INFORMATION MODEL AND INTEROPERABILITY

Project Title	RESOLUTE
Project number	653460
Deliverable number	4.4
Version	9.0
State	final
Confidentially Level	PU
WP contributing to the Deliverable	WP4
Contractual Date of Delivery	M20 (31/12/2016)
Finally approved by coordinator	19-05-2017
Actual Date of Delivery	19-05-2017
Authors	Emanuele Bellini, Paolo Nesi, Pierfrancesco Bellini
Email	Emanuele.bellini@unifi.it, paolo.nesi@unifi.it,
Affiliation	University of Florence
Contributors	Daniele Cenni (UNIFI)

funded by the Horizon 2020
Framework Programme of the European Union

PROJECT CONTEXT

Workpackage	WP4: Backend
Task	T4.3: Data integration and interoperability
Dependencies	This document is preparatory for the WP5 because of the methods and tools to integrate heterogeneous data to be used in the CRAMSS and RESOLUTE Dashboard

Contributors and Reviewers

Contributors	Reviewers
Paolo Nesi, Pierfrancesco Bellini, Davide Cenni (UNIFI-DISIT)	Roberto di Vincenzo (SWARCO)
	Jan Paul Leuteritz (Fraunhofer)

Version History

Version	Date	Authors	Sections Affected
1.0	16/02/2017	Emanuele Bellini	All
2.0	18/03/2017	Emanuele Bellini	All
3.0	22/03/2017	Emanuele Bellini, Paolo Nesi, Pierfrancesco Bellini, Daniele Cenni	All
4.0	15/04/2017	Emanuele Bellini	All
5.0	12/052017	Emanuele Bellini	All
9.0	19/05/2017	Paolo Nesi	All

Table of Content

Project Context 2
Contributors and Reviewers 2
Version History 2
Copyright Statement – Restricted Content 2
Executive Summary 5
1 Introduction 5
1.1 Relation with the project 8
2 Knowledge base 8
2.1 Data integration process 8
2.1.1 Data Mapping 10
2.1.2 Data Reconciliation 11
2.2 Ontological Model for data aggregation 12
2.2.1 Risk and resilience Ontology state of the art 12
2.2.2 Ontology implementation 13
2.3 Ontology extension for real time risk and resilience management 14
2.3.1 Methodology 14
2.3.2 Domain analysis 15
2.3.3 Potential damage modeling 15
2.3.4 Ontology extension implementation 17
2.4 Data model and tool improvement 21
2.4.1 Adding data model to Km4City 21
2.4.2 Adding data management into the ServiceMap 22
3 Semanting aware data processing and analysis 23
3.1 Knowledge driven flooding detection and damage estimation validation tool 23
3.2 Knowledge driven seismic detection and damage estimation validation tool 27
3.3 Human behaviour data analysis 28
3.4 Origin Destination Analysis for people flow 31
3.5 Understanding City Usage from AP data 32
3.6 AP Clustering experimental results 33
3.7 Predicting Access Point Connections 37
4 Interopreability via Smart City API 38
4.1 Accessing triage data through Smart City API 38
4.2 Accessing to OD derived from Wi-Fi data via Smart City API 41
4.3 Accessing to Wi-Fi data via Smart City API 41
5 Conclusions 43

REFERENCES 43
Appendix A: Glossary 46

List of Figures

Figure 1 Adaptive capacity diagram 6
Figure 2 System abilities and monitoring capacity 7
Figure 3 Data integration schema 9
Figure 4 -Risk Ontology extention 19
Figure 5 Risk Ontology details 20
Figure 6 Flooding test web app 24
Figure 7 Seismic test web app 27
Figure 8 OD matrix for Florence downtown: (a) classical view; (b) advanced proposed view 32
Figure 9 Typical AP trend in terms of number of connections along the 24h 32
Figure 10 Optimal number of AP clusters via Elbow criteria (comparing K-means and PAM):within sum of square function 34
Figure 11 Optimal K number of clusters via Gap curve (comparing K-means and PAM) 34
Figure 12 Average BIC for measure models vs K number of cluster, higher values are better, the curves are truncated at the best value for K they found. 35
Figure 13 Map of AP clusters: (a) Monday-Friday, (b) Saturday, (c) Sunday 36
Figure 14 The shapes of AP clusters with K=12 and EM clustering algorithm 37
Figure 15 APs time series with theri respective cluter ranges (see details in the text) 38
Figure 16 APs on the map and typical, realtime, guessed values for the datetime 42
Figure 17 Clustered APs time series: typical (blue), realtime (green), guessed (red). 43

EXECUTIVE SUMMARY

The present document reports the results of the work performed in the Task 4.3 focused on data integration and interoperability. The activities have been divided into the following sub-tasks: a) knowledge base development, b) semantic aware data processing and analysis, c) Data Interoperability.

In particular, a resilience-based ontology has been developed for data integration and fusion extending the Km4City a well known ontology used to manage big data generated by a smart city. The ontology has been developed with the support of domain experts and with the aim of accessing and elaborating data for real time application. In order to validate the ontology a damage modelling for flooding and seismic events has been developed to serve the pilot use case of Florence. The

1 INTRODUCTION

An Urban Transport System is a multi actor system that tends to be organised as a silos with some technological and organizational barriers that prevent to manage resilience effectively at city level. For instance, in Florence the yard planning is a responsibility assigned the municipality technical services dept. since it is responsible for the roads maintenance. These kinds of decisions have multiple implications and repercussions at the city level and to mitigate such impact all the affected stakeholders should be timely informed (e.g. mobility dept., citizens, business operators). For instance, the Mobility dept. needs to be aware about the status of the transport network to re-organize the viability in the area (e.g. changing the directions).

Such a communication ineffectiveness in this case between “maintenance” and “operation” ERMG functions increases the system variability that can lead to the system resonance and thus to a disaster. Hence, if the resources (information) needed for the “operation” function to operate as expected (e.g. to adjust the service and organise new mobility strategy) become available too late or not at all (high variability), the output of the function will be totally unreliable, contributing to propagate the variability in the system. This criticality could have limited impact in case nothing happens. On the other hand, in case of emergency the lack of information prevents the capacity of taking correct decisions in due time, ending up to suffer a disaster.

Another example is related to the traffic management where any traffic plan calculated in advance on the base of scenarios that could come from the “risk assessment” ERMG function, may end up to be useless since it is based on pre-determinate situation and the critical event could require a totally new coping strategy.

Moreover not knowing the real time status of the traffic, the real time position and direction of the people, etc. may lead to wrong on ineffective decisions because of the high level of uncertainty.

Usually, to reduce such uncertainty during an emergency, precious time is dedicated to gain a better understanding about the status of the UTS. An intensive (and sometimes not organised) information gathering from the operators on the ground (fire fighters, urban police), from citizens (number calls to emergency call centre), form control rooms of other institutions, etc. is activated. However the creation of reliable communication channels not previously established requires several steps to avoid misunderstanding that may have the opposite effect of increasing instead of reducing the uncertainty. Such steps are:
a) definition of the touch points (key person in an institution, on the ground, etc.)
b) definition of the communication channels to be used (email, mobile, fax, radio, etc.)
c) definition of a communication protocol in order to give the possibility to the operators to understand when the communication start or finish, which is the content, which is the context, etc. (should I wait a message to trigger an action?).

d) definition of the responsibility framework and the communication flow (Is my duty to inform other operators?)

Because of their complexity, criticality and time demand, it is evident that such steps cannot be accomplished with success during an emergency but should be executed during the preparation phase.

In the meantime a number of decisions are taken separately by each operator involved in the emergency (fire brigades, police, civil protection, etc.) causing overlaps in the interventions, over/under estimation of human resources and means engagement needs, conflicting decisions, etc. In this scenario the capacity to respond and recovery effectively is heavily affected.

It is evident that taking well-informed decision is crucial for UTS resilience management and requires to collect and share more information as possible to the operators acting in the city.

This action can be seen as related to the monitoring enhancement according to the adaptive capacity concept (Figure 1). In particular, as depicted in the D2.2, resilience is focused on sustaining the capacity of a system to adapt in the presence of continuous change. Generating, maintaining, and deploying adaptability processes relies upon the allocation of a wide range of resources and at many different system levels and time scales. As such, adaptability capacities are intrinsically related to the level of resources that a system can allocate and its ability to manage these resources in view of specific adaptive cycles.

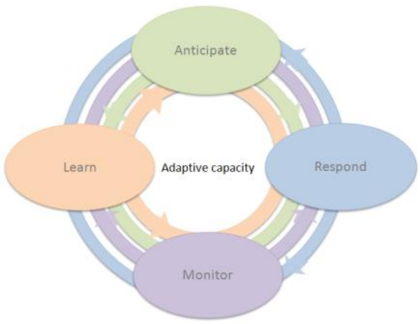


Figure 1 Adaptive capacity diagram

However, a capacity can be considered a precondition for the action. Similarly in the body motion study, every motor ability implies the subsistence of a motor capacity as an enabling condition through which a new ability can be trained and acquired. For instance the ability to control an artefact with the hands during a ballet (e.g. artistic gym) implies the existence of coordinative capacities (e.g. rhythm and equilibrium) for a gymnast. Thus to address UTS resilience both actions are necessary: a) building the adaptive capacity and b) enhancing its *ability* to *plan* for, *absorb*, *recover* from and *adapt* to adverse events **over time** (National Academy of Science definition). In fact these abilities can be exhibited only within a time frame and consequently they can be evaluated solely when a critical event happens. On the other hand, the adaptive capacities are something that requires to be built continuously and can be assessed against the ERMG. According to this perspective, the Data Integration and Interoperability results reported in the present document, aims to strengthen the system **monitoring capacity** in order to enhance the 4 abilities (plan, absorb, recovery and adapt) during the emergency. In Figure 2 is summarised in which extent such a monitoring capacity enhancement enables and improves the aforementioned 4 abilities.

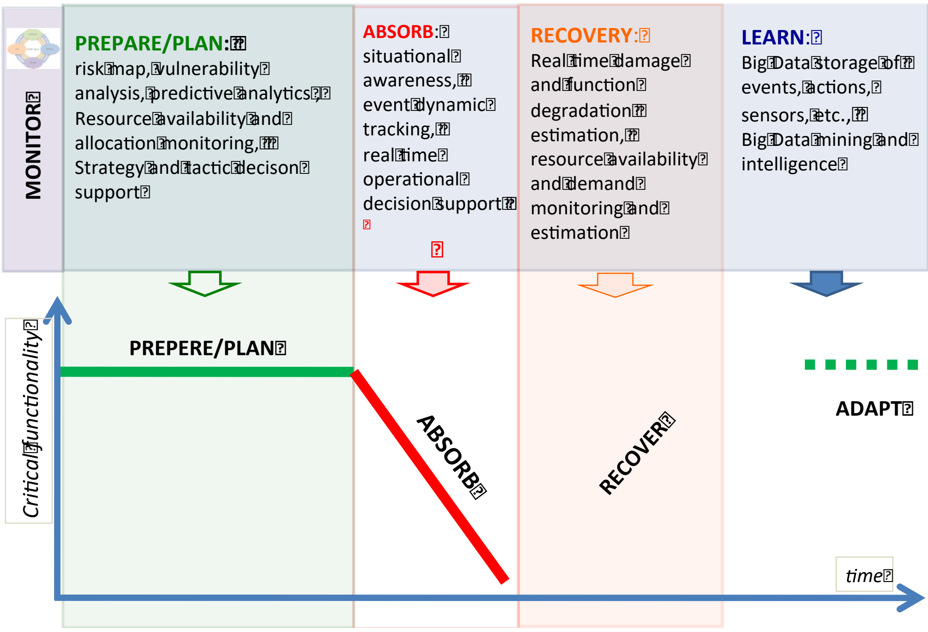


Figure 2 System abilities and monitoring capacity

To enhance the monitoring capacity 3 components has been developed: a) a knowledge base to manage and fuse Big Data, b) semantic processing techniques to analyse data exploiting their semantic relationships c) an interoperability layer to allow data access and consumption.

Knowledge base

In order to enhance and operationalize UTS resilience through data integration it is necessary to start to collect real heterogeneous (big) data from the smart city leveraging technologies like environmental sensors data (e.g. river level), whether forecast, Government Open Data (e.g. Risk maps), Social Networks data(e.g. twitter), Real time Public Transport System Data, City Wifi access, and so forth. A comprehensive review of the data needs and availability has been provided the D4.2. Such big data need to be integrated in a scalable and semantic aware knowledge base in order to exploit the potential of semantic relationships for data retrieval and processing.

The **KM4City** platform (Bellini, Nesi, Rauch, Benigni, & Billero, 2014) implemented in Florence, is an advanced and flexible Big Data semantic aggregator of data generated by a Smart City able to manage huge amounts of static and dynamic data streams and to provide consumption APIs for 3rd party exploitation.

This platform was used as a baseline for the RESOLUTE Data layer (See D4.1) and has been extended with new data and semantic concepts (Risk & resilience based ontology) to address the UTS resilience management requirements. The resulting data layer fuels the Collaborative Resilience Assessment and Management Support System (CRAMSS).

Semantic processing

The semantic reasoning algorithm has been used to generate new knowledge that is integrated into the Knowledge base. These algorithms and procedures can be grouped as:

• **data ingestion, mining and algorithms**: reconciliation, disambiguation, text to RDF/onto, verification and validation of ontology consistency and completeness, text mining, concept disambiguation, semantic enrichment, link discovering/interlinking, data processing (formatting conversion, filtering, stream processing,..), etc.

• **measures to reduce the effects of a disaster before it occurs and thus for activating computing processes such as:** predictions and detection of critical conditions through sensing sensors networks discovering people concentration and flows, emergency room overload, event magnitude (e.g. rainfall) real time monitoring, traffic flows, Public Transport system issues detection (e.g. delays), etc.

• **computing recommendations**, according to cross sector data and similarity distance with the collective profile (static and dynamic), context and condition (on the bus, in the car, raining, snowing, etc.), position in the city, etc.

Interoperability model

The RESOLUTE reference architecture shall facilitate integration, allowing acquisition of data produced by heterogeneous systems, and therefore shall be based on standard enabling technologies and practices, like service oriented and event driven architectures, to achieve flexibility and loosely coupling design. To this end communication from Data layer and CRAMSS is based on standard RESTfull API.

In section 2, an overview of the status of the Km4City data integration before the extension and improvement required by the RESOLUTE project. Section 3 depicts the high level ontological model from which the work starts in modelling the knowledge related to the RESOLUTE aspects, namely: wifi, risk and resilience, flooding, triage, underpasses, etc. Section 4 describe the work in modelling knowledge for risk assessment performed in RESOLUTE projects. As well as, Section 5 reports the activities of data modelling for adding information to the Km4City ontology and managing additional data as data flows from Wi-Fi access points for monitoring and predicting city users flows.

1.1 Relation with the project

This deliverable represents the outcome of the Task 4.3 that is strictly connected with the Pilot definition and execution (WP6). It is part of the backend implementation and integration and enables the frontend development (WP5).

In fact, heterogeneous data integration coming from the UTS and other related sources is an open issue that the project addresses. The objective is to enrich data streams with other data (streams) through semantic relations. This means that big data collection are governed according to a specific strategy. The result represents a relevant outcome of the project. The data integration has been driven by the result of the conceptual model (D2.2), the ERMG (D3.5 and D3.7) and the scenarios discussion (D7.1, D7.2) the data collections (D4.2) and the users requirements (D5.1 and D5.2).

2 KNOWLEDGE BASE

2.1 Data integration process

In general, all the smart city solutions must cope with big data volume, variety, and veracity (Bellini et al., 2013). Open data as static data are not the main source of information in the city. Most of the big data problems is related to their meanings.

