Ref. Ares(2017)1553670 - 22/03/2017

MULTI SOURCE DATA ACQUISITION

Project Title	RESOLUTE
Project number	653460
Deliverable number	D4.2
Version	V10.11
State	FINAL
Confidentially Level	PU
WP contributing to the Deliverable	WP4
Contractual Date of Delivery	M18 (30/10/2016)
Finally approved by coordinator	21/03/2017
Actual Date of Delivery	22/03/2017
Authors	Emanuele Bellini, Michela Paolucci, Paolo Nesi
Email	Emanuele.bellini@unifi.it, michela.paolucci@unifi.it, paolo.nesi@unifi.it
Affiliation	University of Florence
Contributors	Stefano Morelli (UNIFI-DST), Alessandro Battistini (UNIFI-DST), Anastasios Drosou , Alexandros Zamichos, Ioannis Symeonidis (CERTH), Jan Paul Leuteritz (Fraunhofer) Athanasios Tryferidis (CERTH) Ilias Kalamaras (CERTH)

funded by the Horizon 2020
Framework Programme of the European Union

EXECUTIVE SUMMARY

This deliverable presents an analysis of the dataset kinds, and the early identification, assessment and selection of actual datasets that can be used for RESOLUTE pilots. The dataset collection has been carried out according to a specific strategy for reducing the amount of those that are useless. In particular, the activities have benefitted of the work carried out in the WP5 Task 5.1 on user requirements. In fact, a preliminary mapping of the datasets identified to the type of decisions that operators would improve in their daily activities has been provided. The identification and selection results are detailed in the Annex A for Florence Pilot, and Annex B for Athens Pilot. The methodology to ingest heterogeneous datasets in the Data Layer is also presented. Moreover, in order to address the information needs of the operators, several new datasets have been created. In particular, the Flooding susceptibility areas and the wifi data about people flow, and the triage status for Florence city, a generic city weather database for Florence and Athens and a twitter-based dataset for supporting Evacuation DSS operators for Florence. A summary of the lesson learnt during the data selection and collection has been providing in the conclusions.

PROJECT CONTEXT

Workpackage	WP4: Platform Backend
Task	T4.2: Multi source data acquisition
Dependencies	This document is preparatory for the WP5 because of lists the dataset available that will be exploited by models and algorithms

Contributors and Reviewers

Contributors (see Authors in the cover)	**Reviewers**
Stefano Morelli (UNIFI-DST),	Evangelia Gaitanidou, Quality Manager
Alessandro Battistini (UNIFI-DST)	Francesco Archetti, CMR
Anastasios Drosou (CERTH)	Maria Gemou, CERTH
Jan Paul Leuteritz (Fraunhofer)	
Ioannis Symeonidis (CERTH)	
Athanasios Tryferidis (CERTH)	
Ilias Kalamaras (CERTH)	

Version History

Version	Date	Authors	Sections Affected
1.0	28/10/2016	Emanuele Bellini	All
1.0	29/10/2016	Stefano Morelli, Alessandro Battistini	5.1
2.0	24/11/2016	Michela Paolucci	2.4- 2.5
3.0	10/12/2016	Emauele Bellini	All
4.0	14/12/2016	Emanuele Bellini	All
5.0	20/02/2017	Emanuele Bellini	All
6.0	24/02/2017	Emanuele Bellini	All
7.0	27/02/2017	Anastasios Drosou, Alexandros Zamichos	5.2 and 5.6
8.0	28/02/2017	Emanuel Bellini	All
9.0	01/03/2017	Emanuel Bellini	All
10.0	01/03/2017	Paolo Nesi	All
10.6	08/03/2017	Emanuele Bellini	Section 4, 5
10.7	15/03/2017	Paolo Nesi	All
10.8	16/03/2017	Emanuele Bellini	Section 4, 5,6
10.9	17/03/2017	Emanuele Bellini	All
10.10	21/03/2017	Emanuele Bellini	Acronym table, formats
10.11	22/03/2017	Paolo Nesi	All

Copyright Statement – Restricted Content

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

The RESOLUTE project is deeply focused on Big Data as the core mean for enhancing UTS resilience. The objective is to move from the current expertise driven towards the data driven decision making (DDDM). To support such a paradigm shift in resilience management, a new set of decision support systems able to exploit the big data generated by complex systems is necessary. On the other hand, the “garbage in, garbage out“ principle of information systems is still valid also in the big data era. In fact, in order to support DDDM, it is necessary to have reliable data to be collected and processed. This deliverable reports the strategy adopted, the data identification and selection methods based on user requirements, the datasets early identification and selection result (Annex A and B) and the methodology to ingest heterogeneous datasets in the Data Layer.

There are several types of data being collected and managed by the RESOLUTE data layer and used by the CRAMSS: Infrastructure, Urban Transport & Mobility, Events, Social, Environment, Healthcare, Service & POI and Human Behaviour. In particular, Infrastructure data include municipality open data, such as: structure of the city, streets, bridges, underpasses, etc.; mobility data are referred to the traffic flows and Public Transport system data; events are both static and dynamic and come from the several operators such as urban police that dected what happens in the city; social data are referred to info coming from social network such as Twitter or generated by Wi-Fi AP and mobile devices; Environmental data include seismic risk maps, hydrological risk maps, level of the river, pollution, etc.; Healthcare data are related to the position of the hospitals as well as the real time load of the emergency rooms; services & POI data refer to all private and public activities in the city as shops, public office, restaurant, hotels, etc.. Several of them are generated and collected in real time data such as emergency triage status of hospital, environmental sensors, etc. (Bellini, Cenni, Nesi, 2016).

UTS Big Data include: description of the public transportation, busses poles and their timelines, taxi, parking areas and availability, metro status and position, cycle paths, restricted traffic zone, street direction and capabilities, traffic flow information, origin destination matrices for cars, traffic flow movements.

Human behaviour data may be either individual or group-based and include activity related and behavioural personal or collective profiles addressing psychological, habitual and cognitive aspects. These profiles may be extracted based on different kinds of sensors: Wi-Fi network, Bluetooth servers, traffic flow sensors as spires, TV-cameras, mobile cells from telecom operators, mobile Apps, etc., by using data mining, data analytics techniques, processing huge amount of data related to the single movements in the city.

Another important input is data from social networks. A social network crawler can be exploited to manage and analyse all real-time data streaming from the citizens and the city infrastructure (Grasso et al., 2016). The crawler should be language independent utilizing multilingual thesaurus. Text processing and knowledge mining techniques should be used to discover hidden information. In addition to the dynamic data, an interoperable knowledge base contains cross sectors data that can be used to provide services to help the environment to become more efficient in disaster situations. Furthermore, the activities of data analytics and semantic reasoning are used to generate new knowledge that can be integrated into the interoperable knowledge base where cross sectors data are used to help improve resilience in situations of danger (e.g., data ingestion, mining and algorithms, computing models and recommendations).

This deliverable is organised as follow: in chapter 2 the overall Big Data approach and strategy for RESOLUTE data collection is presented; in section 3 Data Ingestion method is described; section 4 introduce the RESOLUTE datasets exploited in the project; in section 5 reports the work done to create new datasets useful needed for the RESOLUTE pilots; conclusion and lesson learnt has been provided in section 6. The details of the datasets selected for the two pilots has been reported in annex A and B.

1.1 Relation with the project

The outcomes of the Task 4.2 are strictly connected with the Pilot definition and execution (WP6), with the back-end implementation and integration (WP4 Task 4.1 and Task 4.3) and with the front-end application (WP5).

In fact, the verification of the data availability is one of the most critical aspects to be addressed in order to extract relevant information to support resilience oriented decisions. The data identification and collection campaign has been driven by the result of the conceptual model (D2.2), the ERMG (3.5 and 3.7) and the scenarios discussion (D7.2, D7,3). A relevant contribution has been provided by the activities of the T5.1 whose outcome, relating to the analysis of the kind of decisions that should be taken by operators, has been used in the present document as a criterion to select datasets.

2 BIG KNOWLEDGE-INFORMATION-DATA (KID) FOR RESILIENCE

2.1 The approach

There is a consensus around the importance of generating system capacities towards enhanced adaptability and on the fact that such capacities must be based on a more flexible management of available resources. The growing presence of technology at every system level pushed by the smart city trend, which is in turn supported by the Internet of Everything (IoE) propagation, renders Knowledge-Information-Data (KID) one of the most critical resources and therefore, one of the key factors to be considered when addressing urban resilience. Information is the basis of the large majority of urban interactions, whether these take place as Human-to-Machine (H2M), Human-to-Human (H2H), or Machine-to-Machine (M2M).

Improved flexibility and adaptability exploiting information technologies constitute a fundamental step towards enhanced systems resilience, particularly by better supporting the ability to cope with increasingly higher degrees of variability and uncertainty, both under "normal" and degraded operational modes. Data has been transformed into information and, together with data mining and semantic processing, fuels the new generation of decision support systems.

The main advancement here is represented by the usage of real data generated by daily activities in the city (cumulating knowledge about critical events, and normal conditions) to assess resilience with respect to classical expert judgment and/or simulated data approaches. In fact, the capacity to cope with expected and unexpected changing conditions can be assessed and quantified by monitoring the resources/assets available in the city, how they are allocated and consumed, and so forth. It is possible to support decision makers in taking better decisions, as monitoring can now be performed in real time, and information and knowledge precision, comprehensiveness and granularity can be increased, while improving the timing, as Early Warning tools. In turn, the decisions can be phrased and communicated benefitting of the same enhancements: precision, comprehensiveness, granularity and timing. In particular, citizens can now be reached through a 4R communication strategy (right person, at the right time, in the right place, with the right message).

The effect of such feedback can immediately be detected through environmental sensors and channels, so that adjustments aiming to reduce unwanted variability (e.g., people escaping in a wrong direction) can be actuated in time. This loop, enabled by new technologies and KID approach, implements what is called the Evidence Driven Adaptive Cycle and needs to be continuously performed. In fact, city managers should monitor resources availability and adapt their decisions not only during an emergency but also in daily activities, in order to prevent or mitigate the cascading effects of unwanted variability. Therefore, the implementation of a Smart City and the corresponding exploitation of the Big KID generated sustain the urban adaptability and resilience since they:

• enhance the monitoring and control capability, by improving the granularity and breadth of knowledge and awareness about the system status and dynamics, continuously collecting Big Data from heterogeneous data sources/streams and sensors as the GPS positions of people, their concentration, typical trajectories in the city, behaviours and sentiment through smart devices and social networks (User Generated Data), Open Data, data from environmental and other kind of sensors (e.g., traffic flows, hydrometers, air pollution, underpasses water level, people flow, temperature), and real time reports such as weather forecast, social media, and so forth;

• enhance the responding capability by providing detailed and timely information to authorities on one side, and to delivering personalised, real time, context aware, and ubiquitous advices to the community;

• enhance the learning capability by applying advanced analysis on Big Data, performing data analytics and data mining to extract knowledge and learn from the events occurred, reaction performed, etc. (e.g., identifying metrics and indicators which may allow you to set up of predictive models and models for early warning and/or anomaly detection and/or selecting suitable reactions; the data analytics performed by using statistic or machine learning approaches);

• enhance the anticipating capability by continuously supporting the assessment of vulnerability and identifying when the system operates nearer to safety boundaries, predicting behaviours and event dynamics, supporting evidence based decisions at strategic, tactic and operation level, thus moving ahead with respect to the current practices based on pre-simulated emergency scenarios (Woltjer, 2006).

