checkbox"/> type |
| <input checked="" type="checkbox"/> subject | <input checked="" type="checkbox"/> homepage | <input type="checkbox"/> wikiPageUsesTemplate | <input checked="" type="checkbox"/> thumbnail | <input checked="" type="checkbox"/> wikiPageExternalLink | <input checked="" type="checkbox"/> wasDerivedFrom | <input checked="" type="checkbox"/> hasPhotoCollection | <input checked="" type="checkbox"/> wordnet_type | <input type="checkbox"/> isPrimaryTopicOf | <input type="checkbox"/> is battles of | <input checked="" type="checkbox"/> is training of |
| <input checked="" type="checkbox"/> is restingPlace of | <input type="checkbox"/> is comune of | <input type="checkbox"/> is after of | <input checked="" type="checkbox"/> is museum of | <input checked="" type="checkbox"/> is title of | <input type="checkbox"/> is origin of | <input checked="" type="checkbox"/> is headquarters of | <input checked="" type="checkbox"/> is location of | <input checked="" type="checkbox"/> is city of | <input type="checkbox"/> is battle of | <input checked="" type="checkbox"/> is see of |
| <input type="checkbox"/> is restingPlace of | <input checked="" type="checkbox"/> is province of | <input type="checkbox"/> is place of | <input checked="" type="checkbox"/> is origin of | <input checked="" type="checkbox"/> is production of | <input checked="" type="checkbox"/> is placeOfBurial of | <input type="checkbox"/> is place of | <input checked="" type="checkbox"/> is nonplace of | <input checked="" type="checkbox"/> is recordable of | <input checked="" type="checkbox"/> is mainShrine of | <input checked="" type="checkbox"/> is route function of |

<http://log.disit.org>

- **Access and query:** dbpedia, europeana, eclap, geonames, gettyvocabulary, etc. etc.
- **Browse** and define rendering graphs
- **Save and share** graphs for reedit via email
- **Embed** linked open data graph in your web pages, selecting also relations you prefer
- See JLVC publication