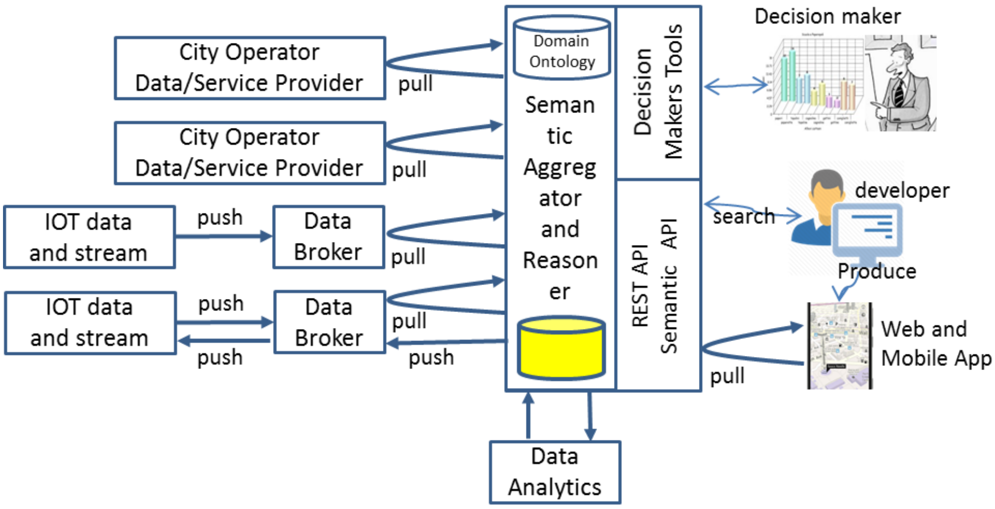


Figure 3 Data integration schema

The situation is depicted in Figure 3. A Semantic Aggregator and Reasoner collects data and services from the UTS and other operators, to aggregate and integrate them in a unified and semantically interoperable model based on a multi-domain ontology. This approach allows re-conciliating data and exploiting a coherent model to reduce the errors, integrating data representing the same concept and coming from different structures, operators, and sources. The ontology can be used by the semantic aggregator to model city domains entities and their relationships and not only metadata of data sets and tables as in non semantic aggregators. The usage of a multi-domain ontology allows the adoption of a model representing relationships of specialization among classes and relationships, aggregation, association, and similarity, that enable the inferential processes in the RDF Graph Database (Kotoulas et al., 2014), (Bellini et al., 2015), (Bellini P. et al., 2014b). Thus, the obtained knowledge base can be used for creating strategies for data quality improvement and for setting up algorithms and reasoning about the several aspects and services belonging to multiple domains (prediction, early warning, etc.). This advantage is also evident on the provided API and in the possibility of providing integrated data and views for the Decision Makers. For the same reason, the obtained Knowledge Base, by populating the ontology with data and inference, can be profitably and easily used for producing smart services such as contextual routing, multimodal contextual routing, suggestions on demand, personal assistants, connected drive, etc. Some solutions fit to this case: CitySDK [CitySDK] partially covering all features has been developed in an EC project involving major cities and providing specific REST API and grounded on OASC (Open & Agile Smart Cities) adopted the FIWARE NGSI API agnostic model [OASC]; and more widely covering features Km4City [Bellini et al., 2014b] exploited by Sii-Mobility Smart City project, providing Smarty City API of Km4City as [ServiceMap], [Bellini et al., 2014b], and SPUD proposed by IBM in [Kotoulas et al., 2014] exploiting commercial non open solution via a non-accessible ontology.

This kind of solutions need mapping tools from data to ontology and to support the reconciliation as performed by DataLift in [Scharffe et al., 2012] and by Km4City in [Bellini et al., 2014b]. In both cases, vocabularies, algorithms, and dedicated languages have been used, as SILK [Bizer et al., 2009]. It should be noted that the semantic aggregator could be implemented on top of solutions structured by using a non semantic aggregator. In fact, they may provide facilities listing data sets, updating their acquisition, and accessing to them with some tabular based API automatically generated, and they can be used to feed the semantic model.

The solution adopted is better ranked with respect to the sub-goals of the Urban Platforms [EIP Requirements] covering aspects connected to the harmonization of data, and production of intelligent services. Moreover, the implementation of user experience for value added services (subgoal 5 of EIP) is only accessible in a few of them as analysed in the following. The semantic based solutions have to cope with Graph Database collecting huge amount of data, thus resulting in Big Data cases and scenarios presenting relevant data such as variety, velocity, veracity, volume, etc. [Bellini et al., 2013], [Kotoulas et al., 2014].

An effective integration at semantic level of the data domain enables the creation of Smart Decision Support Systems that exploit the possibility of making semantic queries on multiple domains, to make probabilistic reasoning on Bayesian decision support [Bartolozzi et al., 2015], [Kotoulas et al., 2014], and to enable the production of algorithms for implementing personalized routing and Personal Assistants in the city. In some cases, the adoption of graph database to store and retrieve smart city data may be not the most effective solution in terms of performance despite it may enable reasoning as inference [Bellini et al., 2015b]. This issue may lead to decide of managing data in multiple data models according to their usage and natural matching between data structure and concepts. For example, semantic relationships among city entities on graph data stores, IOT and time varying data as sensors occurrences in efficient noSQL tabular stores such as CouchDB [1]or MongoDB[2].

2.1.1 Data Mapping

The Mapping Phase deals with the transport of information, previously saved into HBase database, into an RDF datastore, in our case managed by Owlim-SE[3]. The first part of this procedure retrieves information from HBase[4] to put them on a temporary MySQL database (required to use the Data Integration tool chosen), while in the second part data are translated into triples. Transformation is needed to map the traditional structured into RDF triples, based on information contained in a well-defined ontology (DISIT Ontology for Smart City) and all ontologies reused (dcterms[5], foaf[6], vCard[7], etc.). This process may be performed ad-hoc programs that have to take into account the mapping from linear model to RDF structures. This two steps process allowed us to test and validate several different solutions for mapping traditional information into RDF triples and ontology. The ontological model has been several times updated and thus the full RDF storage has been regenerated from scratch reloading the definition (all the other vocabularies, selecting the testing several different solutions) and the instance triples according to the new model under test. Once the model has been generated, RDF triples can be automatically inserted.

The first essential step is to specify semantic types of the data set, i.e., it is necessary to establish the relationship between the columns of the SQL tables and properties of ontology classes. The second step consists in defining the Object Properties among the classes, or the relationships between the classes of the ontology. When dataset has 2 columns that have the same semantic type but which correspond to different entities, thus multiple instances of the same class have to be defined, associate each column to the correct one.

The process responsible to perform the mapping transformation, passing from Hbase to SQL database has been produced as a corresponding ETL Kettle associated with each specific ingestion procedure for each data set. The second phase, of performing the mapping from SQL to RDF, has been realized by using a mapping model:Karma Data Integration tool, which generates a R2RML model, representing the mapping for transport from MySQL to RDF and then it is uploaded in a OWLIM-SE RDF Store instance. Karma initialization phase involves loading the primary reference ontology and connecting dataset containing the data to be mapped.

[1] http://couchdb.apache.org/
[2] https://www.mongodb.com/it
[3] http://bulgariana.eu/display/OWLIMv52/OWLIM-SE+Full-text+Search
[4] https://hbase.apache.org/
[5] http://dublincore.org/documents/dcmi-terms/
[6] http://xmlns.com/foaf/spec/
[7] https://tools.ietf.org/html/rfc6350

This process allowed the production of the knowledge base that may present a large set of problems due to inconsistencies and incompleteness that may be due to lack of relationships among different data sets, etc. For example to join services with the road map using the street address names that are written in different ways (e.g., "Via XXVII Aprile" and "VIA VENTISETTE APRILE") producing '*owl:sameAs*' triples to link them. These problems may lead to the impossibility of making deductions and reasoning on the knowledge base, and thus on reducing the effectiveness of the model constructed. These problems have to be solved by using **a reconciliation phase** via specific tools and the support of human supervisors.

2.1.2 Data Reconciliation

To connect services to the Street Guide in the repository a reconciliation phase in more steps, has been required, because the notation used by the Tuscany region in some Open Data within the Street Guide, does not always coincide with those used inside Open Data relating to different points of interest. In substance, different public administrations are publishing Open Data that are not semantically interoperable.

Furthermore, there are different types of inconsistencies within the various integrated dataset, such as:

- typos;
- missing street number, or replacement with values "0" or "SNC";
- Municipalities with no official name (e.g. Vicchio/Vicchio del Mugello);
- street names with strange characters (-, /, ° ? , Ang., ,);
- street numbers with strange characters (-, /, ° ? , Ang., , (, ,);
- road name with words in a different order from the usual (e.g. Via Petrarca Francesco, exchange of name and surname);
- red street numbers (in some cities, street numbers may have a colour. So that a street may have 4/Black and 4/Red, red is the numbering system 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 Papa Giovanni XXIII).

Thanks to the created ontology, is possible to perform services reconciliation at street number level, i.e. connecting an instance of class *Service* to an external access that uniquely identifies a street number on a road, or only at street-level, with less precision (lack that can be compensated thanks to geolocation of the service).

The methodology used in this reconciliation phase consists of first try to connect each service at street number-level, and then, perform the reconciliation at street-level.

The first reconciliation step performed consists of an exact search of the street name associated to each service integrated. For example, to reconciliate the service located at "VIA DELLA VIGNA NUOVA 40/R-42/R, FIRENZE", a SPARQL query is necessary, to search for all elements of *Road* class connected to the municipality of *"FIRENZE"* (via the ObjectProperty *inMunicipalityOf*), which have a name that exactly corresponds to "VIA DELLA VIGNA NUOVA" (checking both fields: official name, alternative name). The query result has to be filtered again, imposing that an instance of *StreetNumber* class exists and it corresponds to civic number "40" or "42", with the R class code Red.

From this first reconciliation step, the services for which was identified a single instance of the class *Entry* has been selected, and the related reconciliation triples at street number- level, have been created.

A very frequent problem for exact search, is the existence of multiple ways to express toponym qualifiers, that is dug (e.g. Piazza and P.zza) or parts of the proper name of the street (such as Santa, or S. or S or S.ta): thanks to support tables, inside which the possible change of notation for each individual case identified are inserted, a second reconciliation step was performed, based on exact search of the street name, which has allowed to increase the number of reconciled services at street number-level.

The third reconciliation step is based on the research of the last word inside the field *v:Street-Address* of each instance of the *Service* class, because, statistically, for a high percentage of street names, this word is the key to uniquely identify a match.

These first three reconciliation steps have been also carried out without taking into account the street number, and so in order to obtain a reconciliation at street-level of each individual service.

The fourth reconciliation step is realizing thanks to Google Geocoding API[8], through which different services, not yet connected to the *Street Guide* macroclass at street number-level, were searched again.

The next reconciliation step used automated methods to remove strange characters, inside the street number field, or the address field, but unfortunately at this point it is becoming increasingly difficult to obtain unique results in the search for correspondences between instances of the class *Entry* and instances of the class *Service*.

The last reconciliation step implemented, trying to reconcile all those services in which the name of the town is incorrectly used or it is expressed in a not official notation; even in this case it is difficult to get great results from every single reconciliation step.

At present, all services that present typos, street number equal to "0" or to string "SNC", still need to be managed; moreover services with strange char in the street name, are partially managed.

2.2 Ontological Model for data aggregation

2.2.1 Risk and resilience Ontology state of the art

Differently from other domains, the ontologies and the semantic approch are basically underused in the risk and resilience management.

Within the sciences there is the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Gangemi et al, 2002) and the Semantic Web for Earth and Environmental Terminology SWEET, an "upper-level" ontology that collects thousands of concepts describing natural phenomena (physical) properties, unit measurements, human activities and data features (Raskin and Pan, 2005). There are also other ontologies of domain in the field of Earth sciences which have the purpose of facilitating the understanding and sharing of knowledge like Chau (2007) has proposed a knowledge model for water flow and water quality management; Tripathie and Babaie (2008) extended the SWEET ontology by adding concepts and relationships in hydrogeology, especially in the domain of groundwater circulation. Concerning the concept of risk it is worth to mention the MONITOR domain ontology of Kollarits and Wergles, (2008). This ontology collects risk concepts and risk analysis and management and proposes a glossary of terms and concepts related to it. In this ontology a conceptual model derived from the definitions of the terms inherent in the domain of risk is defined. MONITOR is strongly based on the DOLCE ontology. In fact all the concepts described have their equivalent in DOLCE. MONITOR and SWEET are in any case the ontologies that can be considered the state of the art, however, they are not desigined to manage heterogeneous entities, big data stream or to support advanced semantic processing. Thus they are valuable for a systematization of the concepts in the domain but to be operational, efficent and more focused on critical infrastrucutre, a new ontology is needed. Indeed most of the concepts used

[8] https://developers.google.com/maps/documentation/geocoding/

for this work derive from the work published in the risk and resilience domain.

2.2.2 Ontology implementation

In order to create an ontology to support UTS resilience, a large number of data sets have been analysed to see in detail each single data elements of each single data set with the aim of modelling and establishing the needed relationships among element, thus making a general data set semantically interoperable (e.g., associating the street names with toponimous coding, resolving ambiguities). The work performed started from the data sets available in the Florence area up to Tuscany region and then it is continued to include Athens datasets. In total the whole data sets are more than 800 data sets. At regional level, Tuscany Region also provided a set of open data into the MIIC (Mobility Integration Information Center of the Tuscany Region), and provide also integrated and detailed geographic information reporting each single street in Tuscany (about 137,745), and the location of a large part of civic numbers, for a total of 1,432,223 (a wider integration could be performed integrating also Google maps and Yellow/white pages). From the MIIC it is possible to recover information regarding the data identified in the D4.2 as streets, car parks, traffic flow, bus timeline, etc. While from Florence municipality real time data about the RTZ, tram lines on the maps, bus stops, bus tickets, accidents, ordinances and resolutions, numbers of arrivals in the city, number of vehicles per year, etc. From the other open data points of interest can be recovered as position and information related to: museums, monuments, theatres, libraries, banks, express couriers, police, firefighters, restaurants, pubs, bars, pharmacies, airports, schools, universities, sports facilities, hospitals, emergency rooms, government offices, hotels and many other categories, including weather forecast by Lamma consortium. In addition to these data sets, those coming from the mobility and transport operators have been collected as well.

The analysis of the above mentioned data sets allowed us to create an integrated ontological model presenting 6 main areas of macroclasses as depicted in Figure 2.

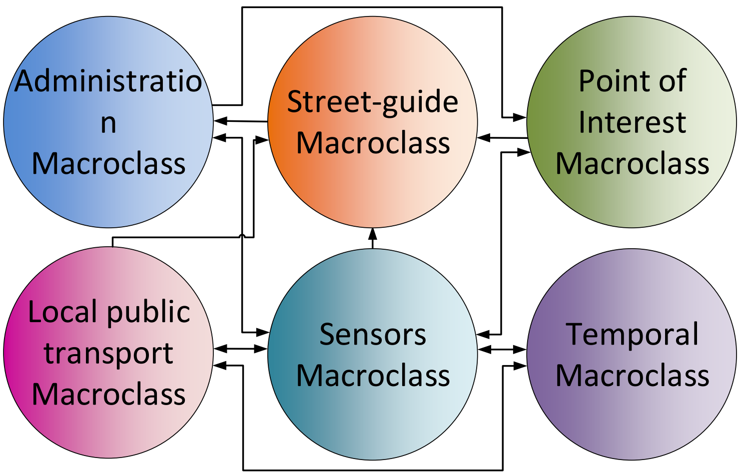


Figure 2 - Ontology Macro-Classes and their connections

Administration: includes classes related to the structuring of the general public administrations, namely PA, and its specifications, Municipality, Province and Region; also includes the class Resolution, which represents the ordinance resolutions issued by each administration that may change the traffic stream.

Street-guide: formed by entities as Road, Node, RoadElement, AdministrativeRoad, Milestone, StreetNumber, RoadLink, Junction, Entry, and EntryRule Maneuver, it is used to represent the entire road system of Tuscany, including the permitted manoeuvres and the rules of access to the limited traffic zones. The street model is very complex since it may model from single streets to areas, different kinds of crosses and superhighways, etc. In this case, OTN vocabulary has been exploited to model traffic [4] that is more or less a direct encoding of GDF (Geographic Data Files) in OWL.

Point 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. The classification of individual services and activities is based on main and secondary categories planned at regional level. In addition, this macro segment of the ontology may take advantage of reusing Good Relation model of the commercial offers[9].

Local public transport**:** includes the data related to major TPL (Transport Public Local) companies scheduled times, the rail graph, and data relating to real time passage at bus stops. Therefore this macroclass is formed by classes TPLLine, Ride, Route, AVMRecord, RouteSection, BusStopForeast, Lot, BusStop, RouteLink, TPLJunction.

Sensors: macroclass concerns data from sensors: ambient, weather, traffic flow, pollution, etc. Currently, data collected by various sensors installed along some streets of Florence and surrounding areas, and those relating to free places in the main car parks of the region, have been integrated in the ontology.

Temporal: macroclass that puts concepts related to time (time intervals and instants) into the ontology, so that associate a timeline to the events recorded and is possible to make forecasts. It may take advantage from time ontologies such as OWL-Time.

This structure has been extended with an ontology dedicated to model risk and resilience analysis that is introduced in the chapter 4.

2.3 Ontology extension for real time risk and resilience management

The Km4City ontology has been created with the objective of integrating heterogeneous data coming from the urban context for Smart City monitoring, thus a conceptualization able to manage risk and resilience missed. The work started identifying the concepts at the base of the vulnerability and risk analysis because it represents a well grounded domain. The objective was to design a model able to support real time computing (queries) and to extract information related to the amount of the damage occurring in a given urban area during an event whose magnitude is measured by georeferenced sensors. The impact is calculated considering multiple exposers (services, assets, people, etc.) and their status (open/close, presence/absence, etc.) at the time of the event.

The scope is to gain a better situational awareness during the emergency, track the event but also to support a better understanding of UTS and city vulnerabilities and the consequent informed decisions to cope with unexpected changing conditions.

2.3.1 Methodology

To formalize domain knowledge in an ontology it is crucial to work with domain expert following a co-design approach. To this end the project has engaged relevant experts since the beginning as:

- **Safety and security operators:** (e.g. Civil Protection, Urban Police) They contributed with scenarios, past and current experiences (work as done), information needs and system improvements.
- **Critical infrastructure managers**: Florence Mobility Department, Attiko metro
- **Environmental Risk and resilience experts: (UNIFI-DST)** Their role is related in particular to vulnerability analysis and asset management
- **IT experts**: of the system that stores data (RDF Store) and the ontology structure that contains the data.