Data-driven Decision Management (DDDM) is an approach born in the business governance that values decisions that can be backed up with verifiable data. The success of the data-driven approach is reliant upon the quality of the data gathered and the effectiveness of its analysis and interpretation. In order to implement such an approach the following steps have been defined and accomplished:

1. Define a strategy

Complex Systems like UTS are generating tons of data and there is the risk of being overwhelmed and getting lost in the noise and hype of surrounding data. Starting with a strategy helps decision makers to ignore the hype and cut to what is going to make a difference for *resilience management.* Thus, instead of starting with what data could or should be accessed, the approach followed started on the basis of what ERMG aims to achieve.

2. Identify the resilience needs for decision makers

Once the strategic objectives have been identified, the next step is to work out which questions are needed to be answered in order to achieve those goals. By working out exactly what is needed to be known, focus can be on the data that is really needed. The data requirements, cost and stress levels are massively reduced when moving from '*collect everything just in case'* to '*collect and measure x and y to answer questions z'*.

3. Identify the data to address resilience needs

This step aims at identifying which data are needed to be accessed or acquired in order to answer at these questions. It's really important to understand that no type of data is inherently better or more valuable than any other type, in general. The focus is on identifying the data that best fits the purpose: the data that could help answering the most pressing questions and deliver on strategic objectives. This step should not be stuck only on what is currently available. In fact, if during the analysis the analysts realize that question can be better answered by a measurement or data that is not current performed or available, this outcome will be used to drive the subsequent improvement of the monitoring capability within the complex system.

4. Identify data availability

Once the data needed have been identified, it is necessary to start looking for online open datasets. If some critical datasets are not released with open licence, specific agreement with the data providers can be put in place.

5. Collect the big data

Much of this step comes down to setting up the processes and people who will gather and manage data. The critical points are related to data format, data rate, metadata, APIs, etc. The big data management requires computational capabilities as well as a system to manage data processing, analysis and reuse.

2.2 The strategy definition

In Decision Analysis, there is a distinction between a good decision and a good outcome, namely that one cannot guarantee a good outcome (i.e., desired effects), but through careful and insightful problem framing, data collection, and sound analysis, one can increase the odds of achieving a good outcome by making a good decision (Linkov, Collier) (Howard, 1988).

There are several decision loops like OODA (Observe, Orient, Decide, and Act), PDCA (Plan-Do-Check- Act), DISA (Sense, Interpret, Decide Act), DDAD (Detect, Decide, Act, Desired Effect), etc. All of them have a centralized "Command and Control" (CAC) perspective that might be too rigid in a socio-technical system as a city, where there are a number of control rooms, heterogeneous responsibilities, people who tend not to follow official recommendations, etc.

Bungay's Directed Opportunism approach (Bungay, 2010) represents the main shift of the strategy, from centralized "command and Control" to "Mission Control". The Bungay's approach is a control loop composed by Plans (e.g. reduction of the car presence in the affected area), Actions (e.g. sets of traffic lights cycle, close streets, send recommendation message to city panels) and Outcome (e.g. actual reduction of the car presence in the targeted area), steps where the objective is to leave people/operator free to take opportunistic decisions exploiting their local knowledge that is considered more accurate. However, to guarantee that such decisions are actually right, the following gaps need to be crossed:

(1) Knowledge Gap: The difference between what we would like to know and what we actually know.

(2) Alignment Gap: The difference between what we want people to do and what they actually do.

(3) Effects Gap: The difference between the expected and the actual results of our actions.

When these gaps are encountered, the intuitive response is to seek increasing our control on these areas by gathering more detailed information, providing more detailed instructions, and installing tighter controls (typically metrics). These intuitive responses typically have the opposite effect of their intent, creating greater confusion and entropy. In fact, in order to reduce these gaps, it is necessary to apply the following actions:

(1) Do not command more than is necessary or plan beyond the circumstances that can be foreseen (Knowledge gap).

(2) Communicate to every one as much of the higher intent as is necessary to achieve the purpose (Alignment gap).

(3) Make sure everyone is empowered to make decisions within bounds (Effects gap).

Such an approach is strictly connected to the Resilience Engineering background since the flexibility, that characterise the Bungay's view, is one of the core qualities of a resilient systems.

2.3 Identify the resilience management needs

Resilience management needs have been analysed during the 1st RESOLUTE workshop in Florence (D7.3) and the 2nd RESOLUTE workshop in Athens (D7.4) where a number of scenarios has been proposed to the stakeholders. In particular the work performed in Athens within the Task 5.1 on user interaction, aimed at identifying the resilience needs in operational terms. More details on the analysis will be reported on D5.2. The preliminary results of this work identified what kind of decisions the operators want to make or improve using the RESOLUTE solutions. The following list reports a summary of such kind of decisions:

1. Making available special equipment or resources (operators, volunteers, funds) in advance (anticipate)
2. Which kind and how many units to dispatch during a critical event (respond)
3. Evacuating population towards a safe area (respond)
4. Closing lanes of road, bridges, underpasses (respond)
5. Delivering timely information to public transport passengers, citizens, etc.
6. Suspending or redirecting public/private (transport) services (anticipate, respond)
7. Investing in infrastructure resilience improvement (anticipate, learn)
8. Improving organizational functionality (learn)
9. Training population (anticipate, learn)

The efficiency and effectiveness of these decisions are affected by the systemic informational and knowledge gaps in the system experienced by operators and decision makers in their daily activities. Most of these decisions require a reach and comprehensive informative layer behind to be reliable and produce a positive impact at systemic level.

In particular the drawbacks emerged by analysing scenarios like water bomb, yards, river and flash flooding, bomb attack (Bellini at al., 2016) (D2.2) suggest the need of managing static as well a dynamic/real time information to cope with unexpected events and support better decisions. Moreover, the zoom needs to be tuned according to the magnitude and the geographical extension of the event avoiding any loss in data granularity and quality.

The scope of RESOLUTE project is to operationalize resilience building such an informative layer to bridge the knowledge gap and to support data driven decisions.

2.4 Identify the data to for resilience management needs

According to strategy and the informative stakeholder needs, emerged during the Florence and Athens workshops, have been identified several class of datasets that can be exploited to enhance situational awareness through CRAMSS and the system resilience as a whole.

In Table 1, has been mapped the datasets identified able to cover the contextual information needs. Contextual information includes data needed to frame the context and that are useful for all kind of decision makers (cross-operators common information). In Table 2 has been mapped the kind of decisions resulted from the user requirement investigation and analysis carried out in T5.1 with the operators, and the dataset able to support that decions. In Table 3 the mapping between the information needed and the ERMG funtions related to that information needs are reported.

Table 1 Contextual information – datasets mapping

Contextual information	Datasets from Annex A
Geographic base city structure (streets, areas, etc.)	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, A5
Location and magnitude of event	F61, F62, F63, F64, F65, F66, F67, F68, A8
Time of critical event, recovery time (estimate / time of arrival of units)	F51
Passengers/ civilians: position, number (in system /train estimate according to peak / off-peak time), Condition: condition of injured people (can they still walk?), number of passengers /trapped, injured, passengers with special needs (e.g. wheelchair etc.)	F17, F18, F19, F45
Meteorological information (forecast from models, intensity from sensors, etc.)	F59, A7
News/event (planned and unplanned) updates: Where are strikes, marathons, car accidents, etc happening in the city and that affect the UTS	F37
Roles and contact list: general Information of responsibilities, Jurisdictions, information graphs: chain of command, who is the process / maintenance owner of certain parts of the UTS (e.g. drainage on streets).	Produced in a separate file

Table 2 Decisions – Datasets Mapping

Decision Maker	Needs/Decisions	Specific Information needs	Datasets
Civil Protection, Fire Brigades, Police, Ambulance services	Which kind and how many units to dispatch	Number and skill of operators available (e.g. Volunteers) Magnitude and dynamic of the event	F58, F60-F68
	If special equipment is needed	Special equipment availability	F58, F60-F68

Decision Maker	Needs/Decisions	Specific Information needs	Datasets
	If increasing the level of resource availability (operators, means, ...)	Number and skill of operators available (e.g. Volunteers) Statistical data from past events Forecasts	F38, F39, F40, F45, F46, F59, F79, F80
Mobility dept.	Close lanes of road	- Traffic information (public transport), possible indicators: passenger volume (planned vs. actual; on which relations is reduced service; delays)*** - Alternative transport means* - Traffic information (individual traffic)- position, direction, congestion, quickest route to arrive to location [*planned*]***	F7, F8, F9, F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F72-F75, F77, F78
	Communication: Inform passengers (also on alternative routes), traffic police, etc.	Position of the people	F14, F16, F17, F18, F19, F26, F27, F32, F33, F36, F42, F76
Public Transport Service manager	Route of buses and changes	Traffic information (public transport) [*passenger volume; service degradation level; delays*] Alternative transport means *[kinds, availability..]*	F72-F75, A1, A2, A6
Police, Civil protection, City Manager	Evacuate people	Meteorological information (forecast from models, intensity from sensors, etc.) Utilities, hospitals, etc. (position, kind, operational readiness, resource availability like beds or bloods bags, contact information) Traffic information (public transport) [*passenger volume; service degradation level; delays*] Alternative transport means *[kinds, availability..]* Traffic information (individual traffic)- position, direction, congestion, quickest route to arrive to location [planned]. Energy supply status. Available from the grid-operator. *[Planned maintenance, outage, …]* General information about buildings: floor plans / escape plans / fire systems of	All the datasets

Decision Maker	Needs/Decisions	Specific Information needs	Datasets
		station / rescue ways / exits / shafts of tunnel, diagram or topographic visualization of all the underground lines / 3D-building-visualization Position of the people	

Table 3 Dataset and ERMG

Kind of information	ERMG
Information about the critical event	
Location and magnitude of the event	. Fundamental for respond related functions. . Facilitates response planning, resource allocation and deployment. . Also provide a basis for initial recovery plans and for mitigation actions
Time of critical event, recovery time (estimate / time of arrival of units)	
Type and dynamic of event, further information (direction of fire (in a tunnel), Special circumstances (e.g. hazardous materials, etc.)	
Position, number, type, condition (operational readiness), general and contact information of ground units (/civilians):	
Position of ground staff *[Police, Emergency medical services, Fire fighting, Civil protection volunteers)*	. Supports various "anticipate" and "monitor" capabilities . Knowing what resources may be available and how they may be deployed (with what efficiency and effectiveness) is fundamental, not only to plan, but also to identify adaptation needs or opportunities (i.e. improve on training or capacity or reallocate resources...) . Some of this information may only be required on the basis of intervals, rather than real-time or continuously. This type of information supports regular reviews of planning and assessment (i.e. risk and operational status assessment)
Persons in charge (liaison officers / water supply / manager of each station/scene): position and Contact information, incl. VHF channel	
Utilities repair units	
Passengers/ civilians: position, number (in system /train estimate according to peak / off-peak time),	
Condition: condition of injured people (can they still walk?), number of passengers /trapped, injured, passengers with special needs (e.g. wheelchair etc.)	