<http://dx.doi.org/10.1016/j.jvlc.2014.10.003>

Smart-city Ontology

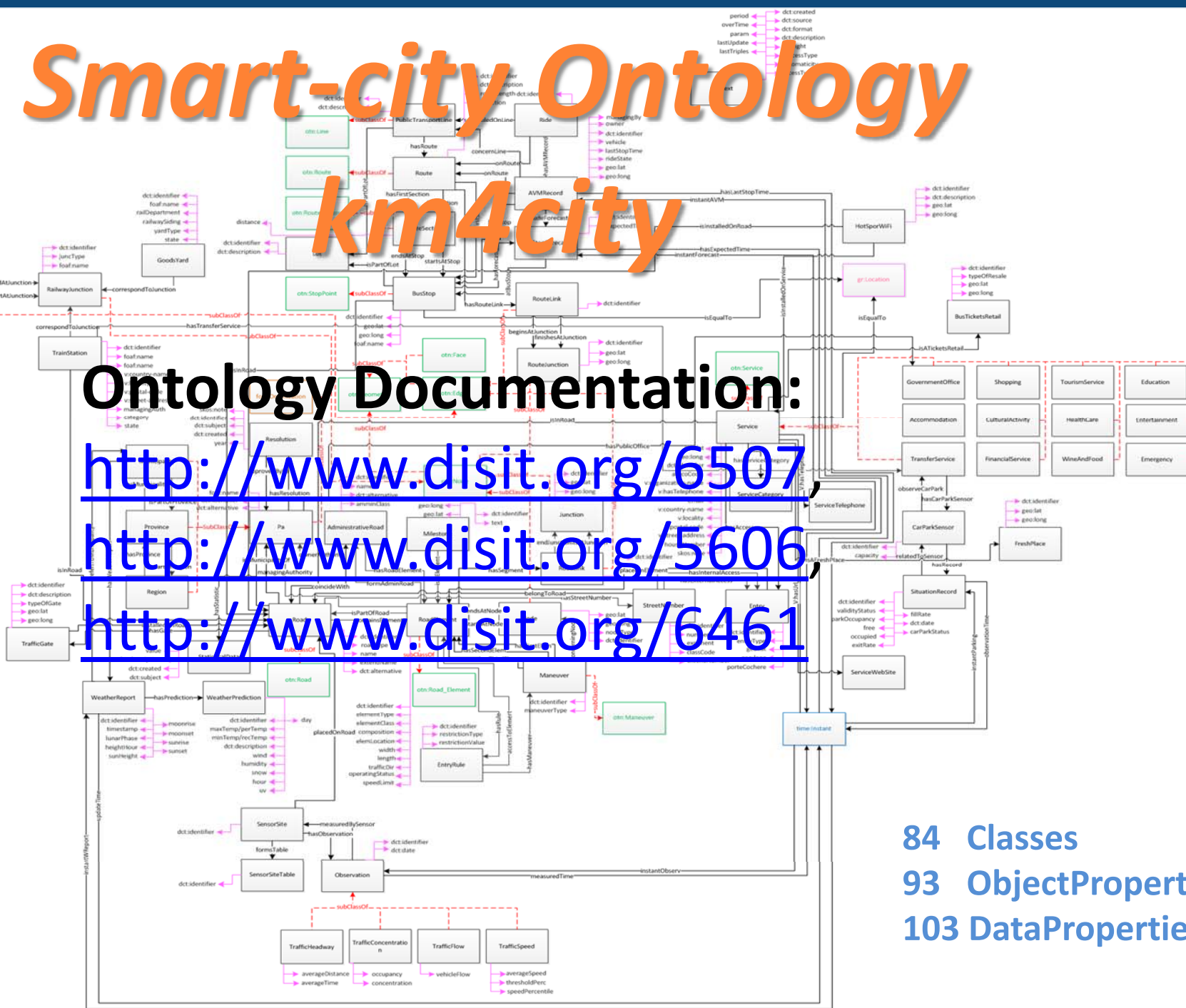
km4city

Ontology Documentation:

<http://www.disit.org/6507>

<http://www.disit.org/5606>

<http://www.disit.org/6461>



84 Classes
93 ObjectProperties
103 DataProperties

Data Status Web Pages

- [Http://www.disit.org/6056](http://www.disit.org/6056) gruppo smart city



DataSet: Arte e cultura
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>
157 Accessi Voto ★★★★★ Azioni



DataSet: Banche
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://localhost/phpMyEdit-5.7.1/Table16.php>
79 Accessi Voto ★★★★★ Azioni



DataSet: Corrieri espresso
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>
63 Accessi Voto ★★★★★ Azioni



DataSet: Emergenze
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>
70 Accessi Voto ★★★★★ Azioni



Dataset: Enogastronomia
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>
56 Accessi Voto ★★★★★ Azioni



DataSet: Imprese del commercio
Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>
40 Accessi Voto ★★★★★ Azioni

DATASET:MODELLO.....NOME DEL DATA SET.....

Questa pagina web contiene informazioni aggiuntive al momento non gestite nella tabella <http://192.168.0.21/phpMyEdit-5.7.1/Table16.php>

SORGENTE

- Identificativo del campo chiave del database
- link per vedere tabella del db
- posizione e nome del file scaricato, URL

PRIMA FASE

- descrizione del processo di ingestion, prima fase
 - descrizione eventuali elaborazioni a mano e criticità
 - posizione URL e nomi di file di elaborazione a mano o java, o perl o che altro
 - fasi del processo: parti a mano, parti con altri tool----> ETL
 - dove sta l'ETL URL, come si chiama
 - descrizione sommaria dell'ETL
 - descrizione e posizione dei file di output
 - contributor name

SECONDA FASE

- descrizione del processo di ingestion, seconda fase
 - descrizione eventuali elaborazioni a mano e criticità
 - posizione URL e nomi di file di elaborazione a mano o java, o perl o che altro
 - fasi del processo: parti a mano, parti con altri tool----> ETL
 - dove sta l'ETL URL, come si chiama
 - descrizione sommaria dell'ETL
 - descrizione e posizione dei file di output (triple)
 - contributor name

SEMANTIC MODEL

- Descrizione del mapping delle triple rispetto al KB
- contributor name


OPEN PROBLEM

- problemi residui:
 - eventuali inconsistenze, elementi da riconciliare
 - suggerimenti per non portarsi questi errori avanti nelle prossime ingestion
 - contributor name for each comment as: "Nesi:"

CONTRIBUTORS

- name of contributors and date of the last change

Major topics addressed

- **Big Data Smart City Architecture**
- **Smart-city Ontology**
- **Data Ingestion and Mining**
- **Distributed and real time processes**
- **RDF processing**
- **Smart City Engine**
- **Development Interfaces**
- **Sii-Mobility** 

Sii-Mobility

- **Title:** Support of Integrated Interoperability for Services to Citizens and Public Administration
- **Objectives:**
 1. Reduction of social costs of mobility;
 2. Simplify the use of mobility systems;
 3. Developing working solutions and application, with testing methods;
 4. Contribute to standardization organs, and establishing relationships with other smart cities' management systems.