[9] http://www.heppnetz.de/projects/goodrelations/

The work has been carried out through the following steps:

- Studying of environmental risks with the support of domain experts with a focus on flooding and seismic;
- Modelling the seismic and flooding cases to be used as a test case for the verification of the correctness of the ontology;
- Modelling an ontology able to describe and manage risks and the real-time damage assessment;
- Extension of KM4CITY with the new ontology
- Instantiation of the new ontology with the data coming from the RESOLUTE data layer;

2.3.2 Domain analysis

It is well known that the product of vulnerability for the value the element exposed allows to derive the potential damage expected in relation to a critical event with a certain intensity/magnitude. The potential damage depends therefore by:

- asset vulnerability exposed to the event;
- asset value exposed to the event.

According to the D2.2 Conceptual model, the Assets considered are: infrastructure, people, and organization. The vulnerability depends on the type of asset, the zone in which is located, its characteristics (materials, design), its functions (e.g. type of usage) and the intensity of the event. The value of the asset depends by the characteristics, functions and can assume an economic and/or social value in relation to the service provided to the community.

From this analysis, the importance of defining a generic and comprehensive data model able to accommodate different kind of information to calculate the damage in a multi-dimensional way, come out from the discussion with the experts. The ontology created allows the modelling of the vulnerability in such a way to obtain a table of vulnerabilities for:

• type of asset;
• geographical area;
• asset characteristics;
• intensity of the exposed phenomenon.

An accurate description of the asset in terms of characteristics, positions, etc. allows a better vulnerability analysis and a more precise estimation of the value of the exposures.

2.3.3 Potential damage modeling

The potential damage estimation in case of seismic event allows to define which are the zones at risk of damage according to the following characteristics:

- geological feature of the territory on which there are the services;
- built characteristics on which the services are connected;
- value associated to the type of service offered to the community.

The potential damage estimation is calculated in the following way:

1. Identification of the services within the area affected by the event;
2. For each service identified, the vulnerability (based on geological and built characteristics) value related to the intensity of the phenomena is extracted;
3. Calculate the value of the services

2.3.3.1 Seismic damage modeling

The damage caused by seismic event is the degree of expected loss as a result of a seismic event. An analytical model was defined for the damage caused by the occurrence of earthquakes. This estimation of the damage model is decomposed through the analysis of the following components:

- Starting from the map of seismic hazard, the town of Florence has been classified in different seismic hazard regions. This classification is based on the maximum horizontal seismic acceleration factor of hard soil or flat. From this classification experts have drawn up tables able to provide a value of vulnerability between 0 and 1 for each of the kinds of services counted in the ontology KM4City, given as input a seismic magnitude measured by the sensor;
- Three macro-regions have been defined within the territory of Florence, each of them representing an area characterized by a homogeneously built (e.g. buildings of 14th century). For each region, it has been defined seismic vulnerabilities related to the magnitude of the event measured.
- For each type of surveyed service in KM4City (related to the Florentine territory), have been assigned a value of importance/criticality. .

Thanks to these data, the ontology is able to support the realtime and simulated damage estimation given a geographic area and a magnitude of the seismic phenomenon. In particular, two activities are considered a) the services situated in the affected area and b) the building typology (age of building) according to the zone classification provided by the heart science dept of the University of Florence.

2.3.3.2 Flooding damage modeling

Whit the term “flooding” we mean the presence of water accumulated in a normally dry place, following the formation of streams or of an external accumulation of water due to natural phenomena and not artificial (such as the breaking of a water system).

In order to obtain a measure of the damage, or the degree of expected loss, it is necessary to exploit the data generated by a rain gauge sensor. The rain sensor is a sensor capable of providing a measurement of the amount of rainfall in a span of time. It is often integrated with a weather station. A model for the evaluation of the potential damage due to heavy rainfall precipitation has been defined taking into account the following data:

- values of rainfall intensity fall in a time in mm / h, in relation to a scale of values of hazard from 0 to 1;
- vulnerability refers to the road elements, based on their structural classification;
- value of the elements involved in risk (roads, services, people, physical assets);
- observations of traffic present on the road network;
- through this data model, the occurrence of a rain event with intensity I in the geographical zone established, we assess the potential damage by taking into account:
- road elements characterized by their vulnerability and economic value according to the structural characteristics;
- the services characterized by a value linked to the type of service. A high value identifying a high importance index for the population;
- road traffic can make a growth factor to the damage caused by flooding estimated potential. A high traffic involves high number of cars, and people, and implies a greater difficulty of intervention in order to secure the exposed assets.

The security operator can then make decisions regarding the priorities for action in order to safeguard the security of goods and people involved, while simultaneously studying such areas have greater potential harm to the occurrence of an event in order to implement preventive policies.

2.3.4 Ontology extension implementation

According to the semantic web reusability principle, the developed ontology uses the following classes present in KM4City:
• **Observation** corresponding to class modelling sensor observations;
• **Service**, which catalogues the surveyed services km4city, which in turn is divided into:

- **DigitalLocation** that identify sites of particular interest for the city, as such as jogging trails, parks and gardens, museums and others;
- **Area** that brings together the services represented by an area on the map;
- **Path** that collects information of specific services;
- **RegularService** where regular services are catalogued;
- **TransversalService** that collects the transverse services connected to the regular ones.

The GIS ontology[10] has been also used for the Geometry class, useful to catalogue areas and points. For the DP domain ontology was created the following classes:

- **PluviometricConcentration** subclass of km4city:Observation, created to catalogue instances of rainfall observations received from the sensors;
  The class has a data property **FloodDetection** and a data type: float, that represent the value measured in mm/h.

- **EarthQuakeObservation** subclass of km4c:Observation, created to catalogue instances of seismic observations received from the sensors;
  The class presents a data property **LandMovement** and its data type float, that represents the magnitude of the seismic event.

- **AssetValue** shaping the values of the catalogued assets;
  - **PhysicAssetValue** used to model economic values / importance for the population related to business physical asset;
  - **ServiceAssetValue** used to model economic values / importance for the population related to the type of service;
  - **SocialAssetValue** to model the value and / or the social importance related to asset / service. This class has been added for future development, when they are available data capable of enhancing the social aspect of the asset;

  The class, as its sub-classes, presents a data property **hasValue** and a data type float that represent the associated value. This class is in relationship with **Asset** through the **hasAssetValue** property.

- **Vulnerability** to model the vulnerability values of assets;
  - **PluviometricVulnerability** subclass of vulnerability, the model of type rainfall linked to damage from flooding;
  - **SeismicVulnerability** as for the previous one, but this time for the seismic damage;

  The vulnerability values are characterized by the following features:
  - apply within a geographical area defined by the report **gis: hasGeometry**;

[10] http://www.opengeospatial.org/standards/geosparql

- you can be linked to an asset by **hasVulnerability** report;
- are linked to a type of marker by **hasIndicator** relationship.

The data properties of this class are:

- **shortDescription** and long description be saved in a description of the type of census vulnerabilities;
- **hasAreaClassification** census for the identification of vulnerability classification;
- **forAssetType** to link the value of vulnerability to a type of asset;
- f**orAssetProperty** to connect the vulnerability to a characteristic of a type of asset;
- **fromMinIntensity** in which I define the minimum value of the measured intensity range for which this vulnerability could be considered;
- **toMaxIntensity** for corresponding maximum value;
- **hasValue** property or the date for the census of the value of V.

Through this data modelling is possible to define for each geographic zone, a vulnerability table for each type of asset.

Range	Building 1 floor	Building 2 floors	Building 3 floors	Building 4 floors
[0:9.99]	0.5	0.4	0.2	0.1
[10:19.99]	0.7	0.5	0.3	0.1
[20:29.99]	0.8	0.6	0.5	0.2
…	…	…	…	…

- **Indicator** This class models a sensor indicator, and offers insight non what observes the sensor;
    - **PluviometricIndicator** to model the rainfall indicators;
    - **SeismicIndicator** for seismic sensors;
    - **TrafficIndicator** to model the indicators of traffic observations;

    This class has the data property **hasUOM**

It has been also added two relations between the km4city class: **SensorSite** and **gis:Geometry:**

- **Monitor**: connects a sensor with a geographic area. In this way it is possible define designated areas where track damage with the sensor observations.
- **isMonitoredBy**: is the inverse relationship than the one defined in point previous one.

Through these two reports it is possible to create monitoring scenarios associating to each sensor one or more geographic regions. The sensor observations will be used to estimate the damage occurred to the assets within these regions.

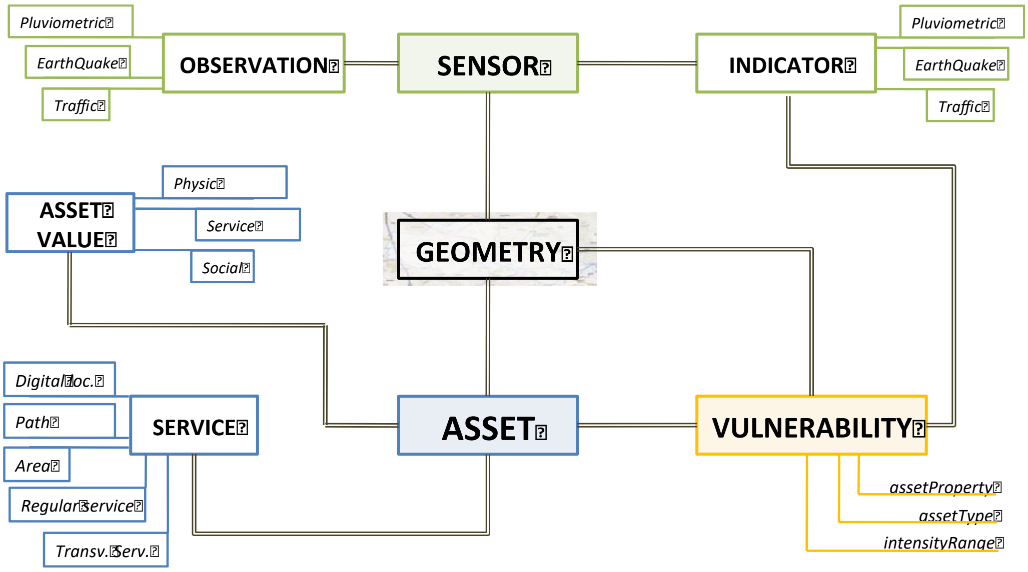


Figure 4 -Risk Ontology extension

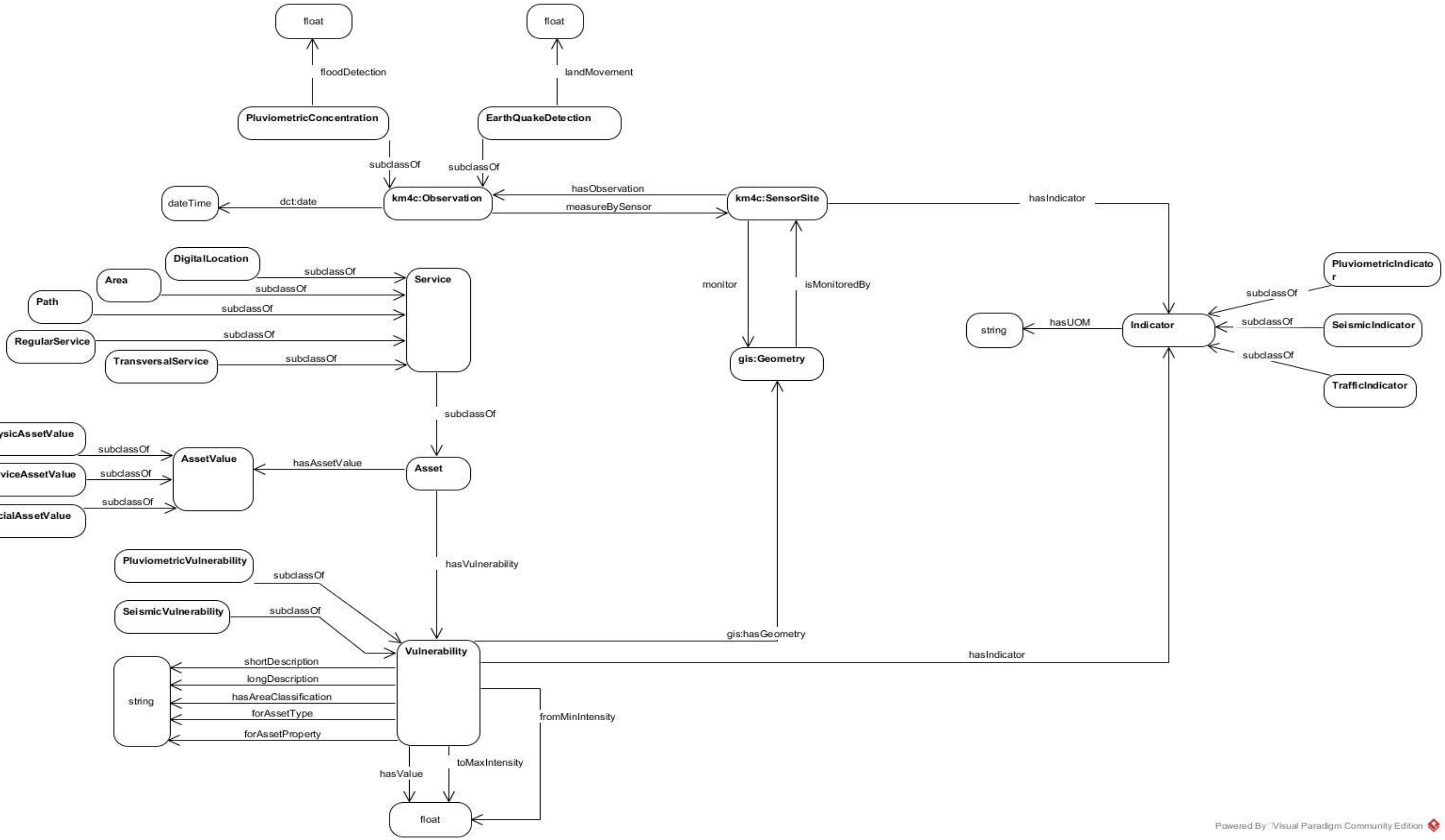


Figure 5 Risk Ontology details

2.4 Data model and tool improvement

This section reports the activities of data modelling for adding information to the Km4City ontology according to the D4.2 and managing additional data as data flows from Wi-Fi access points for monitoring and predicting city users flows.

2.4.1 Adding data model to Km4City

In this section we describe how the Km4City model and tools have been enriched with more data to cope with the information needed to RESOLUTE projects and in particular to manage:

- Hospitalization Structures of Population
- Triage information from hospital to have the status of the hospitals in the area in real time, especially for the first aid sector
- Data from the schools: namely the number of children that should be present in the primary schools in the Florence area
- Areas of Population Assistance: position and shape if any
- Areas of Cartographical Rescuers and Resources: position and shape if any, see previous section
- The waiting areas of the population: position and shape if any

Some ETL processes have been developed to acquire the information from the open data sources. The ETL processes get the data in the specific format, save it and checks if it is different from the one currently available (if present), if it is new the information is acquired and saved on storage and on a shared H-Base instance. The data is also mapped to the km4c ontology and other well-known ontologies (schema.org, dcterms, skos, foaf) to produce the triples to be loaded on the knowledge base.

The triage data management has been reported here as a example of reference.

The triage the data is acquired from three web pages:

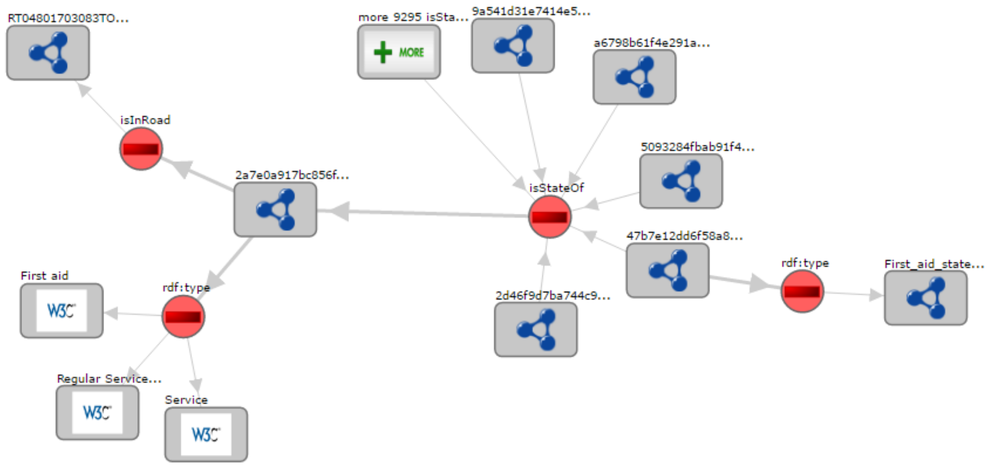
http://www.aou-careggi.toscana.it/internet/index.php?option=com_content&view=article&id=865&lang=it
http://www.ao-pisa.toscana.it/triageWeb/triage.php
http://www.asf.toscana.it/estar/accessi-internet.php

The triage data is acquired regularly using a scheduler that every 10min, the data is processed using an ETL transformation, it saves the data and produces the triples in a specific folder. The frontend machines (having a copy of the whole KB) regularly check if new files with triples are present and in the positive case upload them on the virtuoso 7.2 RDF store.