Number and skill of operators available (e.g. Volunteers)	
Information on the status of urban systems and general information	
Traffic information (public transport) [*passenger volume; service degradation level; delays*]	. This is important for **monitor** capabilities . Real-time indicators (i.e. operational status) should be kept separate from historic or performance indicators. The latter refers mainly to **learning capacities** and therefore addresses different guideline requirements.
Alternative transport means *[kinds, availability..]*	Public Transport plan refers to **strategic planning** functions
Traffic information (individual traffic)- position, direction, congestion, quickest route to arrive to location [planned]	Referred to **monitoring capacity**. Monitoring traffic operation and having early detection of any disturbances
Utilities, hospitals, etc. (position, kind, operational readiness, resource availability like beds or bloods bags, contact information)	This information flow needs to be establishes as the result of any prevention, readiness or response plan. For instance if the resources availability is known and forecasted then it is possible to manage the response capability through the right anticipation like the increment/ trade-off of resource availability.
Energy supply status. Available from the grid-operator.	Being continuously aware about the energy availability status in the city is part of the monitoring capability and it is represented in ERMG with the resource supply function. In fact even if the service of the energy importantly planned maintenance activities in the power grid or foreseeable shortages, and with as much advance as possible
Access to camera surveillance (stations and trains, surroundings), live and also recordings (requested by police and control room operator)	This information regard the monitoring capacity and basically are managed by the ERMG Safety and security
Meteorological information (forecast from models, intensity from sensors, etc.)	Having weather forecasts enhance the anticipation capacity while collecting data form sensors like humidity, pressure, rain, etc. in real time enhance the monitor capacity.
General information about buildings: floor plans / escape plans / fire systems of station / rescue ways / exits / shafts of tunnel, diagram or topographic visualization of all the underground lines / 3D-building-visualization	Most of these information exist but tends to be dispersed because of the data are generated/managed by diverse institutions, some of these data are not digital, others are in not standard format, etc. Identifying, collecting and integrating these data enhance the anticipate and respond capacity and in particular the emergency respond function.
Statistical data from past events (location, type of events: “What went wrong in last operations”. “Where and when are repetitive events or most likely to happen”?)	This is at the core of **learning** capabilities. Should not be limited to statistical data or categorised historic analysis. Access to full records on events are often needed to produce meaningful leaning experience.
News/event (planned and unplanned) updates: Where are strikes, marathons, … happening in the city	To be provided in relation to both monitoring and respond capabilities. This would respond to coordination needs

Social networks analysis (Facebook, Twitter), User Generated Content like pictures or videos uploaded by users on site. Validity /reliability of information provided	This could be useful as statistical source (i.e. estimated number of people in the area, or in contact with…), for event detection and for respond capabilities. Tracking social media data flow
Estimation of Reliability of the Information (for operators, it is critical to know how trustworthy an information is. That may depend on the source of the information: is it a trustworthy source?) *[E.g. “reputation”]*	The source of the information will speak for itself. The user should be in much better position to use information based on its source rather than an system estimation.
Role and contact list integrated with a messenger	
Roles and contact list: general Information of responsibilities, Jurisdictions, information graphs: chain of command, who is the process / maintenance owner of certain parts of the UTS (e.g. drainage on streets). Display hierarchy of decision-making? Who takes what decision? --> [Contacts list]	This information should be collected and continuously maintained up to date as a preparatory action. This could be done by the ERMG Risk Analysis function.

2.5 Identify data availability

Information and data sources can be official and unofficial. Official information/data are those made available by institutions and organizations like city council, public transport operators, private transport operators, civil protection organisations, etc. Some of these datasets are available in Open Data according to the e-gov trends while others that are more sensitive (e.g., position of the people on the ground), require specific agreements between the data owner/provider and the data aggregator. In particular, such agreements define the reusability licences, the data privacy and security constraints, the commitment of the owners in making the data available according to a Service Level Agreement, SLA, etc. However, not all the data useful for supporting informed decisions can be accessible. In fact, private organizations consider their data as a strategic asset for their business and tend to avoid their sharing. In other cases (e.g., Critical Infrastructures), data are covered by national security constraints (e.g., the energy network topology) and could be accessed only though public organizations like city council or civil protection. It is clear that, a relevant role in data identification and assessment has been covered by ATTIKO Metro and Florence City Council since they are the primary data owners/providers in the project and they have the political power to ask to 3rd parties (utilities, public offices, urban polices, industries working with public procurements) to access to the data useful for the pilots.

The unofficial data are generated by channels like social media that have an impressive message distribution speed that can be detected and used to integrate official information sources. The unofficial data are usually highly available (e.g., twitter data) through some given methods (APIs) but the challenges, addressed in the section 2.3 are related to actual capacity to manage their high volume, velocity, variety, and veracity and meaning.

However, in order to better understand the categories of data and their actual availability, the datasets have been classified basing on a set of features (see

Table 4):

- **Name**, name associated to the dataset, Category, one or more taxonomy voices. It is useful to classify the datasets in an easy way (Table 5)
- **Dataset dynamic**, field to classify the datasets from static to real time
- **Frequency update period**, it is connected to the 'Dataset dynamic' field and regards the frequency update period of the dataset. This field is fundamental because informs us about which is the rate that a reading process needs to maintain, to provide data to the platform.
- **Format**, format of the file related to the dataset
- **Interested Area**, geographical area to which the dataset refers
- **Data Provider**, Name and Url related to who (Authority, Municipality, other Private/public provider, etc.) provide the dataset
- **Data URL**, Url of the Provider (link to the place in which the dataset is available)
- **License/Condition of use**
- **Dataset description**
- **Data fields description**, description of each field when it is necessary (this field is not mandatory)
- **Status**, this field is fundamental to understand if the dataset can already be ingested or not

Table 4 Dataset metadata

Field	Admitted values	Description	Field (M-Mandatory, R-Recommended, O-Optional)
Name	textual	textual	M
Category	One voice from Taxonomy	**Taxonomy (Table 5)**	O
Dataset dynamic	S - P - RT	Static (S) Periodic (P) Real Time (RT)	R
Frequency update period	time period	Specifying the unit measure (s, min, h, day, etc.)	R
Format	csv, html, JSON, SHP, XML, SOAP, etc.	file format or API answer format	M
Interested Area	Place, city, region, etc.	area to which the dataset refers	M
Data Provider	Authority name + URL	Who provides data (authority + home page link)	M
Data URL	URL	Specific link where the dataset can be obtained (Web service, FTP, HTTP, etc.)	M
License/Condition of use	Specify the license/ condition of reuse		R
Dataset description	text	synthetic description	M
Data fields description	text	column description	M
Status	Available, Under production		O

Table 5 Resolute Dataset Taxonomy

Resolute Dataset Taxonomy
Infrastructure
Urban Transport & Mobility
Events
Social
Environment
Healthcare
Service & POI
Human Behaviour

The dataset classification phase is fundamental to establish: what kind of data is available, what domain area they cover, what kind of actions can be performed at high level once the dataset will be ingested in the RESOLUTE infrastructure and available for other purposes (via Big Data analysis, statistical analysis, previsions, etc.), what kind of ingestion (and management) priority must be assigned to each single dataset, etc. It consists of a deep and wide search of the relevant data sources (local, regional, national and international sources):

- The large publication of OD (Open Data) has opened the path for the information sharing. Most of the OD are published by governmental organizations, in file formats such as html, xml, csv, shp, etc., and typically provide information that may present links to web resources.
- Open Data coming from PA contains typically statistic information about the city (such as data on the population, accidents, flooding, votes, administrations, energy consumption, presences on museums, etc.), location of point of interests, POIs, on the territory (including, museums, tourism attractions, restaurants, shops, hotels, etc.), major GOV services, ambient data, weather status and forecast, changes in traffic rules for maintenance interventions, etc. Moreover, a relevant role is covered in the city by Private Data coming from mobility and transport such as those created by Intelligent Transportation Systems (ITS) for bus management, and solutions for managing and controlling parking areas, car and bike sharing, car flow in general good delivering services, accesses on Restricted Traffic Zone, RTZ, etc.
- Both Open and Private Data may include Real Time Data, such as the traffic flow measure, position of vehicles (buses, car/bike sharing, taxi, garbage collectors, delivering services, etc.), railway and train status with respect to the arrival, park areas status, and Bluetooth tracking systems for monitoring movements of cellular phones as people, ambient sensors, and TV cameras streams for security and flow.

3 DATA INGESTION

In general, all big data solutions must cope with data volume, variety, and veracity. Open data as static data are not the main source of information in the city. Most of the big data problems connected to the (smart) city platform for resilience monitoring are related to real time data as the vehicle and human mobility in the city, energy consumption, health care, and IOT. The city level architecture should be capable to take advantage of huge amount of big data coming from several domains, at different velocity for exploiting and analysing them for computing integrated and multi domain information, making predictions, detecting anomalies for early warning and for producing suggestions and recommendations to city users and operators.

In the last years, many architectural solutions have been proposed with the aim of making data accessible, aggregated, usable, and exploitable, etc., and many of them failed in posing the basis for creating a smart city open environment for new and smart applications. It is obvious to state that, cloud and distributed systems approaches are at the basis of the big data solutions provided for smart city, as well as IOT solutions at the basis for collecting data from sensors and devices in the urban context. On the other hand, the urban infrastructure is much more complex, and the limited focus on only some of the above-mentioned aspects would create limitations not accepted for the city operators and for the city development of smart services. In Figure 1 is represented the schema describing the method to collect data in a distributed environment. Datasets and flows are collected through the pull method. Service operators, providers or IOT data brokers should make available their data through (public or protected) APIs or dedicated channels while the collector engine call them periodically looking for updates. The frequency of the call depends to the nature of the data. Thus real time data will be requested with a rate of seconds, while very slow changing data like streets map will be requested for being updated in the data layer only when necessary.

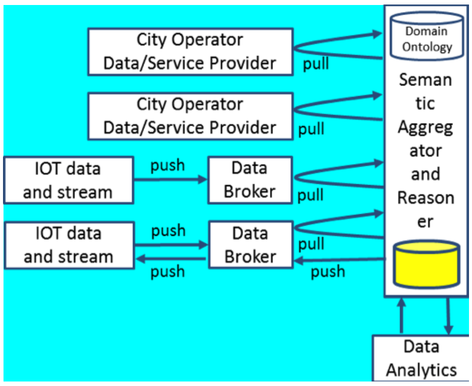


Figure 1 Data gathering schema

For the data ingestion, the problems are related to the management of the several formats and of the several data sets that may find allocation on different segments and domain areas of the Smart City Ontology. The solution has to allow ingesting and harvesting a wide range of public and private data, coming as static, semi-static and real time data as mentioned in the previous sections. For the case of Florence area, we already addressed about 150 different sources of the 1200 available, before start working on RESOLUTE project.