The Sii-Mobility platform will be capable to provide support for SME and Public Administrations. Sii-Mobility consists in a federated/integrated interoperable solution aimed at enabling a wide range of specific applications for private services to citizen and commercial services to SME.

- **Coordinatore Scientifico:** *Paolo Nesi, DISIT DINFO UNIFI*
- **Partner:** ECM; Swarco Mizar; University of Florence (svariati gruppi+CNR); Inventi In20; Geoin; QuestIT; Softec; T.I.M.E.; LiberoLogico; MIDRA; ATAF; Tiemme; CTT Nord; BUSITALIA; A.T.A.M.; Sistemi Software Integrati; CHP; Effective Knowledge; eWings; Argos Engineering; Elfi; Calamai & Agresti; KKT; Project; Negentis.
- **Link:** <http://www.disit.dinfo.unifi.it/siimobility.html>

Comments from
citizens, Social Media



AVM Public
transport



Sensors,
monitoring system

Materials



Sensors for
private transport

Parking
sensors



Traffic monitoring,
highways



Railways



Environmental
parameters

Public Services

Ordinances: events,
large scale works
...



UTC



Infomobility



Telematic
passages, ZTL

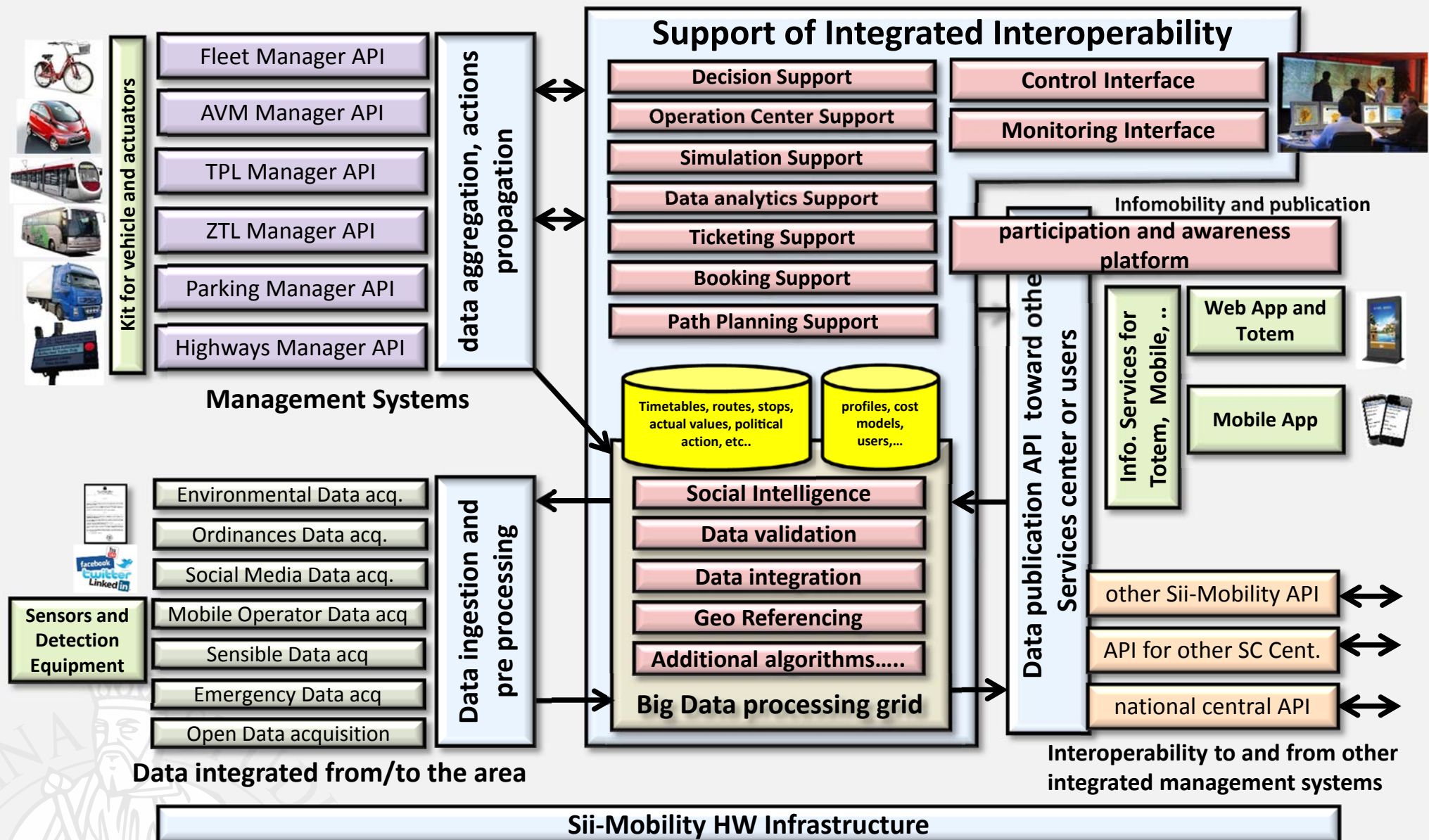
Emergency,
police, 118



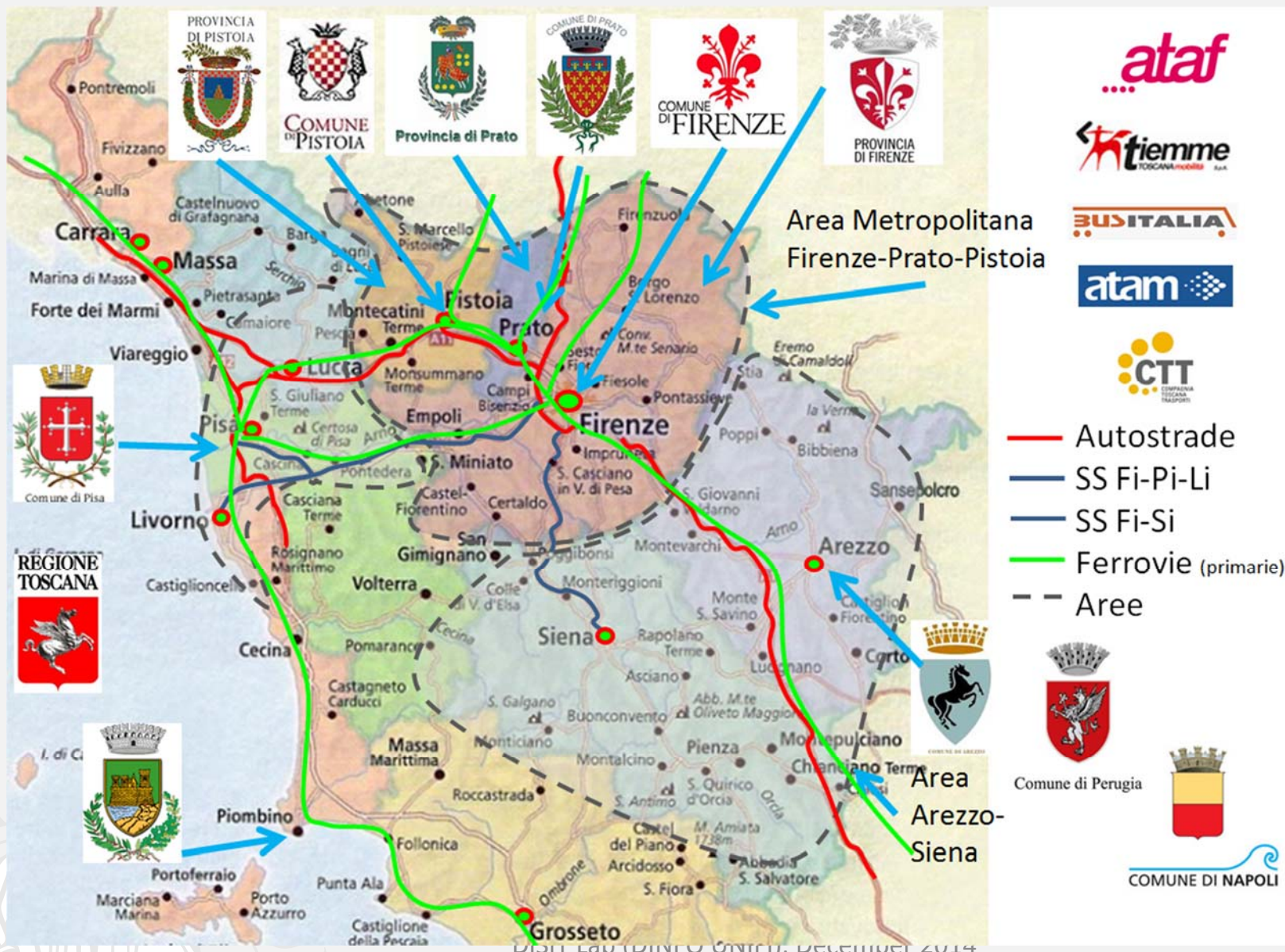
Sii-Mobility: main scenarios

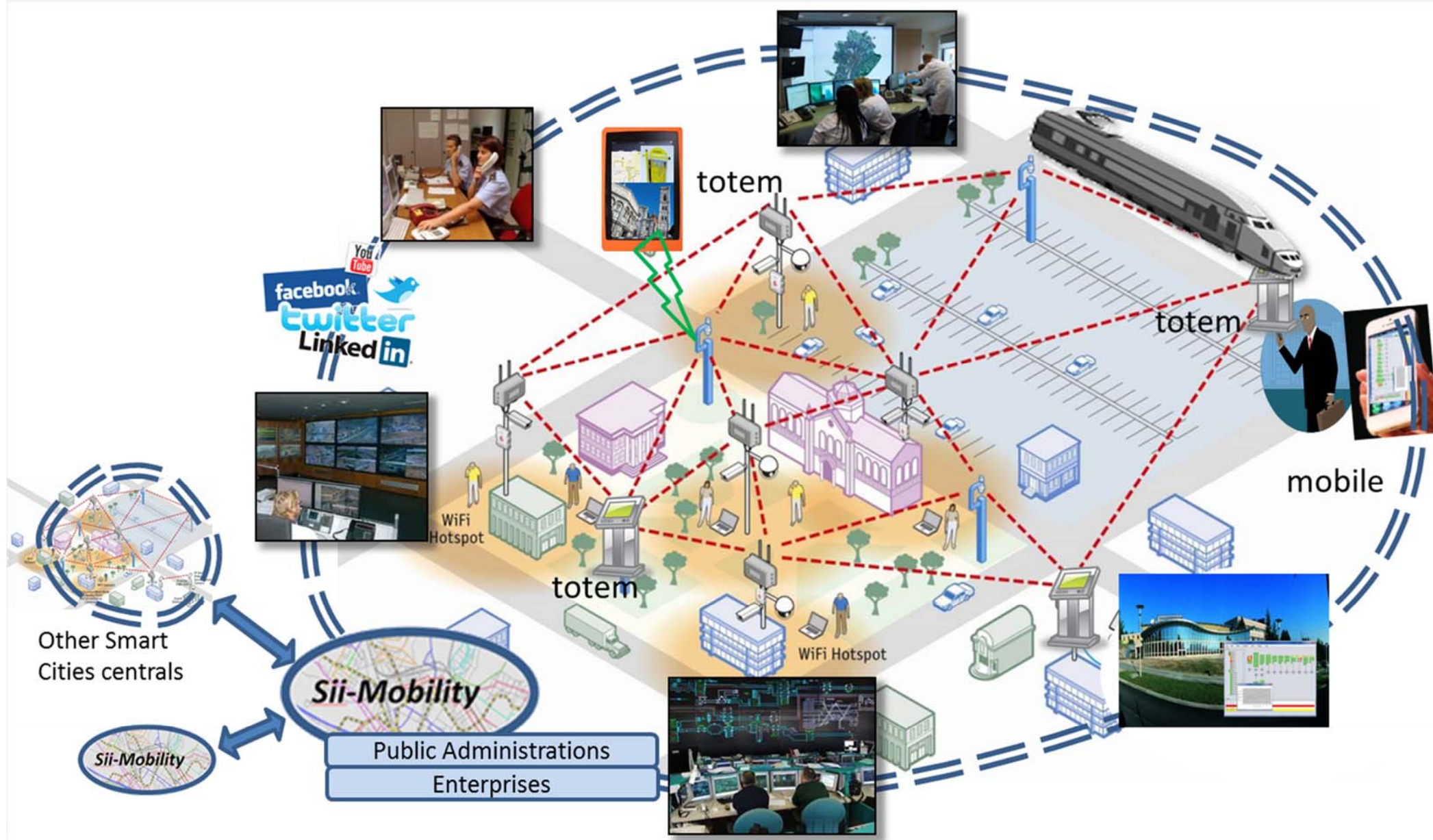
- solutions of connected guide / path
 - personalized services, alarms, vehicle / person receives commands and information in real time, personalized and contextualized;
- Platform of participation and awareness
 - to receive information from the citizen, the citizen as Intelligent Sensor, to inform and educate the citizen, through totem, mobile applications, web applications, etc .;
- personalized management of access policies
 - Incentive policies of deterrence and the use of the vehicle, Credit mobility, flow monitoring;
- interoperability and integration of management systems
 - contribution to standards, testing and data validation, data reconciliation, etc .;
- integration of methods of payment and identification
 - Political pay-per-use, monitoring user behavior;
- dynamic management of the boundaries of the areas controlled traffic
 - dynamic pricing and category of vehicles;
- management shared network data exchange between services (PA and private)
 - data reliability and separation of responsibilities, integration of open data, reconciliation,;
- monitoring of supply and demand of public transport in real time
 - solutions for the integration and processing of data.