For the production of the triples files of the data acquired the km4city ontology has been enhanced adding some new classes and properties:

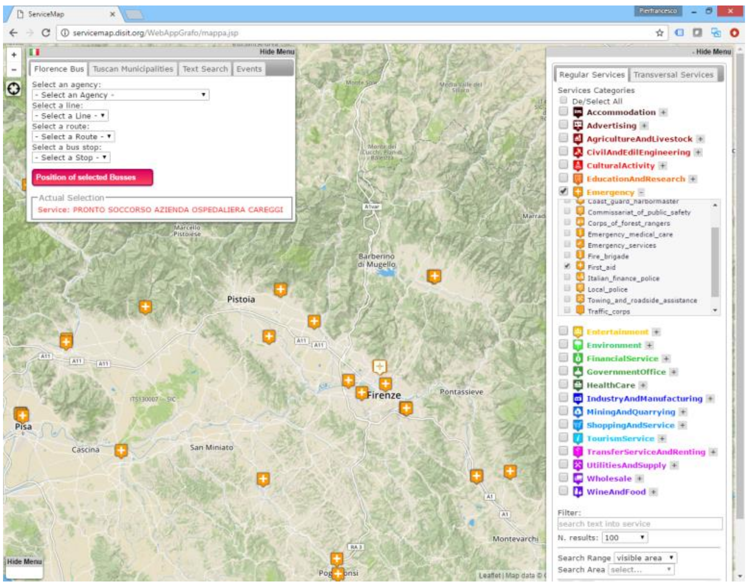
- class **First_aid_state** to describe the state of the triage for a specific First aid service
- properties *isStateOf* and *hasState* to join in a relation the first aid and its state
- data property *firstAidState* with the description of the state (“in osservazione”, “in attesa”, etc.)
- data properties *redCodeCount*, *yellowCodeCount*, *greenCodeCount*, *blueCodeCount*, *whiteCodeCount* with the number of patients with a specific triage colour code and in a specific state.
- the *First_aid_state* uses a dcterm:date property to state the date when the particular state was retrieved.

The relations among the First_aid service and its state can be depicted and analysed using the Linked Open Graph tool available at http://log.disit.org and reported in the following figure where on the left it is present a First_aid service and on the right the different states that are associated with it:



2.4.2 Adding data management into the ServiceMap

The ServiceMap tool available at http://servicemap.disit.org uses the KB where the information on the triage has been added. In particular it is possible to search for the “First_aid” services:



and clicking on a pin it is possible to see the current status of the triage:

On the pin the number of patients in the distinct states and with a specific colour code (red, yellow, green, blue and white) are reported. The possible states are: with destination, waiting, under visit and under temporary observation.

3 SEMANTING AWARE DATA PROCESSING AND ANALYSIS

To validate the Km4city risk and resilience management extension, two dedicated web applications have been developed. These applications were not intended to be included in the resolute layout, on the other hand, the information that they extract from the Knowledge Base (KB) to calculate the damage has been considered useful by the stakeholders even if they are presented in demo mode. Thus a possible evolution of the web app validation tools will be considered for the future.
Moreover, the possibility of accommodate Big Data according to a comprehensive ontology capable to define semantic relationship among the data themselves allows advanced analysis in particular on human behaviour. Dedicated algorithm has been developed to extract valuable insight from the KB to enhance early detection and prediction capability in the monitoring function.

3.1 Knowledge driven flooding detection and damage estimation validation tool

The first one is devoted to test the flooding event. In particular the application is able to retrieve the value of the

pluviometric sensors within a given area and calculate the total potential damage of the all exposed elements considering: a) the physical infrastructures, b) the services, c) the status of the traffic in a specific moment, d) the people on the ground, etc. In the example represented here below are reporter the streets affected with the severity represented by the colour of the row, the relevant services affected, the potential damage estimation with and without the traffic information included and the percentage of loss (68%).
This estimation has been done using SPARQL queries capable to exploit the semantic relationship provided in the ontology.

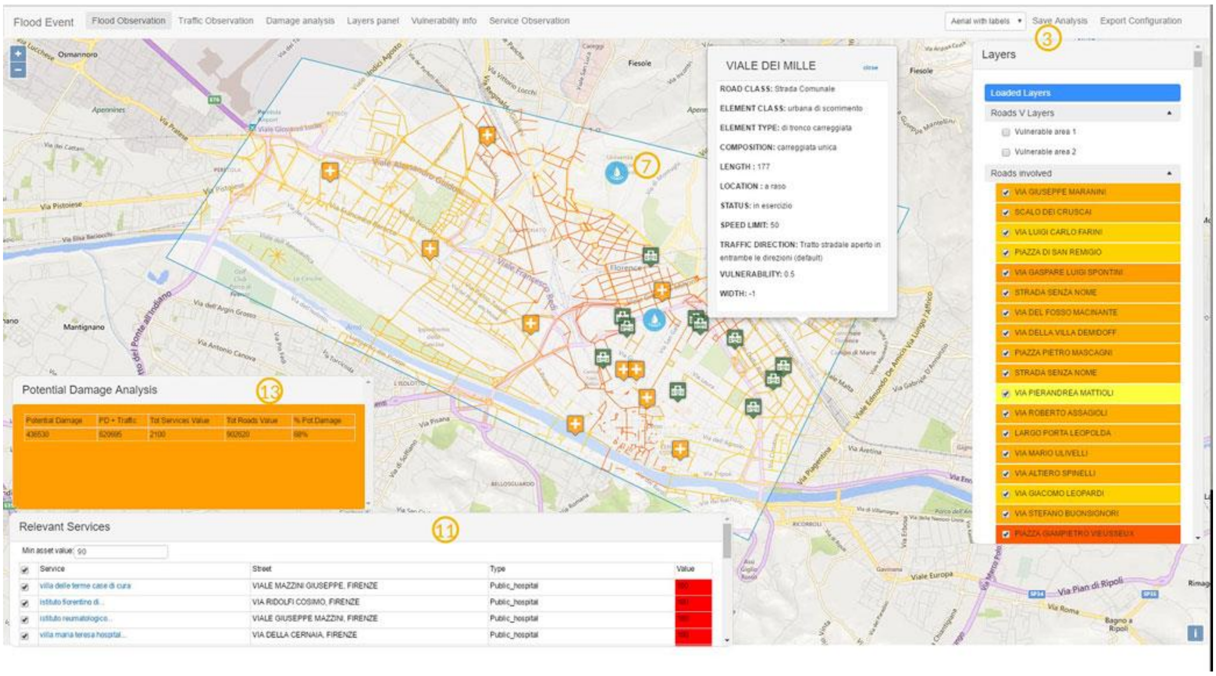


Figure 6 Flooding test web app

Here below is reported the SPARQL queries used to extract and processing the information displayed on the screen (Figure 6):

PREFIX owl: http://www.w3.org/2002/07/owl#
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX km4c: http://www.disit.org/km4city/schema#
PREFIX rrel: http://www.disit.org/risk_ontology/relation/schema#
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX gis: <http://www.opengis.net/ont/geosparql#>
PREFIX w3:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

The following query retrieves the observations for the pluviometric sensors

SELECT * WHERE{
?sensorSite rrel:hasIndicator
<http://www.disit.org/risk_ontology/resource/PluviometricFloodIndicator/id#INDICATOR>
.
?sensorSite rrel:managingBy ?managingBy.
?sensorSite geo:lat ?lat .
?sensorSite geo:long ?long .
?sensorSite <http://schema.org/addressLocality> ?city .
?sensorSite <http://schema.org/streetAddress> ?address .
?sensorSite rrel:hasDescription ?description .

```
?sensorSite rrel:hasObservation ?obs .
?obs dct:date ?dt .
?obs rrel:floodDetection ?floodDetection .
FILTER (
bif:st_within(bif:st_point(?long,?lat),bif:st_geomfromtext('<WKT_Polygon>'))=1).
FILTER(?floodDetection > 0.0).
FILTER(?dt > "<date>"^^xsd:date).}
```

The following query retrieves the information related to the services and their related value within the affected areas

```
SELECT DISTINCT * WHERE {
?service w3:type ?type .
?service geo:lat ?lat .
?service geo:long ?long .
?service <http://schema.org/addressLocality> ?city .
?service <http://schema.org/name> ?name .
?service <http://schema.org/streetAddress> ?address .
?serviceV w3:type
<http://www.disit.org/risk_ontology/risk_ontology/schema#ServiceAssetValue>.
?serviceV rrel:assetValueGroup ?type .
?serviceV rrel:assetValueClass ?classType .
?serviceV rrel:hasValue ?value.
FILTER(?value >= 90).
FILTER (
bif:st_within(bif:st_point(xsd:float(?long),xsd:float(?lat)),bif:st_geomfromtext('WKTPolygon'))=1).
}ORDER BY DESC(?value)
```

The following query retrieves the flooding vulnerability layer used for the damage estimation

```
SELECT * WHERE{
?layer rrel:hasAreaClassification ?classification.
?layer rrel:shortDescription ?description.
?layer rrel:fromMinIntensity ?minI.
?layer rrel:toMaxIntensity ?maxI.
?layer rrel:hasValue ?vulnerability.
?layer rrel:forAssetType "Roads".
?layer rrel:hasIndicator
<http://www.disit.org/risk_ontology/resource/PluviometricIndicator/id#INDICATOR>.
?layer w3:type <http://www.disit.org/risk_ontology/schema#VOfRouteElement>.
?layer rrel:hasGeometry> ?geo.
?geo gis:asWKT ?wktGeo.
}
```

The following query retrieve the traffic observations within a given areas provided in WKT format

```
SELECT  *WHERE {
?sensorSite rrel:managingBy ?managingBy.
?sensorSite dct:identifier ?identifier.
?sensorSite geo:lat ?lat .
?sensorSite geo:long ?long .
?sensorSite <http://schema.org/addressLocality> ?city .
?sensorSite <http://schema.org/streetAddress> ?address .
?sensorSite km4c:placedOnRoad ?road.
?road km4c:containsElement ?roadElement.
?roadElement km4c:elementClass ?roadClass.
?sensorSite km4c:hasObservation ?obs .
?obs dct:date ?dt .
?obs km4c:averageSpeed ?averageSpeed.
?obs km4c:concentration ?concentration.
?obs km4c:vehicleFlow ?vehicleFlow.
```

```
?obs dct:identifier ?idObs.
?obs <http://purl.org/dc/terms/date> ?mtime.
FILTER (
bif:st_within(bif:st_point(xsd:float(?long),xsd:float(?lat)),bif:st_geomfromtext('<WKTPolygon>'))=1).
{FILTER(?averageSpeed>0).}UNION
{FILTER(?concentration!='NAN').}UNION
{FILTER(?vehicleFlow>0)}}ORDER BY DESC(?mtime)
```

The following query retrieves the streets elements situated within a given area provided in WKT format

```
SELECT * WHERE {
?route <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?roadType.
?route <http://purl.org/dc/terms/identifier> ?routeId.
?route km4c:adRoadName ?routeName.
?route km4c:adminClass ?routeClass.
?route km4c:ownerAuthority ?roadAuth.
?route km4c:hasRoadElement ?elem.
?elem km4c:startsAtNode ?starts .
?elem w3:type ?elemSType.
?elem <http://purl.org/dc/terms/identifier> ?elemId.
?elem km4c:composition ?elemCompo.
?elem km4c:elemLocation ?elemLoc.
?elem km4c:elementClass ?elemClass.
?elem km4c:elementType ?elemType.
?elem km4c:endsAtNode ?endsAt.
?elem km4c:formingAdminRoad ?formingAdminRoad.
?elem km4c:length ?length.
?elem km4c:managingAuthority ?managingAuthority.
?elem km4c:operatingStatus ?operatingStatus.
?elem km4c:speedLimit ?speedLimit.
?elem km4c:startsAtNode ?startsAtNode.
?elem km4c:trafficDir ?trafficDir.
?elem km4c:width ?width.
?elemValCompo rrel:composition ?elemCompo .
?elemValCompo rrel:hasValue ?valCompo.
?elemValLoc rrel:elemLocation ?elemLoc.
?elemValLoc rrel:hasValue ?valLoc.
?elemValClass rrel:elementClass ?elemClass.
?elemValClass rrel:hasValue ?valClass.
?elemValWidth rrel:width ?width.
?elemValWidth rrel:hasValue ?valWidth.
?starts geo:long ?longStart.
?starts geo:lat ?latStart.
?starts <http://purl.org/dc/terms/identifier> ?idStart.
?starts <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.disit.org/km4city/schema#Node>.
?starts km4c:nodeType ?typeStart.
?elem km4c:endsAtNode ?stop.
?stop geo:long ?longStop.
?stop geo:lat ?latStop.
?stop <http://purl.org/dc/terms/identifier> ?idStop.
?stop w3:type <http://www.disit.org/km4city/schema#Node>.
?stop km4c:nodeType ?typeStop.
{ FILTER (
bif:st_within(bif:st_point(?longStart,?latStart),bif:st_geomfromtext('<WKTPolygon>'))=1)}
UNION
44 { FILTER (
bif:st_within(bif:st_point(?longStop,?latStop),bif:st_geomfromtext('<WKTPolygon>'))=1)}}
```

3.2 Knowledge driven seismic detection and damage estimation validation tool

This web application allows to select a seismic sensor and a specific region and according to the vulnerabilities and the value assigned of the exposed elements, it perform a calculation.

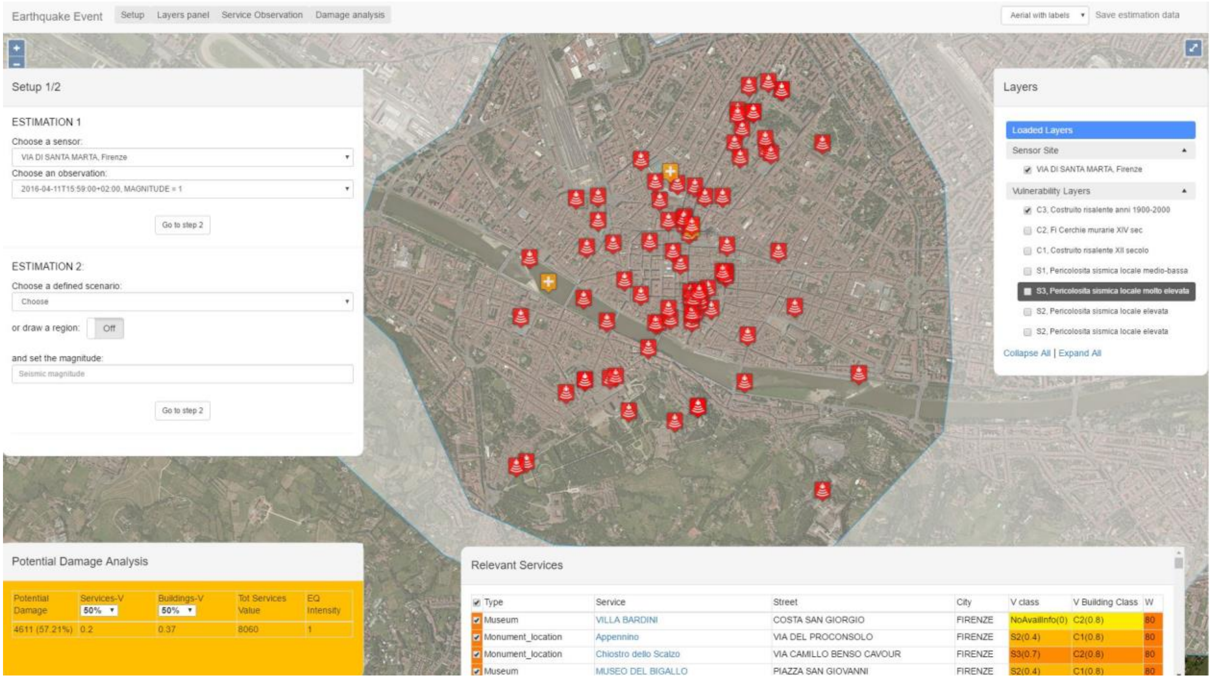


Figure 7 Seismic test web app

```
SELECT * WHERE{
?sensorSite rrel:hasIndicator
<http://www.disit.org/risk_ontology/resource/SeismicIndicator/id#INDICATOR> .
?sensorSite km4c:managingBy ?managingBy.
?sensorSite geo:lat ?lat .
?sensorSite geo:long ?long .
?sensorSite <http://schema.org/addressLocality> ?city .
?sensorSite <http://schema.org/streetAddress> ?address .
?sensorSite rrel:monitor ?geo .
?sensorSite rrel:hasIndicator ?indicator.
?geo gis:asWKT ?wkt.
?sensorSite rrel:hasObservation ?obs .
?obs dct:date ?dt .
?obs rrel:landMovement ?lMDetection .
FILTER(?dt > "<fromDate>"^^xsd:date).
FILTER(?dt < "<toDate>"^^xsd:date).}
```

The following query extract the services in the selected area:

```
}
UNION
{
?service rdf:type <type>.
?sValRes <http://www.disit.org/risk_ontology/relation/schema#assetValueClass>
?classType .
?sValRes rrel:assetValueGroup ?type.
?service <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat .
```

```
?service <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long .
?service <http://schema.org/addressLocality> ?city .
?service <http://schema.org/name> ?name .
?service <http://schema.org/streetAddress> ?address .
}
UNION
{
```

In the following query the is extracted the value of the assets:

```
?sValRes w3:type
<http://www.disit.org/risk_ontology/risk_ontology/schema#ServiceAssetValue> .
?sValRes rrel:hasValue ?assetValue .
?sValRes rrel:assetValueGroup <serviceType>.
?sValRes rrel:hasValue ?assetValue .
```

It is worth to notice that the instance *sVal Res* is extracted with *assetValueGroup* equal to the type of service selected. Such instance is a *ServiceAsset Value* type.