Static and semi-static data include points of interests, geo-referenced services, maps, accidents statistic, etc. This information is typically accessible as public files in several formats, such as: SHP, KML, CVS, ZIP, XML, etc. The most cases, the static and semi-static data sources are ingested using specific data transformation processes (one for each data source). Each Open Data ingestion process retrieves information and produce records in a noSQL Hbase for bigdata, logging all the information acquired to trace back and versioning the data ingestion. Data are then completed, other columns are updated dynamically with other process steps, and finally data obtained are placed on an HBase table.

Real time data includes data coming from sensors (e.g., parking, weather conditions, pollution measures, busses, triage, Wi-Fi, sensors, etc.) that are typically acquired from Web Services, REST Calls, as well as more static data as road graph description, etc. For example ingestion of data relating to traffic sensors consists of a ETL (Extract Transform and Load) transformation (e.g., Penthao Kettle) that invoke the web service via HTTP Post, retrieve the XML data and extract the data fields as *measurementSiteReference*, *measurementTimeDefault*, *concentration*, *occupancy*, *vehicleFlow*, *averageDistanceHeadway*, *averageVehicleSpeed*, *measurementSiteTableReference*, *supplierIdentification* and *publicationTime.* In most cases, the real-time data are directly pushed in the mapping process to feed the a temporary SQL store. They are typically streamed into the traditional SQL store and then converted into triples in the RDF final store. In alternative, in some smart city, real time data are streamed into a broker and from the it to some noSQL database.

In almost all cases, each single data set is ingested by means of a different ETL process defined by using Pentaho Kettle formalism because, among the several existing solutions, this formalism is quite diffused and also used by Information Systems Directorate of Florence, as well as by many municipality. When the Penthao Kettle ETL language and tool may presented limitation, external processes in Java or Python may be adopted.

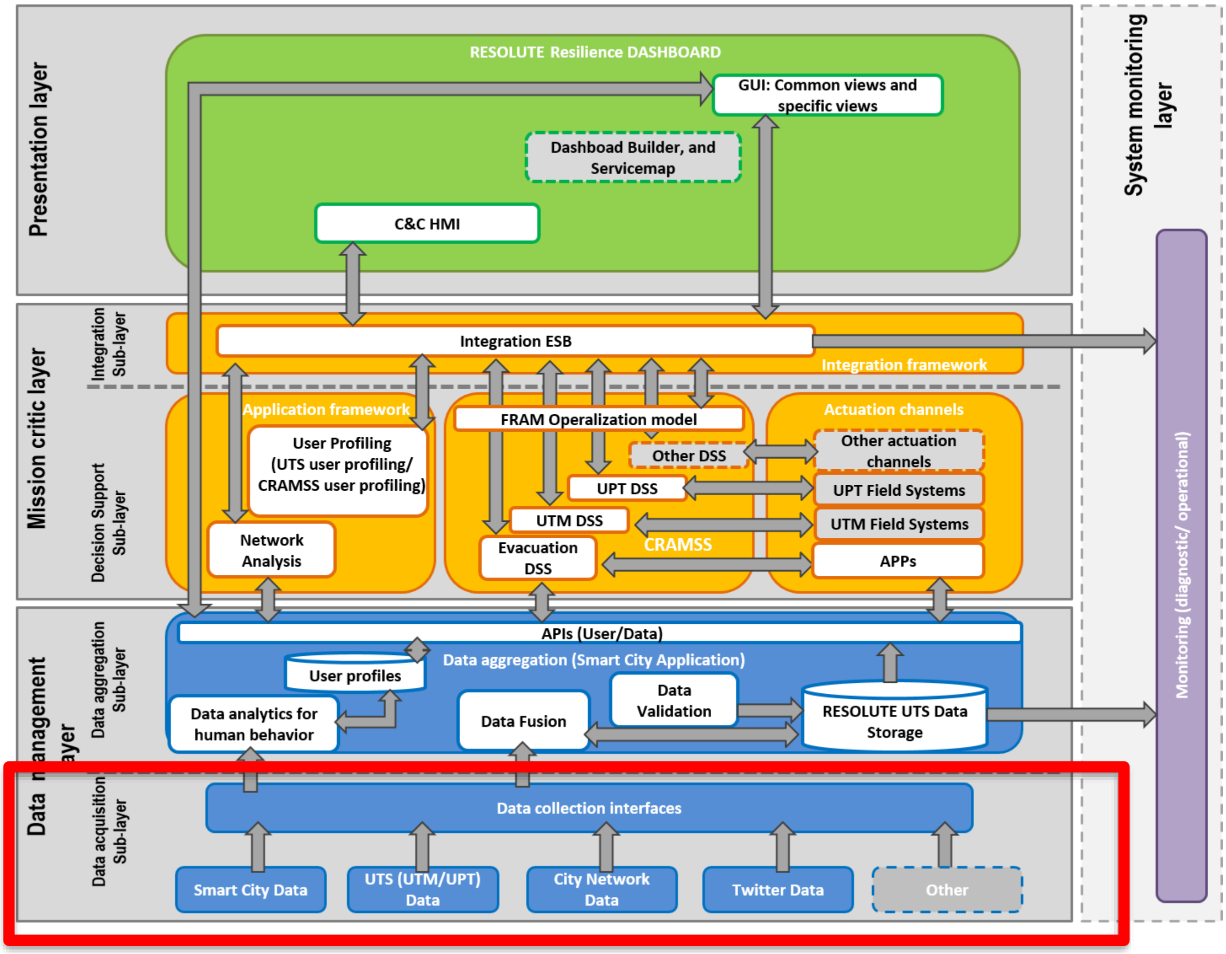


Figure 2 RESOLUTE Architecture with the data collection layer highlighted

Once established what the dataset useful to the purposes of the project are, a set of ETL processes are realised for their collection. Each dataset goes through various stages of processing, starting from ingestion up to aggregation and loading in RESOLUTE Knowledge base [km4city]. The process consists of the following sub-layers (see Figure 2):

<u>Data acquisition sub-layer:</u>

Ingesting Data from different kinds of sources (both Public and Private providers), necessarily involves solving aspects such as: variability, complexity, variety, geo-spatial aspects integration and size of these data (a "Big Data" problem). This problem can be partially solved by using specific reconciliation processes to make these data interoperable with other ingested and harvested data. The velocity of data is related to the frequency of data update, and it allows distinguishing static from dynamic data (real time data). The Dataset Transformation is necessary to realize the Harmonization and the Quality Improvement of the datasets. Different types of datasets can result in methodologies for ingestion of information. To better understand this aspect, the following categories of data have been identified:

- ***Smart City Data***: it consists of a collection of Open and Private Data coming from the city and territory. The major part of this kind of information is published by governmental organizations as Open Data, in different file formats such as html, xml, csv, shp, etc., and typically provides information that may present links to web resources. Moreover, the information is usually static, but can also be distributed in real time or semi-real time modality (e.g. number of visitors in a museum, city tour that are starting in a specific time and are going to make a guided visit of the city, weather forecasts for the different municipalities, events in the city, etc.). To acquire a wide range of these different kinds of data, above described, a set of ETL processes will be realized. The velocity of data ingestion is related to the frequency of data update, and it allows distinguishing static from dynamic data. In general data coming from the Public Administrations is available, typically covering:

  - location of point of interests (POIs) on the territory (including, museums, tourism attractions, restaurants, shops, hotels, etc.)
  - major GOV services: admin, schools, services, expenses, ordinances, etc.
  - environmental data: temperature, pollution, pollination, humidity, pressure, etc.
  - weather status and forecast
  - dataset strictly connected with the resilience aspects such as: i) position and status of underpasses, bridges, etc. ii) risks and vulnerability analysis; iii) relations with Critical Infrastructures ; iv) wi-fi data; v) flood levels; vi) landslides and earthquakes, etc. vii) accidents, flooding, presences at schools, etc.
  - etc.

- ***UTS (UTM/UTP) Data:*** it consists of a collection of a wide set of data regarding mobility and transport aspects in a smart city. In particular:
  - Intelligent Transportation Systems, ITS, for bus/train/ferry/tram/etc., management.
  - Public Transport plans and real time status
  - Traffic flows, people flows
  - Parking, car sharing, movements of public vehicles, ..
  - etc.
- ***City Network Data:*** it will collect a set of data coming from mobile network. This kind of channel, in a smart city context, it is fundamental to take care of the people/citizen's presence and needs. It can be useful to receive real time information coming from:
  - Cellular data collected from the mobile operators, that may collect data about the number of people connected into a quite large are in the city. The authorized monitoring are in the order of 1000 meter for side, and thus are too large for extracting detailed information about city user movement into medium and small cities; they can be used to understand wider movements.
  - the Wi-Fi or sensors networks having information related to the citizens' habits, collecting their activities and that are fundamental. For example to study the citizen's flows, to establish what are the citizens' preferred services, what can be improved for them, to collect their ideas and necessities, etc. Wi-Fi data provide results only about the their specific location, and the city is typically not systematically covered by AP, so that an high resolution is provided, but with surely small coverage in the whole area.
  - Special Apps that can be used to track the user behaviour and to perform measures on the city such as "Firenze dove cosa.." in Florence are available on all devices. In this case, monitoring may be performed only on the population installed the "city app" that would not be difficult harmonising the offer and the advertising. The advance would be to have a diffuse coverage and dense measures.

The data, collected through this component will be directly used and managed to elaborate new knowledge in the 'Data Analytics for human behaviour block' (in the Data Aggregation layer), origin destination matrices, extrapolation on larger areas, etc. This kind of data is sensitive so that has to be treated in a separate manner with respect to the classical data about infomobility, tourism, etc.

- ***Twitter Data***: it will collect and analyse data streams from social media (in the specific, Twitter data) through the creation of dedicated thematic channels, which can be tuned to monitor one or more Search Queries on Twitter with a sophisticated and expressive syntax. The twitter Data Module is realized thanks to the use of the following tools, already developed and at disposal for the RESOLUTE project:

- Twitter Vigilance (http://www.disit.org/tv): it continuously operates in two different working modalities at the same time (offline and real time http://www.disir.org/rttv). In the former mode, it accumulates Twitter data and performs different types of analysis, such as statistics and trends about messages and users, Natural Language Processing (NLP) and Sentiment Analysis (SA) of text messages, in order to provide analysis results, temporal trends and statistics (at level of channel and search, users etc.), sentiment polarity at channel and search level for keywords, adjectives and verbs. By this way, the contextual knowledge of the RESOLUTE Data Storage will be enriched. In the latter mode, this component will perform a real-time analysis (both statistics/trends and NLP-SA) on specifically defined channels, in order to monitor and detect critical conditions and providing alert signals or other type of actions, suggestions etc.
- Twitter Vigilance Real Time, specifically realized for RESOLUTE and presented in draft at the ' Mugnone 2016 ' simulation, studied and created in collaboration with the Civil Protection of Florence, is at disposal in order to be taken as a sample for possible future actions related to resilience activities in the RESOLUTE context, (http://disit.org/rttv/index.php?p=chart_singlechannel&canale=mugnone2016).
- Twitter Sensing Module for Evacuation DSS: sensing continuously Twitter for detecting anomaly or emergency situations. The data collected are displayed as graphs using the k-partite method and the force directed method creating visual clusters where further analysis can be applied. For instance, root cause analysis of abnormal phenomena can be performed and/or even anomaly detection.