Sii-Mobility Architecture



- Experimentations and validation in Tuscany
- Integration with present central station and subsystems





- Smart City Group on DISIT and several slides: www.disit.org
- P. Bellini, M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Building vs Data Harvesting and Cleaning for Smart-city Services", International Journal of Visual Language and Computing, Elsevier, <http://dx.doi.org/10.1016/j.jvlc.2014.10.023>, P. Bellini, P. Nesi, A. Venturi, "Linked Open Graph: browsing multiple SPARQL entry points to build your own LOD views", International Journal of Visual Language and Computing, Elsevier, 2014, DOI information: <http://dx.doi.org/10.1016/j.jvlc.2014.10.003> ,
- A. Bellandi, P. Bellini, A. Cappuccio, P. Nesi, G. Pantaleo, N. Rauch, "ASSISTED KNOWLEDGE BASE GENERATION, MANAGEMENT AND COMPETENCE RETRIEVAL", [International Journal of Software Engineering and Knowledge Engineering, World Scientific Publishing Company](#), press, vol.32, n.8, pp.1007-1038, Dec. 2012, DOI: 10.1142/S021819401240013X
- P. Bellini, M. Di Claudio, P. Nesi, N. Rauch, "Tassonomy and Review of Big Data Solutions Navigation", as Chapter 2 in "Big Data Computing", Ed. Rajendra Akerkar, Western Norway Research Institute, Norway, Chapman and Hall/CRC press, ISBN 978-1-46-657837-1, eBook: 978-1-46-657838-8, july 2013, pp.57-101, DOI: 10.1201/b16014-4
- P. Nesi, G. Pantaleo and M. Tenti, "Ge(o)Lo(cator): Geographic Information Extraction from Unstructured Text Data and Web Documents", SMAP 2014, 9th International Workshop on Semantic and Social Media Adaptation and Personalization, November 6-7, 2014, Corfu/Kerkyra, Greece. technically co-sponsored by the IEEE Computational Intelligence Society and technically supported by the IEEE Semantic Web Task Force. www.smap2014.org
- P. Bellini, P. Nesi and N. Rauch, "Smart City data via LOD/LOG Service", [LOD2014](#), Workshop Linked Open Data: where are we?, organized by W3C Italy and CNR, Rome, 2014

Big Data Smart City processes and tools

Distributed Data Intelligence and Technologies Lab
Distributed Systems and Internet Technologies Lab

Prof. Paolo Nesi

DISIT Lab

Dipartimento di Ingegneria dell'Informazione

Università degli Studi di Firenze

Via S. Marta 3, 50139, Firenze, Italia

tel: +39-055-2758515, fax: +39-055-2758570

<http://www.disit.dinfo.unifi.it>

paolo.nesi@unifi.it