In the following query the vulnerability layers associated to the seismic hazard are extracted:

```
SELECT ?vos ?value ?sVMin ?sVMax ?sVClass ?description ?sGeoV ?vSWKT WHERE {
?vos km4c:forAssetType \"Land\" .
?vos km4c:hasIndicator
<http://www.disit.org/risk_ontology/resource/SeismicIndicator/id#INDICATOR>.
?vos <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.disit.org/risk_ontology/schema#VOfService> .
?vos <http://www.disit.org/risk_ontology/relation/schema#hasValue> ?value.
?vos rrel:fromMinIntensity ?sVMin .
?vos rrel:toMaxIntensity ?sVMax .
?vos rrel:hasAreaClassification ?sVClass .
?vos rrel:LongDescription ?description .
?vos km4c:hasGeometry ?sGeoV .
?sGeoV <http://www.opengis.net/ont/geosparql#asWKT> ?vSWKT
```

The following query extracts the vulnerability value:

```
SELECT ?value ?vob ?bVMin ?bVMax ?bVClass ?description ?bGeoV ?vBWKT WHERE {
?vob rrel:forAssetType> \"Buildings\" .
?vob rrel:hasIndicator>
<http://www.disit.org/risk_ontology/resource/SeismicIndicator/id#INDICATOR>.
?vob w3:type <http://www.disit.org/risk_ontology/schema#VOfBuildings> .
?vob rrel:hasValue ?value.
?vob rrel:fromMinIntensity ?bVMin .
?vob rrel:toMaxIntensity ?bVMax .
?vob rrel:hasValue ?V .
?vob rrel:hasAreaClassification ?bVClass .
?vob rrel:LongDescription ?description .
?vob km4c:hasGeometry ?bGeoV .
?bGeoV <http://www.opengis.net/ont/geosparql#asWKT> ?vBWKT
}
```

3.3 Human behaviour data analysis

The Wi-Fi network can be used for tracking city users' behaviour with the needed resolution (in space and time), by accessing to data anonymously and exploiting them according to an informed consent with the users when they connect to the Wi-Fi. At this regard, Florence has a free Wi-Fi network (Firenze Wi-Fi) consisting of about

1500 APs. One relevant issue is that Firenze Wi-Fi APs were installed with the aim of providing a good Wi-Fi coverage in the city's centre and in relevant city services as hospital and university.

As a first step, we identified the most active places and areas to the monitored, on which the above presented methodology would be applied. This action has been performed by interviewing the municipality and by using data collected from mobile App (Florence, Where, What?), available for Android, iOS and Windows Phone stores [Bellini at al., 2014]. That App work with smart city API based on Km4City [Badii et al., 2016] and provides general information to the city users almost uniformly in the city and on multi-domain since it provides information and suggestions on: public and private mobility, culture, energy, accommodation, restaurant, tourism, free Wi-Fi, bus lines, car parking, pharmacies, ATMs, events, etc. These services are accessible with geo information.

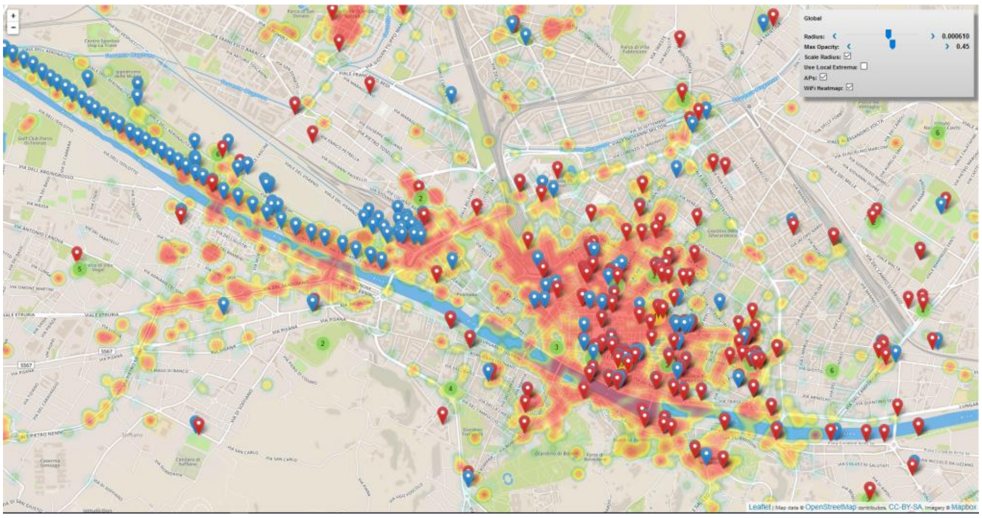


Figure 6 – Heat-map comparing city users' most frequented places vs the position of the 1500 Wi-Fi APs of the whole network (using a colour gradient scale to discriminate between different densities of measures)

Figure 6 reports the heat-map derived from the city users' movements in the city by using the App with overlapped the position of the 1500 AP of the Wi-Fi network. Considering the architectural and environmental constraints of the historical centre of Florence (that is part of the UNESCO World Heritage list), you cannot place APs wherever you want: in most cases, we have to switch on the nearest AP to the predicted one, rather than effectively place the desired AP.

The resulted analysis allowed us to select about 345 APs to be configured and used as probes over the 1500. The data related to the user behaviour tracking via Wi-Fi has been collected in the period from May 2016 to December 2016. They consist of about 56 Million of events of connection and disconnection. Typically, the 60% of connected users are excursionists that stay in the network only for less than 24 hours. In the last 6 months, about 1.15 Million distinct users have been detected, which means about 2.3 million of distinct user per year in a city with about 14 million of new arrivals per year and 350.000 inhabitants. So that we tracked about the 16% of people flow. We compared the predictions from the positioning methodology with the existing APs data, finding the APs to be added and those that were useless for the study. According to the selected AP, the resulting heat-map describing the distribution of measures performed by the AP is reported in **Figure 7**. The developed tool allows customizing the provided map, for example varying the radius and the opacity of the heat spots.

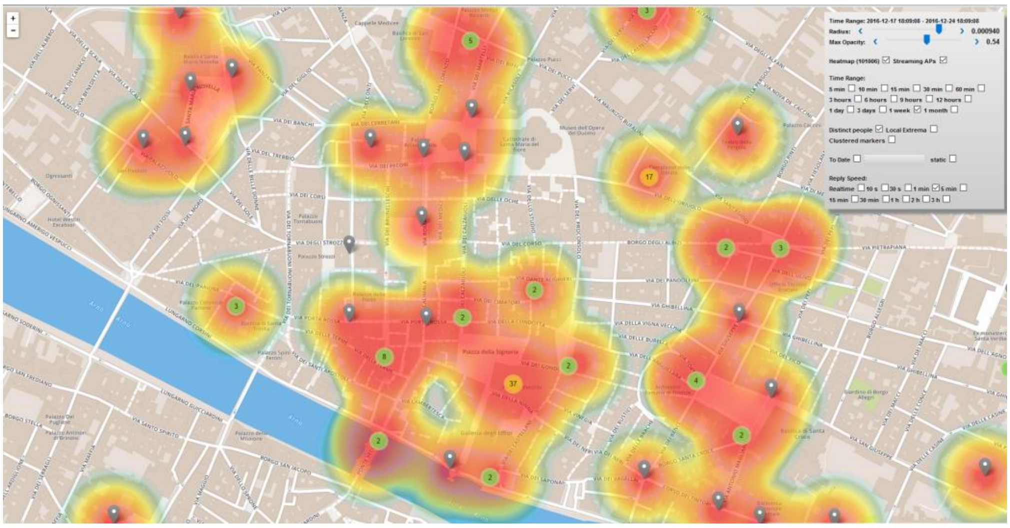


Figure 7 - Segment of the heat-map reporting the hottest places detected by using selected Firenze Wi-Fi APs, in Florence downtown.

The data analysis allows identifying the hottest places (in terms of events on the APs) as reported in **Figure 8**, where the names of the locations and the precise latitudes and longitudes have been truncated for safety reasons. On the other hand, they are also well known location to everybody in the world.
Similarly, a number of visual analytics graphs are produced, such as: the numbers of distinct users during the day, the average connection time per AP, the number of working APs in the last minutes, the recency (percentage of new users with respect to the already seen users) and frequency of users. This last view is of particular importance since it allows estimating the number of new users coming into the city. Indeed, it is worth noting that for cultural cities like Florence, newcomers are typically tourists (excursionists) or business people that stay in the city only for a few hours and days.

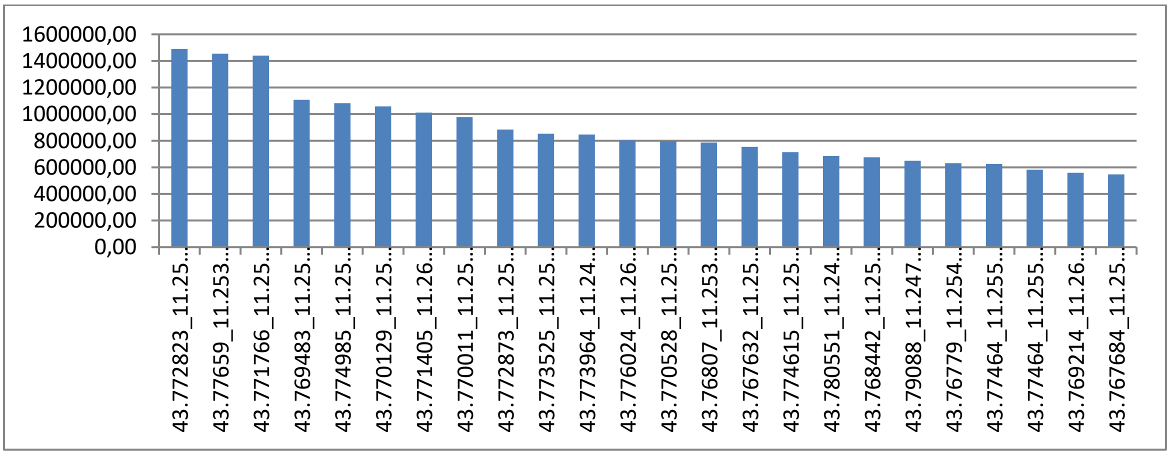


Figure 8 – Distribution of hottest places in the city (truncated series), number of Wi-Fi accesses in last 180 days.

Every working day the network identifies about 34.000 distinct users and among them, about the 10% are new users for the network in the period. For the present analysis, we assumed that new users exploit the city up to 10 days before leaving, while old users continue to exploit the city beyond that limit.

Figure 9 reports the users recency found in the range 1-28 days. Every column in the histogram shows the number of distinct users (y-axis) that at most returned in the city within a defined number of days (x-axis). It is evident from this pattern, that most of the users using the Wi-Fi network are exploiting the city for a few days before leaving. This kind of analysis can be performed at large scale (i.e., considering the whole city) or simply by observing the user behaviour in some zones of interest. For example, the analysis of recency in the historical city centre (which is normally the most exploited part of the city) can provide valuable insights, since it allows understanding which cultural attractions people prefer to visit, or where and how often they return to them.

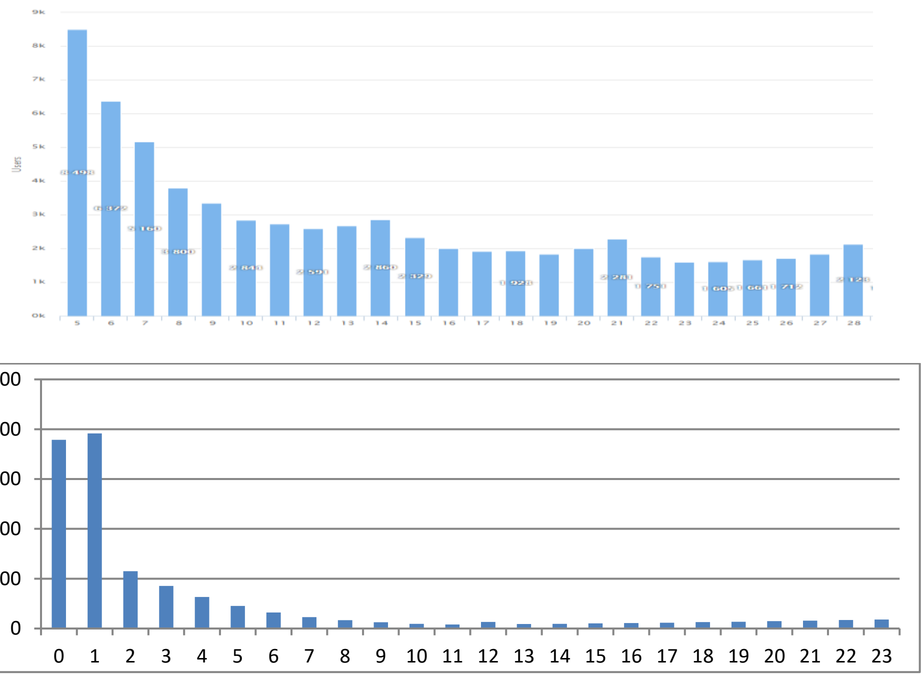
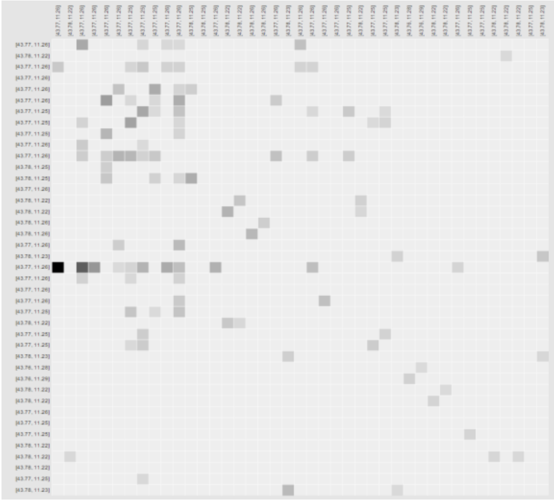


Figure 9 – Number of distinct users accessing to the Wi-Fi network, recency from 1 to 28 days and within the 24 hours (where 0 means from 1 to 59 minutes, 1 from 1 hour and 1:59 minutes, etc.

3.4 Origin Destination Analysis for people flow

To better understand the movements in the city, it is mandatory to perform flow analysis to effectively evaluate user's behaviour. Since in the downtown the APs are also overlapped this issue has to be taken into account. The measures performed by the mobile APP (as described in first part of Section V) have been also used to define a compromising size for each area collecting accesses to the Wi-Fi. On the basis of the tracked city users among the APs of the Wi-Fi network it is possible to computer the OD matrix according to the origin and destination area defined by the distribution of the APs in the city. On the other hand, the OD matrixes are typically quite sparse as one can see in Figure 8**a**, where the OD matrix for Florence is reported.

Figure 8**b** reports a new approach for depicting and analysing the OD matrixes. It is a visual analytic approach for depicting an OD matrix as what we call *OD Spider Flow* in which the analyst may identify the hottest areas of the city as those with larger and darker points/dots. When a dot is selected the graph reports the major (in/out) flows from that origin to the most probable destinations, also providing the percentage of probability on the destination dots. Every flow is depicted with an arrow and a coloured circle reporting the total number of occurrences and their percentage with respect to the total flows.