All these different types of information have to be collected via the data collection interface and available for the next phases related to the Data Aggregation Sub-layer.

Data Aggregation Sub-layer:

- ***Data Analytics for Human Behaviour***, it makes analysis *on the* data coming from the ***City Network Data*** phase, related to *User Profiles and* Behaviours, such as:
    - flows of people configuration from Apps tracking and from wi-fi among geographical clusters
    - active citizens;
    - users or profile activity;
    - time of the day activity;
    - engagement modalities;
    - recommendations (sent, viewed, disliked):
        - automatic production of recommendations for users of the App, machine learning solutions for users' behaviour;
        - users' behaviour trends, general statistics, single user statistics, etc.;
        - heat map for category of city users (e.g., citizens, tourists, students, operators,..), etc. dynamic maps are available, many types with parameters.
- ***Data fusion:*** it works on the datasets that have been previously ingested, thanks to the following phases:
    - *Quality Improvement*, it is fundamental because the datasets ingested can contain a set of different errors and inaccuracies. For this reason, it is necessary to increase the quality of the datasets, if needed, to produce reliable and useful information for next applications, considering the following aspects: completeness, consistency, accuracy, absence of duplication, integrity, etc.
    - *Reconciliation:* this task solve the lack of coherence among indexed entities referring to the same concept but coming from different data sets. Entities may present mismatch in semantics

on names of the elements, dates, GPS coordinates, emails, telephone numbers, area codes, etc.

- *Triple generation and Aggregation*: this phase generates RDF triples from QI data for every dataset that has been processed. The triples created are stored Resolute Semantic Data Storage and are based on a model built on relationships defined within referring to a City Ontology (that will be realized starting from the multi-domain and Open Source Ontology Km4City, http://www.disit.org/6506) with the purpose to semantically integrate all the different information coming from the Data Acquisition sub-layer in an RDF Data store. The model proposed, makes use of the Km4City Ontology ([km4city]) and of an RDF store, integrating all information coming from the different datasets and allows applications to use the data in order to provide new advanced services to the citizens and city administrators exploiting inferential capabilities. The Km4City Ontology (Km4City, http://www.disit.org/6506), models many aspects of a smart city: some of them are static (or quasi-static) data such as
  (i) road graph modelling the roads and their relationships
  (ii) public administrations,
  (iii) "services" that are present in the city (e.g., hospitals, car parks, Electric vehicles' charging columns, train stations, taxi areas, etc.) that are associated with the road graph and organized in a hierarchy,
  (iv) train/bus/tram/etc. stops, and the respective lines of transportation,
  (v) road sensors that are present on the roads.

  Moreover, dynamic information that change over time is also modelled, such as:

  (i) weather forecasts for the different municipalities,
  (ii) timeline and status/position of the bus/tram/train with eventual forecasts for the arrival at the stops,
  (iii) status of the car park lots (e.g., number of free places),
  (iv) traffic sensors values
  (v) events in the city
  (vi) Environmental data
  (vii) Status of the first aid, triage.

  The starting ontology model, has been enriched considering the following aspects related to the resilience activities: i) underpasses, bridges, etc. ii) risks and vulnerability analysis; iii) relations with Critical Infrastructure; iv) wi-fi flows of people; v) flood levels; vi) landslides and earthquakes, etc.

- ***Data validation:*** it is created to apply validation and verification techniques that allow checking the correctness and consistencies of the system and its data. Given the complexity of structure implemented for each dataset, a validation of the triples has been performed, to verify that the realized RDF stored is well-formed. This validation is executed by running a set of SPARQL queries that verifies the consistency of the fundamental constraints.
- ***Dataset Load/indexing into the Knowledge Base***, it loads the RDF triples on the RESOLUTE storage

After this kind of processes, the data enriched and aggregated will be available thanks to a system of API for the Mission Critic Layer.

4 RESOLUTE DATASET

4.1 Introduction

Taking informed decisions for managing system resilience may require datasets that simply do not exist at the time of the analysis. This might usually happen because that need has not emerged before. Sometimes the information required already exists in the organization but is fragmented (e.g., into different databases), and may be incomplete. In other cases, the data should be created from scratch. In this chapter, the kinds of data collected and used in the RESOLUTE are introduced. The work of recognition and selection carried out according to the strategy identified above (See Chapter 2), was about the real dataset related to the type of data here reported and has produced two detailed lists of data sets for Florence and Athens pilots (See Annex A and B). The result can be considered as a first recognition and systematization of the material available. During the project further datasets might be added or removed according to the feedbacks coming from the pilot execution (WP6)

4.2 (Infra)structure

The structural datasets are referred to tangible entities in the city. Data regarding infrastructure can be classified in static and dynamic according to their update rate and may refer to several different kind of groups:

- **Structural description**
    - street graph, for example: open street map (static, or quasi static since the ordinances in the city are changing every day a new work on the city is activated: open and close a street, change its features, etc.). The structural description also includes: direction, size, kind, reservation level if any, velocity max, etc.;
    - restricted traffic zone, that may change shape over time in the day;
    - physical elements: bridges, underpass, pillars, etc., and their status;
    - special areas: pedestrian areas, gardens, parking, etc.;
    - emergency related data: people gathering areas, waiting areas, shelter areas;
    - etc.

In Figure 3, a GIS map composed by shape files of the streets, green areas, building areas is shown in overlay as an example of the datasets available from Florence Open Data portal[1].

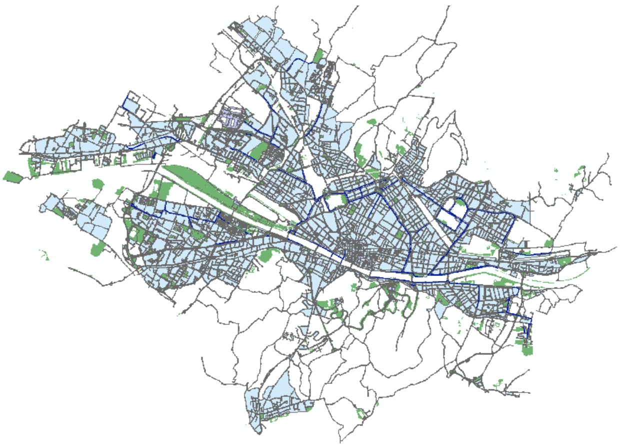


Figure 3 Map of infra-structure of the Florence city

[1] http://opendata.comune.fi.it

The dataset of streets once integrated and enriched with other data in the datalayer present the following information

- dct:**identifier**
  RT04801711064ES
- http://www.disit.org/km4city/schema#**composition**
  carreggiata unica
- http://www.disit.org/km4city/schema#**elemLocation**
  a raso
- http://www.disit.org/km4city/schema#**elementClass**
  urbana di scorrimento
- http://www.disit.org/km4city/schema#**elementType**
  di tronco carreggiata
- http://www.disit.org/km4city/schema#**length**
  33.0 (mt)
- http://www.disit.org/km4city/schema#**operatingStatus**
  in esercizio
- http://www.disit.org/km4city/schema#**speedLimit**
  30 (km/h)
- http://www.disit.org/km4city/schema#**trafficDir**
  Tratto stradale aperto nella direzione negativa (da NOD_FIN a NOD_INI)
- http://www.disit.org/km4city/schema#**width**
  > di 7,0 mt

Other attributes are:

Type: Road

PlacedinElement: services and other elements semantically connected to the street arc.

Structural data are usually static with quite rare updates managed by city council or at level or province, or at level of region, as administrative sources. They are usually used as a background for analysis (e.g., vulnerability analysis of road network of Task 4.4) or calculation (e.g., evacuation rooting of Task 5.2). Dynamic data communicated daily or even more frequently can be associated with:

- changes in the street graphs;
- status and shape of the restricted traffic zone;
- status of the underpasses that may be closed for critical conditions;
- etc.

4.3 Urban Transport and Mobility

Data regarding transport mobility can be classified in static and dynamic and may refer to several different kinds of groups:

- **Public transport** information, with: busses, trains, ferry, metro with their paths, stops, schedule, and real time position/delays with respect to scheduled timeline. Avery day a change may be performed. In addition, in case of special events or disaster changes are dynamically applied to reconfigure the paths and schedules.
- **Traffic flow sensors**, estimating: velocity, density, etc., for a number of categories (estimating flow including private and public vehicles)

- Network of **red lights**: position and status, in the view of Connected Drive solutions. They may be controlled in some cases by busses and/or public vehicles to control the green and reducing the red light stops for busses. The temporal setting and sync among them may change according to the plans defined at level of municipality to cope with changes from ferial and week end, from regular and large events in the city, etc. In the future, according to the Connected Drive paradigm the red light would be capable to provide information at vehicles about the next change of light colour and time to do it.
- **Public vehicles**: planned path, real time position, etc.; for example for garbage collection, ambulance, municipality vehicles, police vehicles, civil protection vehicles, etc. Some of them have a regular time line schedule;
- **Traffic flow reconstruction** on the rest of the city on the basis of the scattered data collected by sensors, by busses velocity, by other monitored vehicles with some On Board Device (OBD, vehicular kit), such as those of the AVM/ITS, or by using the mobile APP movements (see for example those collected from Google Map, Tom Tom), etc.;
- **Traffic flow trajectories**: preferred or typical paths in the city on the basis of the punctual scattered estimations or on the basis of On Board Device (vehicular kit), such as those of the AVM/ITS, or by using the mobile APP movements (see for example those collected from Google Map, Tom Tom), etc.
- **Origin destination matrixes**: describing the probability that a given flow of vehicle in a point of the city would be directed to arrive a second point of arrival of the city. For example, the 20% entering in the city from the highway are reaching the Time Square within 1 hour.
- **Parking status**: which may communicate the number of free slots.

4.3.1 Public Transport information

Mobility based datasets are related to the information about traffic status, mobility strategies, real time bus position, bus lines and time tables, etc. Since the domain is still not well standardised mobility data are released in a number of different formats. A first standardization attempt is represented by the General Transit Feed Specification (GTFS)[2], a de-facto standard promoted by Google to display static as well as dynamic (with GTFS extension) information on a map. The GTFS data model is represented in the Figure 4. GTFS "feeds" allow public transit agencies to publish their transit data and developers to write applications that consume that data in an interoperable way. The feeds are represented in a series of text files that are compressed into a ZIP file, and include information such as fixed-route schedules, routes, and bus stop data. GTFS datasets are used in a variety of types of applications, including trip planners, such as Google Maps, mobile applications, timetable generation software, tools for transit planning and operations analysis. In Km4City of Florence, they are ingested for the whole Tuscany area.