(a)

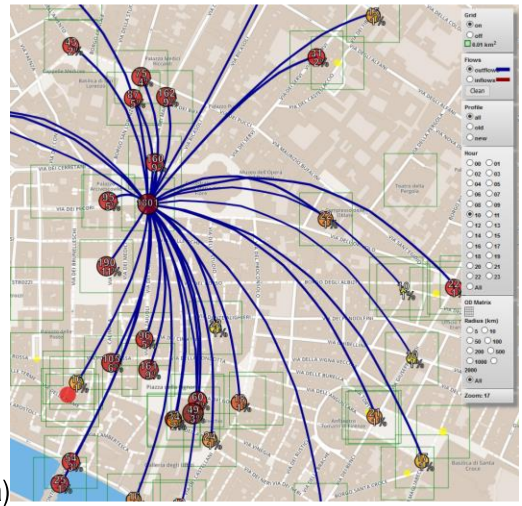


Figure 8 OD matrix for Florence downtown: (a) classical view; (b) advanced proposed view

The analysis can be performed for the whole city users or only for the new arriving users (with respect to the last 10 days), for each time slot of the day or for the whole day, for incoming and outgoing flows, and at different level of resolution (zoom). Zooming in/out the map redraws the flows with a different cluster zone, making possible to depict more detailed or aggregated flows between the various zones. The classical OD matrix can be shown as well from the same tool, also calculated with a customizable range within the city's centre, for the chosen flow configuration (i.e., cluster area's size, hour of the day, user profile). This kind of derived information can be used for running the services in the city, to plan the cleaning, to distribute the security people, etc.

3.5 Understanding City Usage from AP data

From the analysis of the OD matrices and/or OD Spider Flows it is evident that different parts of the city are differently used by different city users. AP presents different kind of trends in the usage of the Wi-Fi along the 24 hours and in the different days of the week (Jiang et al., 2012). For example, we may have some areas by which the people typically arrive (station) in the morning and leave in the afternoon while they are less accessed at lunch time. For example, some APs could have a huge workload only during mornings or evenings (when people go/back to/from work), others only on late evenings (when people go out for entertainment), others only of festive days etc.

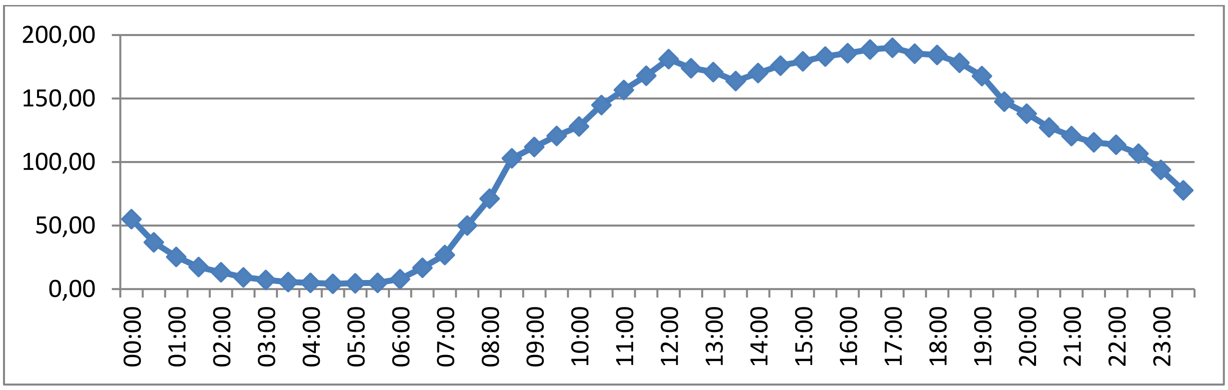


Figure 9 Typical AP trend in terms of number of connections along the 24h

In Figure 9, an example of trend for a certain AP along the 24 hour of the day. The trend of Figure 11 has been estimated by computing the averaged value per time slot of a certain AP every working day, extracting data from the 56 million of data described above.

In Florence, as in many other touristic cities, the issue is much more complex, since a lot of different city users' kinds (with different aims) use the city at the same time during the working days, and as well as on Saturday and Sunday.

Therefore, in order to tune the services in the city (security, cleaning, transport, etc.), it is very important to infer patterns and analyse city user's behaviour. In the present scenario, the major interest is related to understand how the city is used by city users which in turn can be re-conduced to the problem of understanding how APs work and are used. The idea is to exploit some data mining techniques clustering AP on the basis of their normalized temporal pattern. This will allow grouping them in areas and put in evidence the flows and the service exploitation in the different city's zones. Clustering the APs' behaviours can help to understand if there are zones having a similar usage and exploitation and hence similar flow patterns, and needs in terms of services.

According to the data collected from the Wi-Fi network described at the beginning of Section V, the averaged trend along the 24 hours of the day, for each AP, for each day of the week has been computed. Since the main interest is to find the similar patterns for each AP a Scale Factor and the normalized averaged pattern (from 0 to 1) has been computed. This resulted in 345 APs, on 7 days, on 48 time slot for the day (one every 30 minutes) (from 00:00 to 00:30, from 00:30 to 01:00 and so on until 23:30). A preliminary analysis of AP patterns showed a marked difference between festive and ferial days. For this reason, we chose to cluster the time series by keeping track of their respective day of week, thus considering working days, Saturdays and Sundays as three distinct groups. From the statistical point of view, the temporal pattern for each AP presents an average and an interval confidence for each time slot as depicted in the examples reported in Figure 11.

Since we are interested in finding similar patterns for the APs, a clustering approach has been adopted to find similarities in time series as in the Dynamic Time Warping [Wang et al., 2010], and by using different clustering algorithms and metrics to evaluate both the better ranked clustering algorithm and the proper number of clusters.
Among the clustering algorithms we compared the results obtained by using: k-means clustering algorithm minimizes the within-class sum of squares for a given number of clusters [MacQueen, 1967], [Hartigan and Wong, 1979], hierarchical clustering [Suzuki and Shimodaira, 2014], density-based clustering or subspace clustering. Unlike k-means clustering, hierarchical clustering builds a bottom-up hierarchy, and does not need to specify the number of clusters. For the clustering, the closeness of cluster elements can be determined by using (a) complete linkage clustering (i.e., finds the maximum distance between points of two clusters), (b) single linkage clustering (i.e., finds the minimum distance between points of two clusters), (c) mean linkage clustering (finds all pairwise distances for points of two clusters, calculating the average), (d) centroid linkage clustering (i.e., finds the centroid of each cluster and then calculate the distance between the centroids of two clusters).

3.6 AP Clustering experimental results

In this section, the comparative analysis among some of the above mentioned different clustering methods is reported. It should be noted that, different clustering techniques and, even for the same algorithm the selection of different parameters or the presentation order of data objects may greatly affect the final clustering partitions. Thus, the adoption of rigorous evaluation criteria is mandatory to trust the cluster results: selection of model and clusters number.

With the Elbow method (as reported in Figure 10), the solution criterion value (within groups sum of squares) will tend to decrease substantially with each successive increase in the number of clusters: after 8 clusters the observed difference in the within-cluster dissimilarity is not substantial. Consequently, we can say with some reasonable confidence that the optimal number of clusters to be used seems to be 7. Note that identifying the point in which a “kink” exists is not a very objective approach and is very prone to heuristic processes. For these reasons, we computed the Gap statistics [Tibshirani et al., 2001] to assess the optimal number of clusters in the data. From this analysis reported in Figure 11**, t**he estimated number of clusters K=12.

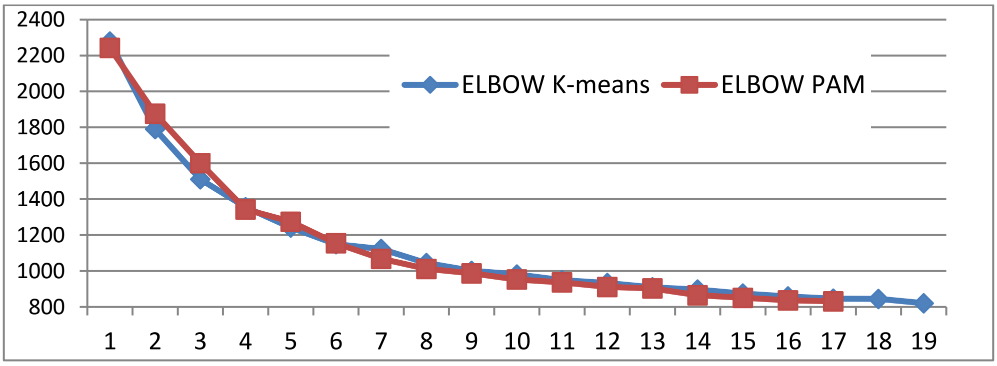


Figure 10 Optimal number of AP clusters via Elbow criteria (comparing K-means and PAM):within sum of square function

.

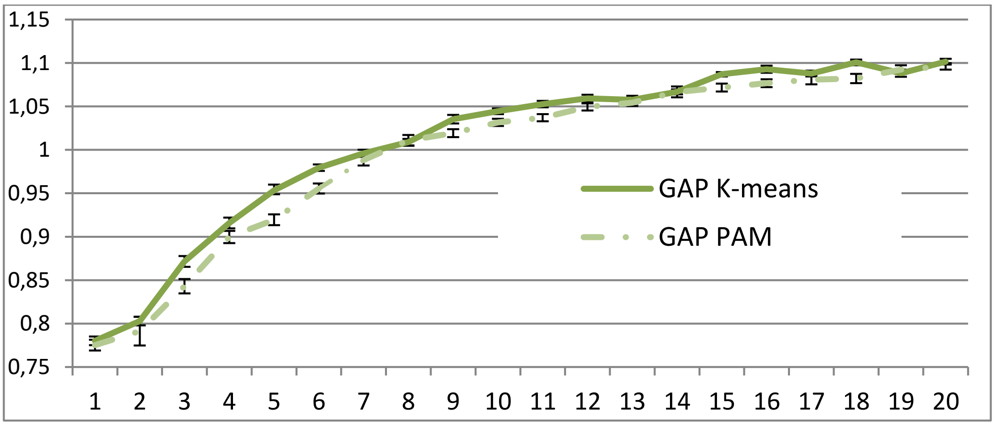


Figure 11 Optimal K number of clusters via Gap curve (comparing K-means and PAM)

Finally, Figure 12 shows the average BIC (Bayesian Information Criteria) values for six different mixture models using the model-based approach over a range of different numbers of clusters [McLachlan and Peel, 2000]. With the VEE mixture model, the maximum average BIC score is reached at 10 clusters. In addition, the VVE mixture model also achieves higher BIC values than the VEE model up to 10 clusters. Therefore, the model-based approach favours the diagonal model which produces higher quality clusters. The BIC analysis selects the VVE model at 10 clusters. Note that although the BIC analysis does not select the best model, it allowed selecting the better number of clusters in this data set.

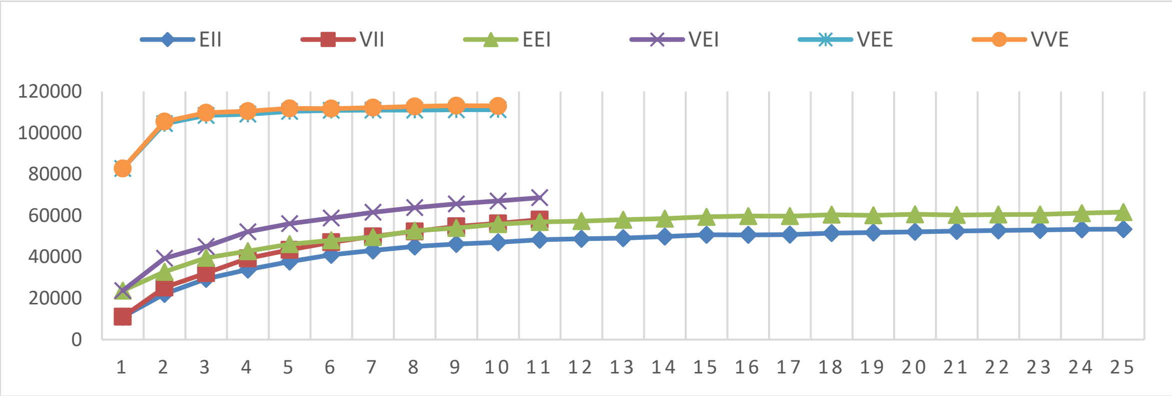


Figure 12 Average BIC for measure models vs K number of cluster, higher values are better, the curves are truncated at the best value for K they found.

We used the Dunn index [Dunn, 1974] as a measure to assess the validity of cluster techniques. Dunn index is based on inter-cluster distance and the diameter of cluster hypersphere. It can be seen that PAM clustering performs the best with 12 clusters (Dunn index for PAM is equal to 0.0798, for K-means is equal to 0.0730 and for Model-based is equal to 0.0478).

Cluster Id	Avg. Std. Dev.	Population
1	0.2379	W: 172, Sa: 23, Su: 24
2	0.0849	W: 23, Sa: 43, Su: 43
3	0.0882	W: 8, Sa: 42, Su: 34
4	0.1820	W: 3, Sa: 30, Su: 26
5	0.1059	W: 20, Sa: 15, Su: 14
6	0.0822	W: 38, Sa: 15, Su: 8
7	0.1311	W: 9, Sa: 57, Su: 34
8	0.1374	W: 2, Sa: 23, Su: 55
9	0.1226	W: 4, Sa: 32, Su: 38
10	0.1460	W: 52, Sa: 12, Su: 3
11	0.2487	W: 11, Sa: 13, Su: 21
12	0.1617	W: 1, Sa: 28, Su: 31

Table III – Standard deviation and population for AP clusters. W: Working days, Sa: Saturday, Su: Sunday

As a final result, the EM algorithm with 12 clusters has been adopted for massive and continuous computing. On this regard, **Table III** reports the average standard deviation and the related population of each AP cluster.

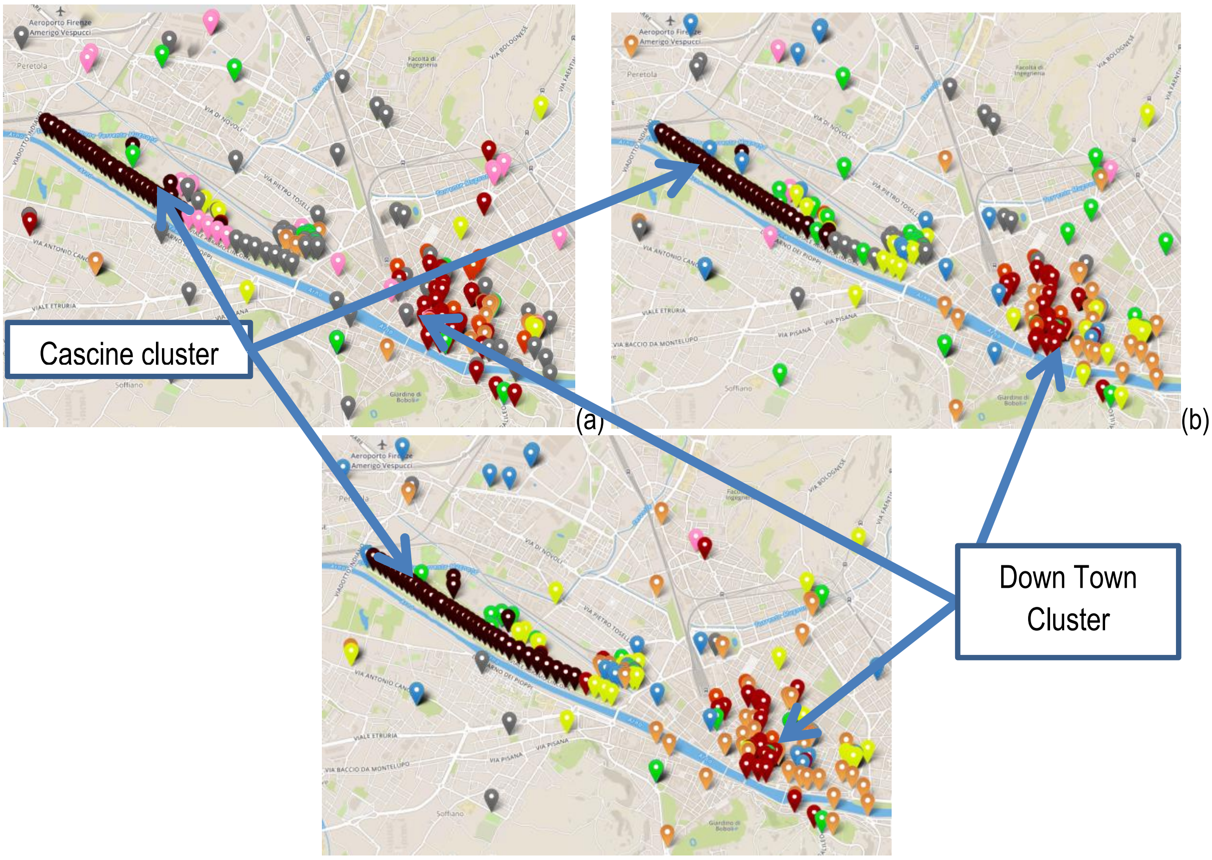


Figure 13 Map of AP clusters: (a) Monday-Friday, (b) Saturday, (c) Sunday

(c)

In Figure 13, the distribution of clustered AP in the Florence map for day kind: Monday-Friday, Saturday and Sunday in which AP of the identical colour belong to the same cluster disregarding the day kind. From Figure 13, it can be noticed that group of APs located in the Cascina park (black) is enlarging passing from working days to Sunday, while the cluster of downtown (dark red) is losing some of its APs passing from working days to Sunday. While some of them remain stable: mainly those located in the major attractions for tourists.