[2] https://developers.google.com/transit/gtfs/

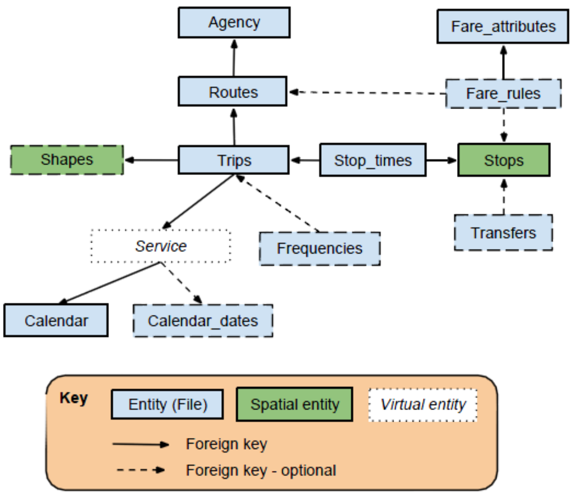


Figure 4 GTFS data model

An example of stop_time file is here provided:

trip_id	arrival_time	departure_time	stop_id	stop_sequence	shape_dist_traveled
1893_156527	13:42:00	13:42:00	FM9004_103	1	0

- **Trip_id**: Contains an ID that identifies a trip. This value is referenced from the trips.txt file.
- **arrival_time:** Specifies the arrival time at a specific stop for a specific trip on a route. The time is measured from "noon minus 12h" (effectively midnight, except for days on which daylight savings time changes occur) at the beginning of the service date
- **departure_time:** Specifies the departure time from a specific stop for a specific trip on a route. The time is measured from "noon minus 12h" (effectively midnight, except for days on which daylight savings time changes occur) at the beginning of the service date.
- **stop_id:** Contains an ID that uniquely identifies a stop. Multiple routes may use the same stop. The stop_id is referenced from the stops.txt file. If location_type is used in stops.txt, all stops referenced in stop_times.txt must have location_type of 0.
- **stop_sequence**: Identifies the order of the stops for a particular trip. The values for stop_sequence must be non-negative integers, and they must increase along the trip

The Public Transport System data in Florence (lines, time tables, bust stops expected time, etc.) published with such formats come from the operators of tramvia (GEST), and buses (CAP, BLUBUS, PIUBUS, ATAFLINEA, AMVBUS, etc.) at local and regional level. The data are collected periodically and each time a change is applied. The publication of real time position of the vehicles is less standardised and each operator tends to create its own data stream, a standard such as DATEX II could be used for providing such as information. Then to ingest such data into the data layer a dedicated ETL for each data source is needed.

In Athens similar data are provided by the OASA portal[3] where a number of RESTful APIs are published. In particular it is possible to retrieve Lines, Routes,, Stops, etc.

For instance the API Endpoint: http://telematics.oasa.gr/api/?act=webGetLines

Send back the following response in JSON:

[3] http://oasa-telematics-api.readthedocs.io/en/latest/

[{ "Line_Code":"815", "Line_ID":"021", "Line_Descr":"\u03a0\u039b\u0391\u03a4\u0395\u0399\u0391
\u039a\u0391\u039d\u0399\u0393\u0393\u039f\u03a3 - \u0393\u039a\u03a5\u0396H", "Line_Descr_Eng":"PLATEIA
KANIGKOS - GKIZI" }, ...]

The OASA portal also presents real time information via APIs as getBusLocation. An example of the data obtained with the following APIs call: http://telematics.oasa.gr/api/?act=getBusLocation&p1=2045 is provided here below:

[{"VEH_NO":"30885","CS_DATE":"Mar 6 2017
11:25:07:000AM","CS_LAT":"37.9385080","CS_LNG":"23.6328230","ROUTE_CODE":"2045"},
{"VEH_NO":"20470","CS_DATE":"Mar 6 2017
11:29:00:000AM","CS_LAT":"37.9446020","CS_LNG":"23.6663980","ROUTE_CODE":"2045"},
{"VEH_NO":"20582","CS_DATE":"Mar 6 2017
11:29:19:000AM","CS_LAT":"37.8477270","CS_LNG":"23.7534250","ROUTE_CODE":"2045"},
{"VEH_NO":"20585","CS_DATE":"Mar 6 2017
11:29:16:000AM","CS_LAT":"37.9151060","CS_LNG":"23.7067670","ROUTE_CODE":"2045"}]

In particular, given the route code (2045) it is possible to have the GPS coordinates (lat, long) of the vehicles with the time stamp.

Unfortunately, the data models used for the JSON response are not in a standard GTFS format and thus cannot be reused the same ETL created to ingest the Florence dataset. A huge work for data interpretation and reconciliation would be necessary during the ingestion process.

4.3.2 Traffic Sensors

As we said, in the mobility category also data from traffic monitoring are included, in particular real time data from sensors distributed in the city. Such sensors are georeferenced and are able to generate information as: Avg. Distance (average distance between vehicles), Avg. Time, Occupancy (the percentage of the average length of a vehicle respect to the length of the street arc of reference), vehicle Concentration, Vehicle Flow, Avg Speed. These measures can be expressed in terms of generic vehicles or for each kind of vehicle category. This kind of sensors can be spire in the ground, or laser solution as well as virtual spire realized by using tvcameras.

Usually these sensors interact with a traffic manager system (e.g., SWARCO TMS installed in Florence) that is able to collect data and propose to mobility manger some pre-calculated strategies to achieve mobility continuity in the city by regulating red lights. Indeed, such information sensed from the ground are relevant not only for mobility department but also for all the other actors involved in resilience management such as first responders, civil protection, etc. Thus collecting and integrating this kind of data with others coming from different control rooms (urban police, fire brigades, etc.) and from other kind of sources (user as a sensor, environmental sensors, weather conditions, etc.) try to bridge the informative gaps existing between the actual status of the system and the representation of the system provided by the monitoring system.

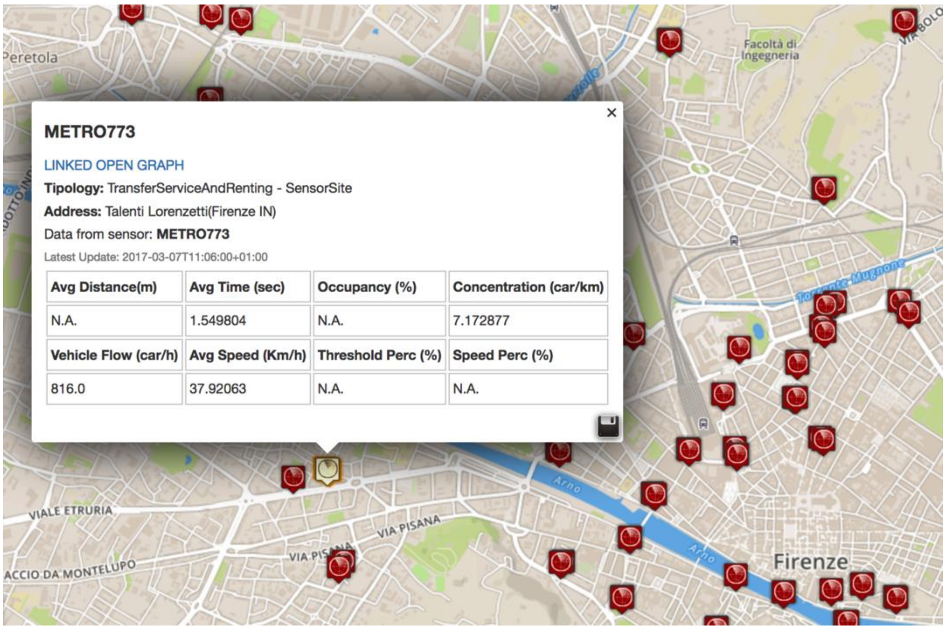


Figure 5 Traffic Sensor data displayed from ServiceMap

In Figure 5, a map with the traffic sensors georeferenced and connected to their data is represented. The level of precision about the status of the traffic in the city depends also to the density and distribution of the sensors on the ground. This means that investments in the hard infrastructures to monitor traffic dynamics in the urban context are necessary to obtain reliable and fine grain data.

4.4 Events of different kinds

In the category of **Main City Events** we can find pre-scheduled public events like exhibitions, sport (football match, marathons), manifestations, visits of political parties, etc. Knowing if a specific event is planned in a specific area of the city allows a better preparation in terms of safety and security of the attendees and mobility continuity for the city. For example, the city changes the organization of network red-light and the public transport scheduling according to the major event occurrence.

In a **Civil Protection Event** category one may also expect to have events like accidents, not planned strikes, flooding, heart quake, bomb attack, etc. These kind of events needs to be considered as a continuous stream generated by different sources that have the authority for transforming low level events (sensors and indicators levels goes over a given threshold, twitter trends,) into an official alert to be shared along the other operators (fire brigades, civil protections, police, etc.). Specifically speaking of the Civil Protection Alert they may manifest the possibility of having critical condition in the future: wind, snow, strong rains, strong temperature, etc. And thus they be used to prepare the population. In Florence are this information is distributed via several different Apps.

To the **Mobility Event** category is also associated the event dataset (F83). This data represents a list of the events updated in real time of what is happening in the city (of Florence) in relation to the mobility (e.g. car accidents, roadway reduction, etc.). The dataset is fuelled by Urban Police and Mobility dept. and it is managed by Citta' Metropolitana di Firenze.

An example of data element describing the presence of an incident happened at 12.58 and going to be solved after 2 hours in the in the lat/long 11.27133,43.76594 position, is reported:

<gml:featureMember>
<topp:infomob_eventi_di_trafficofid="infomob_eventi_di_traffico.fid-61bd4a22_15a94a7d9bb_2bd9">
<topp:id>1055867</topp:id>
<topp:descr>INCIDENTE UNTIL 03/12/17 AT 14:58.</topp:descr>

<topp:date_from>2017-03-12T12:58:51</topp:date_from>
<topp:inizio>12/03/2017</topp:inizio>
<topp:dataini_iso>2017-03-12 12:58:51</topp:dataini_iso>
<topp:date_to>2017-03-12T14:58:51</topp:date_to>
<topp:fine>12/03/2017</topp:fine>
<topp:datafin_iso>2017-03-12 14:58:51</topp:datafin_iso>
<topp:lastupdate>2017-03-02T14:55:02</topp:lastupdate>
<topp:severitycode>10</topp:severitycode>
<topp:icon>traff/dob_acc</topp:icon>
<topp:geom><gml:Point srsName="http://www.opengis.net/gml/srs/epsg.xml#4326">
<gml:coordinates decimal="." cs="," ts=" ">11.27133,43.76594</gml:coordinates>
</gml:Point>
</topp:geom>
</topp:infomob_eventi_di_traffico>
</gml:featureMember>

This XML reports an incident that started at 15.58 and finished (or will finish) at 14.58.

4.5 Social

The dataset derived by Social media are basically derived by social networks as Twitter, Facebook, etc. The data are usually collected through RESTFull APIs provided by the social network itself. In particular, since Twitter generates an enormous stream of information, it is crucial to define what you want to collect in advance through dedicated queries. Defining the right filter is critical to reduce the amount of useless data collected that can generate noise in the channel.

Tools as Twitter Vigilance[4] of DISIT lab, are able to collect, process and analysis such complex data source defining the filters (channels), analysing the streams with techniques of Natural Language Processing (NLP) to detect relevant variations in the channel in terms of numbers of tweet, sentiment, etc.. For instance in the Figure 6 is shown a relevant variation with a tendency to the negative sentiment in the channel “Meteo” created to monitor and detect weather related events.

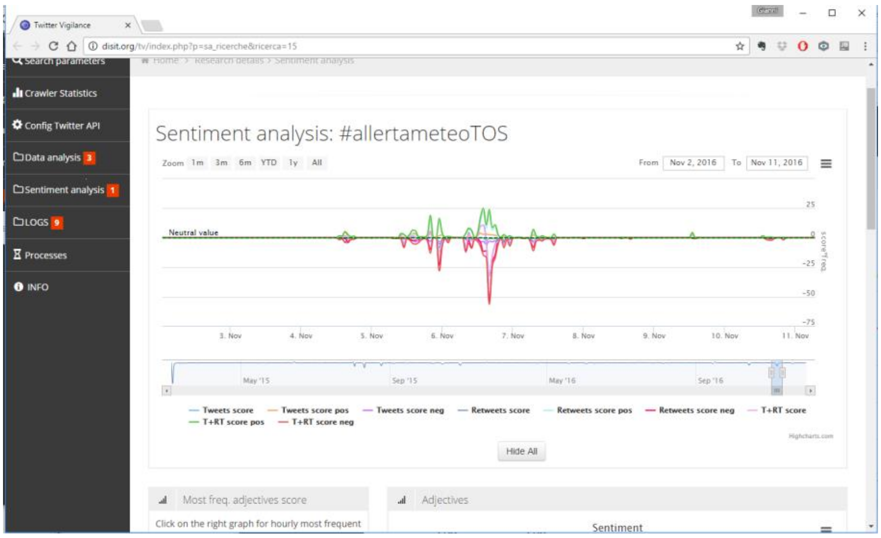


Figure 6 Sentiment Analysis signal processing

[4] http://www.disit.org/tv/

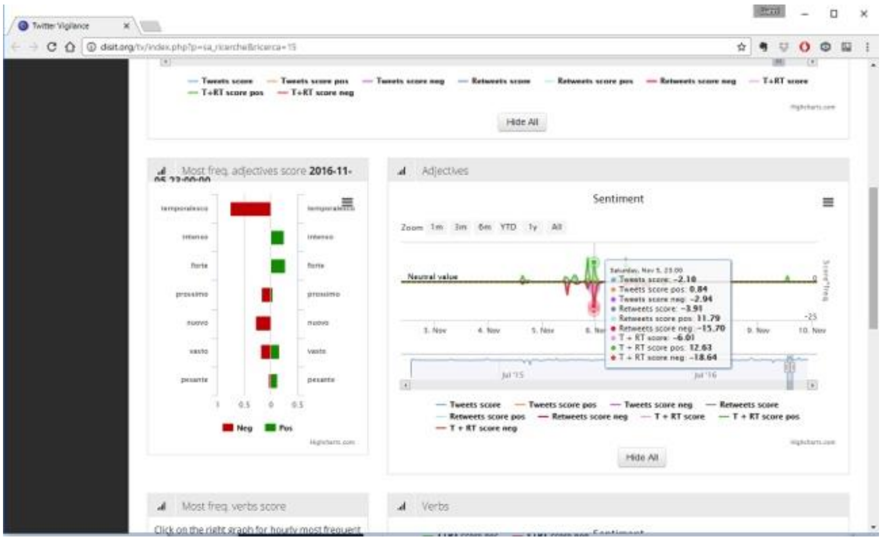


Figure 7 Sentiment Analysis terms exploration

The tool performs a deep analysis in the channel looking at the verbs, adjective, nouns and to detect the acceptation (negative/positive) of the terms (Figure 7). In the picture below, it is shown that to the term “temporalesco” (heavy raining) has been associated an issue (negative). In this case, the Twitter dataset was useful because was able to detect an issue related to the rain before the official authorities. In Figure 8, it is possible to see the effect of the alerting delay in the Arno river.

Figure 8 Arno swollen

On the other hand, this kind of data needs to be treated carefully. In fact, since social network based data are dependent to the presence of the users on the ground that act as a sensor, the absence of people when an event happens makes the tool not reliable for early warning. This is what happened in Florence during the Arno levee collapse caused by water aqueduct disruption (see Figure 9). The event happened at 6.15 AM of the 25/05/2016 but not relevant variation has been detected that that time because of there were not twitter users present to comment. Looking at the Figure 10, is it possible to see that the first relevant variation was detected in the afternoon of the same day and an increment has been detected in the subsequent hours and days with a numbers of retweet.

Figure 9 Arno Levee collapse - source Repubblica.it

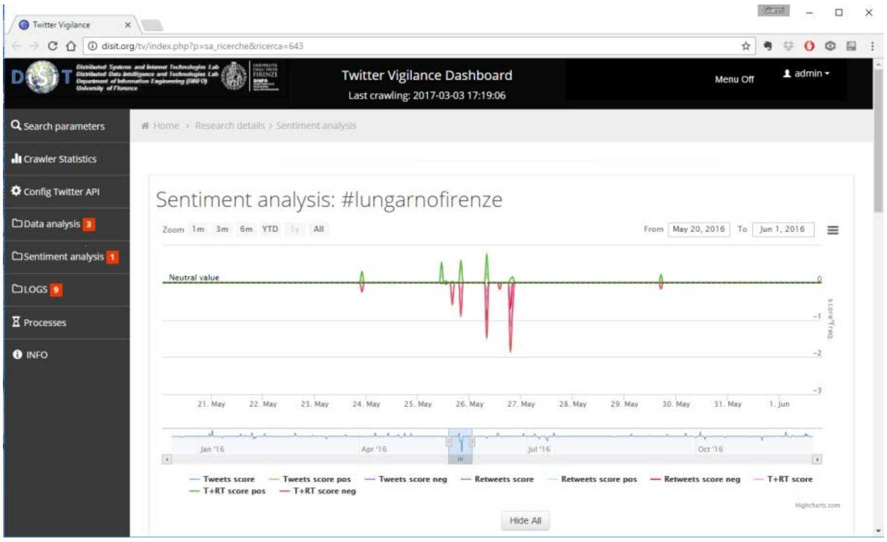


Figure 10 Sentiment Analysis on Arno Levee

According to this and other experiences (Grasso et al. 2016; Crisci et al. 2016), the social media data is a powerful tool that requires a deep understanding of its dynamics, its capacity of meaning and the method of how to extract such meanings. Social media data are collected with a dedicated architecture of distributed crawlers on a cluster of N nodes that perform data gathering and extraction by using Twitter Search API (see Figure 11). In the case of daily Twitter Vigilance, it is a multiuser tool that may be exploited by several authorities for monitoring different aspects and for different purposes. On the other hand, the real time version is much more computationally heavy, so that a special version for Florence only has been realized.

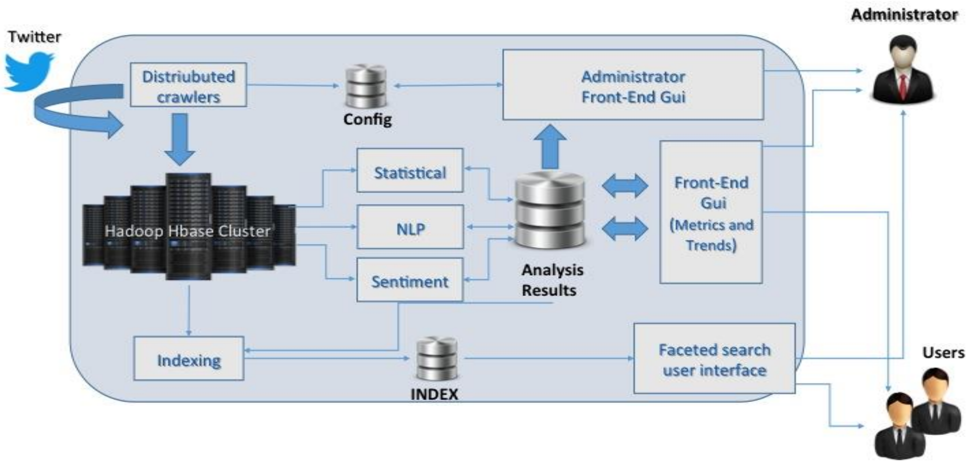


Figure 11 Twitter Vigilance social media collection architecture

In Twitter Vigilance, The data gathered from twitter are stored on Apache HBASE, a well-known column-oriented database management system that runs on top of Apache Hadoop which provides both highly scalable distributed filesystem and data processing framework, i.e., map-reduce. The data acquisition approach is based on the concept of Twitter Vigilance Channel, consisting in a set of simple and complex search queries which can be defined by a registered user by combining keywords, hashtags, user's IDs, citations, etc., in a structured logical syntax, according to the search syntax of Twitter. Both configuration parameters and statistical results are accessible from the front-end interface for the user. Collected tweets are made accessible to the back-office processes, which implement statistical analysis, natural language processing (NLP) and sentiment as well as general data indexing. Data processing is implemented as set of scheduled map-reduce jobs which read from HBASE table and store results in the Hadoop Distributed File System while data indexing is provided by Apache Solr operating in cloud mode and relying on the aforementioned filesystem. Solr is an open source search server based on the full text search engine Apache Lucene. The metrics resulted by the back-office processes are stored on a dedicated database and made accessible to the front-end graphic user interface, which allows customizing the search query, dashboards, reports and file export features (e.g., to Excel or CSV format) for visual analytics, temporal trends and time series visualizations, data results navigation, Twitter user's statistics and analysis. All these kinds of analysis are performed at both Twitter Vigilance Channel level and at single search level.

4.6 Environment

Dataset catalogued as Environment are those related to weather forecast, idrogeological risk areas, earthquake risk areas, basin level, weather forecast, river network, potential flooding areas, susceptibility flooding area, etc. In particular, in Florence most of these datasets are released by Arno River basin authority while weather forecast data are produced by Environmental Modelling and Monitoring Laboratory for Sustainable Development (LAMMA consortium).