The maps reported in Figure 13can be easily accessed by a real-time tool accessible for the municipality of Florence. Each cluster has a different colour, and clicking on an AP opens a popup with detailed data about the specific AP and the cluster at which it belongs to (i.e., cluster id, maximum, minimum, average flow and standard deviation). In this way, we are able to see in an intuitive manner if there are adjacent zones that show similar AP daily patterns.

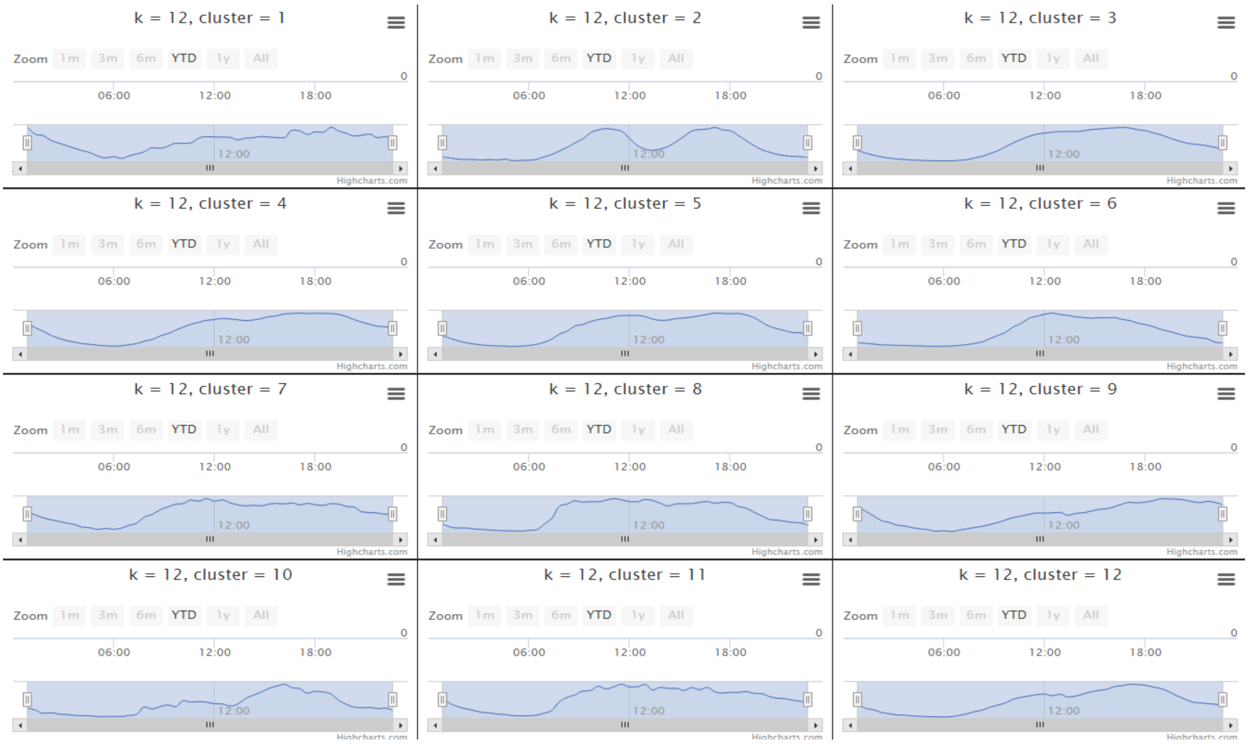


Figure 14 The shapes of AP clusters with K=12 and EM clustering algorithm

Figure 16 - The shapes of the AP clusters with k = 12 and EM clustering algorithm.

The Figure 14 reports the normalized shapes of the 12 clusters identified which resulted from the best clustering algorithm, the EM. It can be noticed that the second cluster presents APs with relevant activity during the morning and afternoon respecting a break for lunch. Moreover, some clusters provide an evident activity in the afternoon with respect to the morning or vice versa, but with different proportions. A few of them present significant activity also after dinner and in the first hours of the night, as clusters number 1 and 9. So that, it is evident where the city is active during the night.

3.7 Predicting Access Point Connections

The data collected from the Firenze-Wi-Fi network have been analysed to derive a number of information and knowledge: heat map as most frequent places and hottest areas in the city, daily user behaviour patterns in the city area to understand how the city is used, OD matrix to extract people movements. In this section, the development of a model for predicting the number of connections of each specific AP in the city is presented. The number of connections of an AP is directly related to people presences. And thus, it can be used for planning in advance, as well as, it poses the basis to be used as an instrument for early warning: that is for detecting dysfunctions as un-expected patterns in the city users' behaviour. To this end, the autoregressive integrate moving average approach (ARIMA), have been adopted as solutions to set up accurate predictive models in order to detect dysfunctions. The autoregressive part (AR) of model creates the basis of the prediction, which can be improved by a moving average modelling for errors made in previous time instants of prediction (MA). The order of ARIMA modelling is defined by the parameters (*p*,*d*,*q*): *p* is the order of autoregressive model; *d* is the degree of differencing, and *q* is the order of the moving average part, respectively. The predictive model has been developed by using Box-Jenkings methodology as ARIMA modelling [Box et al., 1994], and the solution has been compared in terms of performances with a set of other models.

We chose to consider the time series by dividing the week in three distinct groups, thus considering working days, Saturdays and Sundays, in order to maintain consistency with the cluster analysis.

For the analysis, we have applied different predictive ARIMA models for each 30-minute interval, for each groups

of days and for each AP. Note that, for each time interval we estimate the best ARIMA model according to the AIC, Akaike Information Criterion [Akaike, 1987]. In most cases, the best predicting model has been an ARIMA [5,1,0] (it is an ARIMA model), meaning that the model considers 5 observations from the past, and by the difference of the last two observations. The best AICs have been obtained in the range of 1000-1300 in different time slots. Better predictive results have been obtained for the AP in which a significant number of accesses are typically present.
The ARMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors.

In Figure 15, two examples of AP time series with prediction are reported. Each of them reports: in blue line the average value of the cluster at which the AP belong; the light blue bound describes the interval confidence of the reference cluster of the AP; the red line the actual value of the day; the orange bound describe the interval confidence obtained by the distribution of the value of the AP in the past; finally, the RED segment (second part segment) is the effective prediction by using the ARIMA model. Please note that, the adopted ARIMA model does not take into account the value collected by the same AP in the day, since we would like to use the predictive model for detecting dysfunctions and not to follow the most probable next values. The detection of critical situation can be obtained making the difference from those two approaches/ estimations.

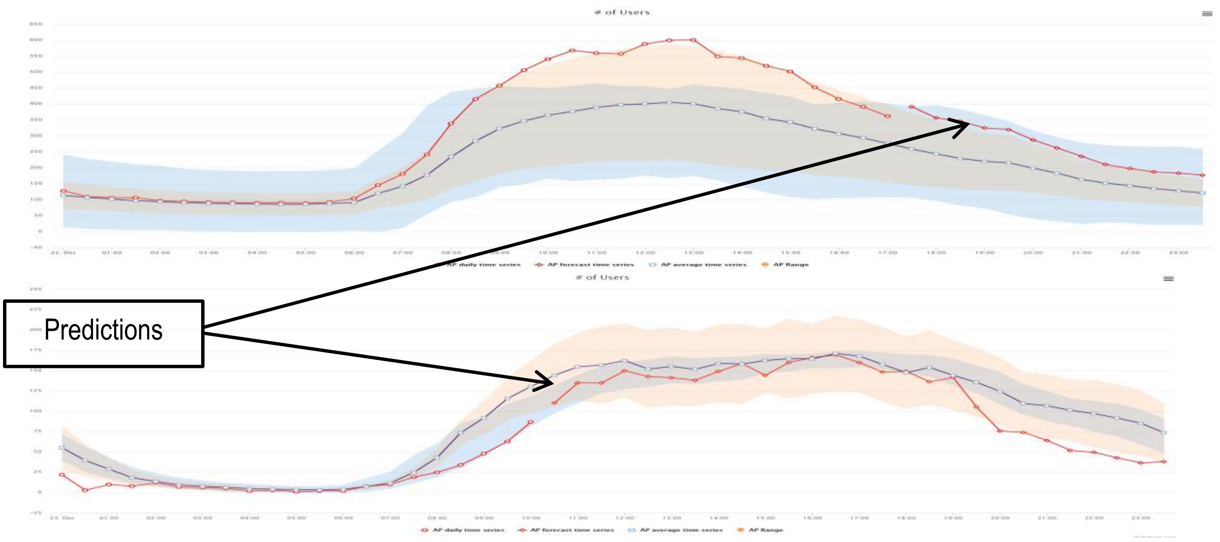


Figure 15 APs time series with their respective cluster ranges (see details in the text)

4 INTEROPREABILITY VIA SMART CITY API

In order to achieve data interoperability a number of RESTful standardised APIs has been developed to provide data access to the front end application. Through such decoupling approach, flexibility and scalability are secured. In fact data intensive processing is entirely delegated to the data layer that is managed through a cloud based infrastructure. In the following paragraphs examples of how to use such APIs are provided.

4.1 Accessing triage data through Smart City API

The data on the triage can be retrieved using the Smart City API, in particular it is possible to find First_aid services with the following REST API:

http://servicemap.disit.org/WebAppGrafo/api/v1/?categories=First_aid&selection=43.7195;10.7775;44.0382;11.5740&maxResults=0&format=json

where the **categories** parameter select the services in the First_aid category, the **selection** parameter allows to indicate the geographical area where the service have to be found, and **maxResults** indicates the maximum number of results to be retrieved. The selection parameter can be:

- a rectangular area identified by two GPS points (latitude;longitude)
- a single GPS position with an additional maxDists parameter stating the maximum distance of the service in km from the point
- a complex geographic area described using a WKT polygon or linestring
- a previously saved WKT area identified with and identifier

the data is provided as GeoJSON:


```
{
    "Services": {
        "fullCount": 11,
        "type": "FeatureCollection",
        "features": [{
            "geometry": {
                "type": "Point",
                "coordinates": [11.091001, 43.879616]
            },
            "type": "Feature",
            "properties": {
                "name": "PRONTO SOCCORSO OSPEDALE MISERICORDIA E DOLCE",
                "tipo": "First_aid",
                "typeLabel": "First aid",
                "serviceType": "Emergency_First_aid",
                "distance": "6.78941877912738",
                "serviceUri": "http:\/\/www.disit.org\/km4city\/resource\/6dd80ecd9964dbc299db51314e418035",
                "multimedia": ""
            },
            "id": 1
        }, {
            "geometry": {
                "type": "Point",
                "coordinates": [11.24566, 43.80311]
            },
            "type": "Feature",
            "properties": {
                "name": "PRONTO SOCCORSO AZIENDA OSPEDALIERA CAREGGI",
                "tipo": "First_aid",
                "typeLabel": "First aid",
                "serviceType": "Emergency_First_aid",
                "distance": "10.11229842769542",
                "serviceUri": "http:\/\/www.disit.org\/km4city\/resource\/2a7e0a917bc856f2102ec1b57e0ee390",
                "multimedia": ""
            },
            "id": 2
        }, …  ]
```

    }
}

When a First_aid is identified, the triage status can be retrieved using the serviceUri that identifies the service. The REST API is:
http://servicemap.disit.org/WebAppGrafo/api/v1/?serviceUri=http://www.disit.org/km4city/resource/2a7e0a917bc856f2102ec1b57e0ee390&format=json
The JSON returned provide more information on the First_aid service, and in the realtime section the values of the triage status.

{
    "Service": {
        "type": "FeatureCollection",
        "features": [
            {
                "geometry": {
                    "type": "Point",
                    "coordinates": [11.24566, 43.80311]
                },
                "type": "Feature",
                "properties": {
                    "name": "PRONTO SOCCORSO AZIENDA OSPEDALIERA CAREGGI",
                    "typeLabel": "First aid",
                    "serviceType": "Emergency_First_aid",
                    "phone": "0554279644",
                    "fax": "0554279644",
                    "website": "",
                    "province": "FI",
                    "city": "FIRENZE",
                    "cap": "50100",
                    "email": "",
                    "linkDBpedia": [],
                    "note": "",
                    "description": "",
                    "description2": "",
                    "multimedia": "",
                    "serviceUri": "http://www.disit.org/km4city/resource/2a7e0a917bc856f2102ec1b57e0ee390",
                    "address": "VIALE MORGAGNI",
                    "civic": "85",
                    "wktGeometry": "",
                    "photos": [],
                    "photoThumbs": [],
                    "photoOrigs": [],
                    "avgStars": 0.0,
                    "starsCount": 0,
                    "comments": []
                },
                "id": 1
            }

```
        ]
    },
    "realtime": {
        "head": {
            "vars": ["measuredTime", "state", "redCode", "yellowCode", "greenCode", "blueCode", "whiteCode"]
        },
        "results": {
            "bindings": [{
                "measuredTime": {"value": "2017-02-27T18:37:26Z"},
                "state": {"value": "Con Destinazione"},
                "redCode": {"value": "1"},
                "yellowCode": {"value": "1"},
                "greenCode": {"value": "0"},
                "blueCode": {"value": "0"},
                "whiteCode": {"value": "0"}
            }, {
                "measuredTime": {"value": "2017-02-27T18:37:26Z"},
                "state": {"value": "In Attesa"},
                "redCode": {"value": "0"},
                "yellowCode": {"value": "4"},
                "greenCode": {"value": "5"},
                "blueCode": {"value": "0"},
                "whiteCode": {"value": "0"}
            }, ….]
        }
    }
}
```


more details on the Smart City API can be found at http://www.disit.org/6597

4.2 Accessing to OD derived from Wi-Fi data via Smart City API

FROM GPS COORDINATE FROM-TO TO PERCENTAGE OF PEOPLE THAT WOULD PASS FROM THE TWO POINTS

This REST API allows the user to get the number of people passing from one zone to another in the city.

Real data are restricted to LAN IP addresses only; clients querying the service from the outside will get fake/obfuscated values only.

Example of usage:

http://www.disit.org/wifi-firenze/ap/streaming-realtime/api/cityUsersPerLocationAPI.php?fromLatitude=43.811004&fromLongitude=11.25231&time=11:30&day=0&data=flow

4.3 Accessing to Wi-Fi data via Smart City API

FROM GPS COORDINATE TO FLOW

This REST API allows the user to get the guessed, real-time and typical number of users in the Firenze Wi-Fi network at a particular time, in a particular day, and place.

Realtime, Guessed and Typical correct data are restricted to LAN IP addresses only; clients querying the service from the outside will get fake/obfuscated Typical values only.

Realtime, returns the number of users averaged at 30 minutes on current time. Requesting it at 10:05, is going to provide people in the last 30 minutes slot (10:00-10:30). Real time are real time values.

Guessed, returns the number of predicted users averaged at 30 minutes slot. For example, being on Friday at 09:00, requesting at datetime for Friday 10:05, is providing the guessed number of people in the slot for 10:00-10:30 for Friday. Guessed are based on SARIMA predictive models.

Typical, returns the number of typical users averaged at 30 minutes. For example, being on Monday, requesting at datetime of Friday 10:07, is providing the typical averaged number of people in the slot 10:00-10:30 for Friday. Typical are based on average clustered models.

Example of usage:

http://www.disit.org/wifi-firenze/ap/streaming-realtime/api/cityUsersPerLocationAPI.php?latitude=43.811004&longitude=11.25231&time=11:30&day=0&data=typical

Where:

- latitude, the decimal latitude of the location;
- longitude, the decimal longitude of the location;
- time, the time in hour:minute format;
- day, the day of the week (0 = Sunday, 1 = Monday, 2 = Tuesday, … , 6 = Saturday)
- data, the type of data to be returned [guessed, real-time, typical]

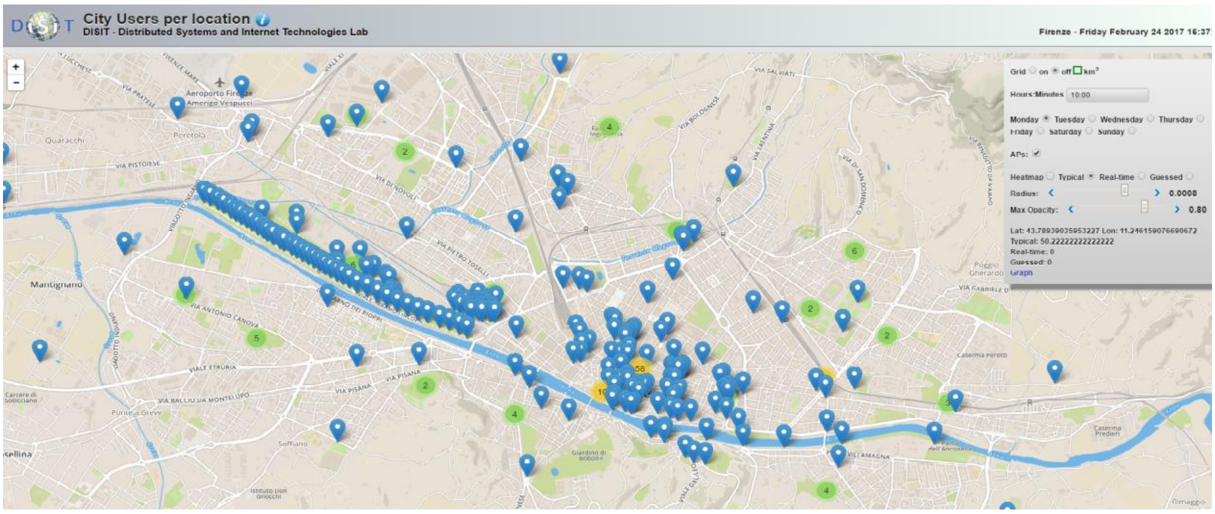


Figure 16 APs on the map and typical, realtime, guessed values for the datetime

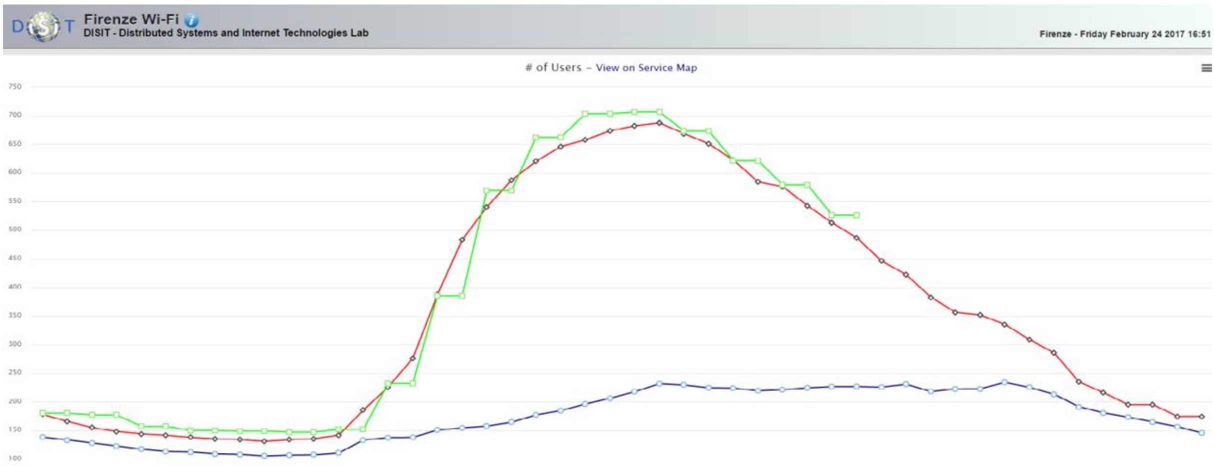


Figure 17 Clustered APs time series: typical (blue), realtime (green), guessed (red).

5 CONCLUSIONS

This deliverable reports the work carried out in the Task 4.3 on data integration and interoperability. Data integration has been achieved starting from the Km4City, a well-known ontology designed to manage big data generated by a smart city. The ontology has extended with new data and new concepts to manage resilience UTS. The data collected has been exploited through semantic processing and data mining in a number of analysis such as: human behaviour (flows, trajectories, velocity, presence), OD matrix, real time damages estimation from flooding and seismic events.

In order to enable access and interoperability with the data, some REST-full APIs has been developed. In particular, through them it is possible to access to the OD metrics. The OD matrix calculation is based on the WIFI data stored in the knowledge base.

The final result is a flexible, scalable and ready to be used data layer able to collect and fuse new and existing resilience related heterogeneous information from the environment and to allow computational intensive advanced processing. The data collected and the processing results are made available to front-end application through APIs. This solution represents the core of the RESOLUTE platform.

REFERENCES

- Akaike H. Factor analysis and AIC. Psychometrika. 1987 Sep 1;52(3):317-32.
- Apache HBase: A Distributed Database for Large Datasets. The Apache Software Foundation, Los Angeles, CA. URL http://hbase.apache.org.
- Ashok, K. and Ben-Akiva, M. E. (2000) Alternative approaches for real-time estimation and prediction of timede-pendent origin-destination flows, Transportation Science 34: 21-36
- Auer, Sören, Jens Lehmann, and Sebastian Hellmann. "Linkedgeodata: Adding a spatial dimension to the web of data." In The Semantic Web-ISWC 2009, pp. 731-746. Springer Berlin Heidelberg, 2009.
- Badii C., P. Bellini, D. Cenni, G. Martelli, P. Nesi, M. Paolucci, "Km4City Smart City API: an integrated support for mobility services", 2nd IEEE International Conference on Smart Computing (SMARTCOMP 2016), St. Louis, Missouri, USA, 18-20 May 2016.
- Banerjee A. and R. N. Dave, "Validating clusters using the Hopkins statistic," 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542), 2004, pp. 149-153 vol.1.

- Barry Bishop, Atanas Kiryakov, Damyan Ognyanoff, Ivan Peikov, Zdravko Tashev, Ruslan Velkov, “OWLIM: A family of scalable semantic repositories”, Semantic Web Journal, Volume 2, Number 1 / 2011.
- Bartolozzi, M. P. Bellini, P. Nesi, G. Pantaleo and L. Santi, "A Smart Decision Support System for Smart City," 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, 2015, pp. 117-122, December 2015, Cina, IEEE press.
- Bellandi Andrea, Pierfrancesco Bellini, Antonio Cappuccio, Paolo Nesi, Gianni Pantaleo, Nadia Rauch, ASSISTED KNOWLEDGE BASE GENERATION, MANAGEMENT AND COMPETENCE RETRIEVAL, International Journal of Software Engineering and Knowledge Engineering, Vol.22, n.8, 2012
- Bellini E., P. Nesi, L. Coconea, E. Gaitanidou, P. Ferreira, A. Simoes, A. Candelieri, "Towards resilience operationalization in Urban Transport System: the RESOLUTE project approach", ESREL 2016, European Safety and Reliability, University of Glasgow, 25-29 September 2016. (pending) http://esrel2016.org/
- Bellini P., D. Cenni, P. Nesi, "AP Positioning for Estimating People Flow as Origin Destination Matrix for Smart Cities", The 22nd International Conference on Distributed Multimedia Systems, DMS 2016, Italy, (in press), 2016.
- Bellini P., Di Claudio M., Nesi P., Rauch N., "Tassonomy and Review of Big Data Solutions Navigation", Big Data Computing To Be Published 26th July 2013 by Chapman and Hall/CRC
- Bellini P., I. Bruno, A. Cavaliere, D. Cenni, M. DiClaudio, G. Martelli, S. Menabeni, P. Nesi, G. Pantaleo, N. Rauch, "Km4City as Smat City Semantic Model and Tools",Open Track, http://challenge.semanticweb.org/2015/submissions/, Semantic Web Challenge of the 14th International Semantic Web Conference (ISWC2015), Bethlehem, PA, USA, October 11-12, 2015.
- Bellini P., M. Benigni, R. Billero, P. Nesi and N. Rauch, "Km4City Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", International Journal of Visual Language and Computing, Elsevier, 2014, http://dx.doi.org/10.1016/j.jvlc.2014.10.023
- Bellini P., M. Benigni, R. Billero, P. Nesi and N. Rauch, "Ontology Bulding vs Data Harvesting and Cleaning for Smart-city Services", 20th International Conference on Distributed Multimedia Systems, DMS 2014, Wyndham Pittburgh University Center, Pittsburgh, USA, 27-29 Aug. 2014.
- Bellini P., P. Nesi and N. Rauch, "Knowledge Base Contruction Process for Smart-city Services", IEEE 19th International Conference on Engineering of Complex Computer Systems (ICECCS 2014). 2014, Cina.
- Bellini P., 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
- Bellini Pierfrancesco, Monica Benigni, Riccardo Billero, Paolo Nesi, Nadia Rauch, Km4City ontology building vs data harvesting and cleaning for smart-city services, Journal of Visual Languages & Computing, Volume 25, Issue 6, December 2014, Pages 827-839
- Box GE, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. John Wiley & Sons; 2011 Sep 20.
- Calinski T., Harabasz J. (1974), A Dendrite Method for Cluster Analysis", Communications in Statistics, Theory and Methods, 3(1), 1-27
- Caragliu A., Del Bo C., Nijkamp P., Smart cities in Europe, 3rd Central European Conference in Regional Science CERS, 2009
- Cascetta, E. and Postorino, M. N. (2001) Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks, Transportation Science 35: 134-147
- Danalet, Antonin, Michel Bierlaire, and Bilal Farooq. ”Estimating Pedestrian Destinations using Traces from WiFi Infrastructures”, Pedestrian and Evacuation Dynamics 2012. Springer International Publishing, 2014. 1341-1352.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society Series B. 1977;39:1–38.
- Doblas, J. and Benitez, F. G. (2005) An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix, Transportation Research, Part B:

Methodological 39: 565-591

- Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". Journal of Cybernetics. 3 (3): 32–57.
- Embley, David W., Douglas M. Campbell, Yuan S. Jiang, Stephen W. Liddle, Deryle W. Lonsdale, Y-K. Ng, and Randy D. Smith. "Conceptual-model-based data extraction from multiple-record Web pages." Data & Knowledge Engineering 31, no. 3 (1999): 227-251.
- Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association, 17:126–136, 2001.
- Giffinger, Rudolf; Christian Fertner; Hans Kramar; Robert Kalasek; Nataa Pichler-Milanovic; Evert Meijers, Smart cities Ranking of European medium-sized cities, 2007 Available at: http://www.smartcities.eu/
- Gong, Z. (1998) Estimating the urban o-d matrix: a neural network approach, European Journal of Operational Research 106: 108-115
- Gupta S., P.Szekely, C.Knoblock, A.Goel, M.Taheriyan, M.Muslea, "Karma: A System for Mapping Structured Sources into the Semantic Web", 9th Extended Semantic Web Conference (ESWC2012).
- Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." Journal of the Royal Statistical Society. Series C (Applied Statistics) 28.1 (1979): 100-108.
- Jiang, Shan, Joseph Ferreira, and Marta C. González. "Clustering daily patterns of human activities in the city." Data Mining and Knowledge Discovery 25.3 (2012): 478-510.
- Kaufman, Leonard, and Peter J. Rousseeuw. "Partitioning around medoids (program pam)." Finding groups in data: an introduction to cluster analysis(1990): 68-125
- Kim, H., Beak, S. and Lim, Y. (2001) Origin-destination matrices estimated with a genetic algorithm from link traffic counts, Transportation Research Record 1771, Transportation Research Board, Washington, D.C.:156-163
- KM4City Sentient city Ontology. http://www.disit.org/km4city/schema http://www.disit.org/km4city/schema
- Lundgren, J. T. and Peterson, A. (2008) A heuristic for the bilevel origin-destination matrix estimation problem, Transportation Research, Part B: Methodological 42: 339-354
- Ma, Wenyuan, et al. "Detecting Pedestrians Behavior in Building Based on Wi-Fi Signals." 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). IEEE, 2015.
- MacQueen JB (1967), Some Methods for Classification and Analysis of Multivariate Observations, In LML Cam, J Neyman (eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pp. 281-297
- McLachlan GJ, Peel D. Finite mixture models. Wiley; New York: 2000.
- Nie, Y., Zhang, H. M. and Recker, W. W. (2005) Inferring origin-destination trip matrices with a decoupled GLS path flow estimator, Transportation Research, Part B: Methodological 39: 497-518
- Ontology of Trasportation Networks, Deliverable A1-D4, Project REWERSE, 2005 http://rewerse.net/deliverables/m18/a1-d4.pdf
- Pan, Feng, and Jerry R. Hobbs. "Temporal Aggregates in OWL-Time." In FLAIRS Conference, vol. 5, pp. 560-565. 2005.
- Patil, P. H., and A. A. Kokil. ”WiFiPi-Tracking at mass events.” Pervasive Computing (ICPC), 2015 International Conference on. IEEE, 2015.
- Pentaho Data Integration, http://www.pentaho.com/product/data-integration
- Piorkowski, M.; Sarafijanovic-Djukic, N.; Grossglauser, Matthias, ”A parsimonious model of mobile partitioned networks with clustering”, Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International, pp.1,10, 5-10 Jan. 2009
- Report on the state of play of the outcome of the work of the High Level Group, European Commission, October 2014
- Schauer, Lorenz, Martin Werner, and Philipp Marcus. "Estimating crowd densities and pedestrian flows using

wi-fi and bluetooth." Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.
- Suzuki R, Shimodaira H, pvclust: Hierarchical Clustering with P-Values via Multi-scale Bootstrap Resampling. R package version 1.2-2
- Tibshirani R., Walther G., Hastie T., Estimating the number of clusters in a data set via the gap statistic, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001, 63:411-423
- Vilajosana, I. ; Llosa, J. ; Martinez, B. ; Domingo-Prieto, M. ; Angles, A., "Bootstrapping smart cities through a self-sustainable model based on big data flows", Communications Magazine, IEEE, Vol.51, n.6, 2013
- Wang X., Mueen A., Ding H., Trajcevski G., Scheuermann P., Keogh K., "Experimental comparison of representation methods and distance measures for time series data", Data Mining and Knowledge Discovery, 2010: 1–35
- Raskin, R. G., and Pan, M. J, (2005), Knowledge representation in the semantic web for Earth and environmental terminology (SWEET), Computer & Geosciences, 31, 1119–1125.

- Gangemi A., Guarino N., Masolo C., Oltramari A., Oltramari R., and Schneider L., Sweetening, (2002), Ontologies with DOLCE, in Gómez-Pérez, A. and Benjamins, V. R., Eds., Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Proceedings of the 13th International Conference, EKAW 2002 Siguenza, Spain, October 1-4, 2002, 223–233, Berlin
- Chau, K. W., (2007), An ontology-based knowledge management system for flow and water quality modeling. Advances in Engineering Software, 38, 172–181.
- Tripathi, A., and Babaie, H. A.,(2008), Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies, Computer & Geosciences, 34, 1022–1033.
- Kollarits S. and Wergles, N., (2008), MONITOR - An ontological basis for risk management.

APPENDIX A: GLOSSARY

AB	Advisory board
API	Application Program Interface
AXCP	AXMEDIS Content Processing Tool
CA	Consortium Agreement
CAS	Conditional Access System
CC	creative common
CC0	Creative Common 0, a sort of public domain
CEN	European Committee for Standardization
CSV	A comma-separated values file stores tabular data (numbers and text) in plain-text form.
DCAT	Data Catalogue vocabulary
DICCOF	Collection of tools for data analytics and semantic computing of UNIFI DISIT
DISIT	Distributed Systems and Internet Tech lab & Distributed Data Intelligence Lab of UNIFI
DMP	Data Management Plan (reference document and template of the European Commission)
DPR	Data Protection Rights
DRM	Digital Rights Management
EB	Executive Board
EC	European Commission
ETL	Extraction, Transformation, and Loading
FOAF	Friend of a friend
GA	General Assembly
HTTP	HTTP protocol
IEC	International Electro technical Commission
IODL	Italian Open Data License, equivalent to CC-BY-SA (>3.0), and to ODC-ODbL.

	http://www.formez.it/iodl/
IPR	Intellectual Property Rights
ISO	International Standard Organization
ITS	Intelligent Transport Systems
JS	JavaScript
KML	Keyhole Markup Language, XML notation, expressing geographic annotation and visualization, two-dimensional maps and three-dimensional Earth browsers.
LIP	License Information Public (Fr), http://www.patrimoine-immateriel.fr/licences-information-publique/licence-information-publique-2-0-lip/
LOD	Linked open data
LOG	Linked Open Graph, a service and tool of UNIFI DISIT for accessing graphically to SPARQL entry point of LOD
MO	mobility operator: public transportation operator, parking operators, vehicle sharing, etc.
NGO	Non-governmental organization
NLP	Natural Language Processing
NoSQL	No Structured Query Language
OD	Open Data
ODC	Open Data Commons, http://opendatacommons.org/ license
ODC-By	ODC Attribution License — “Attribution for data/databases”
ODC-ODbL	Open Database License — “Attribution Share-Alike for data/databases”
ODC-PDDL	Public Domain Dedication and License — “Public Domain for data/databases”
OGD	Open Government Data
OGL	Open Government License for public sector: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/
OSIM	Open Space Innovative Mind, tools of UNIFI DISIT
OTN	Ontology on Transportation Network
OWL	Ontology Web Language
PA	Public Administration, such as municipality, city administration
PC	Project Coordinator
PDF	Portable Document Format is a file format used to present documents in a manner independent of application software, hardware, and operating system.
RATP	Régie autonome des transports parisiens
RDF	Resource Description Framework
REST	Representational state transfer
RTTI	Real-time Travel & Traffic Information
SCA	Smart City Actors
SHP	The Esri shapefile, or simply a shapefile, is a popular geospatial vector data format
SKOS	Simple Knowledge Organisation System
SME	Small or Medium Enterprise
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
SPARQL	PARQL Protocol and RDF Query Language
SQL	Structured Query Language
UGE	User Group of Experts
UML	Unified Modelling Language
USDL	Unified Service Description Language
VM	Virtual Machine on cloud
W3C	World Wide Web Consortium
WSDL	Web Services Description Language
XML	Extensible Markup Language