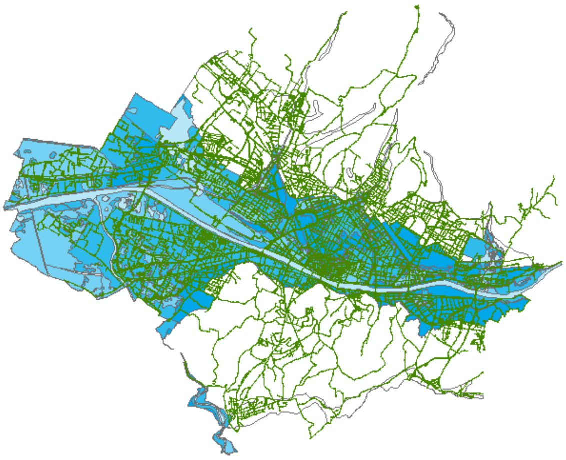


Figure 12 Idrogeological risk map of Florence

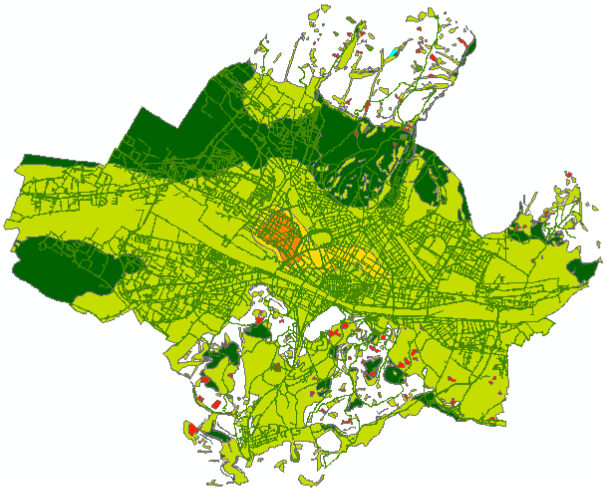


Figure 13 Seismic risk map of Florence

The data generated by the Sensors represent a critical flow that needs to be collected and processed continuosly. In fact, sensors are the tools able to detect the dynamics of the surrounded Environment. This class includes sensors to monitor the river level in different points of the river (Figure 14, Figure 15), the pluviometric data for a day and cumulated (Figure 16),

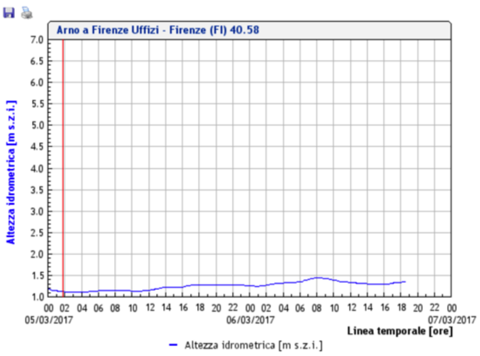


Figure 14 Arno river level at Uffizi check point

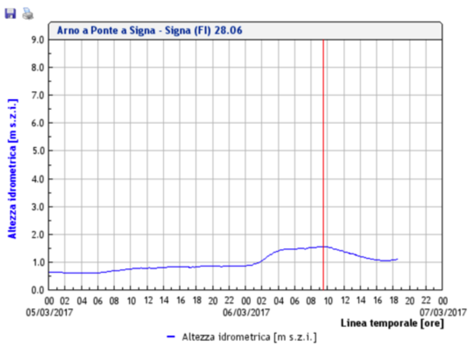


Figure 15 Arno river level at Ponte a Signa check point

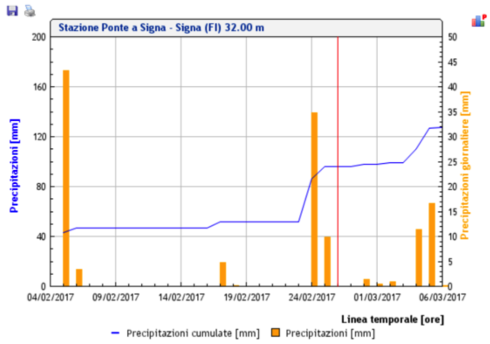


Figure 16 Rainfall levels per day and cumulated at Ponte a Signa

There are a number of sensors data available in the city of Florence (and in Tuscany region) like hygrometer, thermometer, anemometer, hydrometer that are collected in the data layer. The integration of such data combined to data coming from the mobility, allows a close monitoring of the dynamics and impact of extreme events on the ground. In fact, a water bomb has a different impact if it hits an urban area in which at the moment there is a heavy traffic flows respect to when the streets are quite empty. Thus knowing the magnitude of the event, the area affected and the impact on the ground allows better decision during and after the respond actions. There are other interesting environmental sensors data collected as the air quality. Real time information about NO2, SO2, H2S, CO, Benzene, PM2 and PM10 and the annual PM10 exceed count are collected in real time and integrated in the data layer and can be displayed through ServiceMap (see Figure 17).

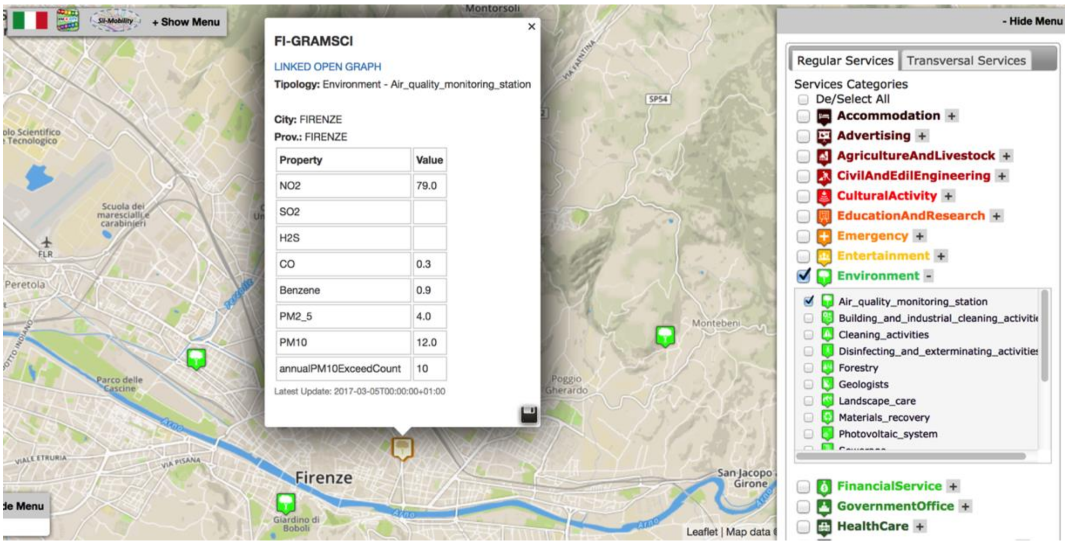


Figure 17 Air quality monitoring from ServiceMap Http://servicemap.disit.org

Monitoring the air quality is crucial in UTS resilience. In fact, the level of PM10 is become a critical parameter that can trigger a decision to suspend the mobility in a city for a while. Passing the PM10 maximum threshold has health impact for the population but also legal consequences for the city. Thus, a decision to block the traffic has a relevant impact on the mobility continuity and business, thus appropriate countermeasures (e.g. temporarily strengthen the public transport system) to mitigate the impact of such decisions, needs to be carefully evaluated and prepared in advance.

4.7 Healthcare

The Healthcare system can be considered a resource whose availability may affect the resilience of the system. Knowing if an emergency room of a given hospital is overloaded and it is publically accessible can make the difference during an emergency. In fact, the ambulances can be redirected towards hospitals with less loaded emergency rooms, or the authority can recommend to the citizens one emergency room respect to another according to its real time status. The RESOLUTE project aims at collecting from the emergency rooms in several hospitals the number and type of accesses in real time. The accesses are divided according to the colour assigned during the first triage (white, blue, green, yellow, red). The dataset gives also the information about a) waiting queue, b) visit in progress, c) patient under observations and d)patient with a destination assigned..

Pronto Soccorso
Ospedale del Mugello

Bianchi	Azzurri	Verdi	Gialli	Rossi	
0	2	0	0	0	In Attesa
0	2	1	1	0	In Visita
0	0	2	0	1	Oss. Temporanea
0	0	1	0	0	Con Destinazione
0	**4**	**4**	**1**	**1**	**Totali**
					Totale Accessi : 10

Moreover, all the public (hospitals) and private (e.g., Red Cros,) emergency management places are georeferenced and stored in the data layer. These datasets are available in Tuscany and area and have been addressed specifically for RESOLUTE integrating them into the KB and ServiceMap.

4.8 Services and Point of Interests

A city can be considered as composed by a number of services such as business activities, schools, hotels, banks, government office, cinemas, and so forth. In ServiceMap are georeferenced and classified to 20 different classes more than 80.000 unique services for Florence, 380.000 in the whole Tuscany. Such fine granularity allows gaining a better knowledge about the assets exposed to the critical events. On the other hand, some of those assets, during the emergency can be used as a resource to host population like hotels, schools, etc. Thus it is crucial to be aware of what, where and how many assets/services should be considered asset exposed to a critical events with a risk of damage and what, where and how many of them should be considered a resource. In particular areas with high concentration of services related to leisure time (shops, cinemas, restaurants) are representing the so called soft-target for terrorism attack but also element at risk during extreme events because of the presence of high number of people in specific moment of the day.

The Figure 18 is a screenshot of the ServiceMap application where are displayed a number of different services retrieved from data layer according to the category selected and a given distance (http://servicemap.disit.org).

This information has to be accessible for the operators, and better accessible for the whole population.

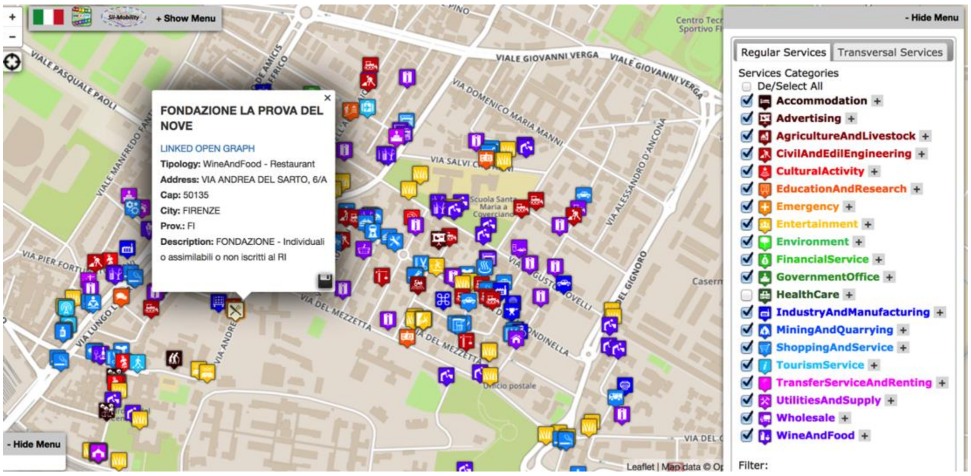


Figure 18 Service discovery on ServiceMap http://servicemap.disit.org

4.9 Human Behaviour

The capacity of monitoring the human behaviour in the city is a capacity that can enhance the resilience of the UTS, and in general of the City. In fact, the human being plays a key role in the effectiveness of resilience-based decisions. In order to track and understand where the people are, and how they usually move, to understand which is their usual habits as well as their behaviour during the emergency is very relevant. This information can be obtained by using Wi-Fi and mobile App obtaining into the city the needed resolutions and real time data. Indeed, it is not an easy data to obtain and to treat because of several technical preconditions and legal constraints.

The first precondition is related to the real existence of a wide enough Wi-Fi infrastructure (usually managed by public administrations) in the city accessible for the citizens and visitors for free, able to provide connectivity. The second one is related to the fact that such an infrastructure should be recorded and managed with a registry. Legal constraints are referred to the privacy issues but also the security aspect should be taken into account. The privacy issue come out since the log generated by the WI-FI track the mac address of the devices connected to the Wi-Fi. And through the mac address it is possible to know who is the owner of the device. In order to avoid such problem, specific anonymization strategies have to be applied directly at the source (log file stream), so that the information sent to the data layer is compliant with the privacy requirements. Anyhow a specific agreement between the users as data owner and the data aggregator (municipality) are needed to define the boundaries of data exploitation. The informed consent is one legal way as adopted in Florence. In fact, knowing where the people are moving in the city, which are their behaviour, etc., is a strong valuable data useful for a number of purposes like geo-marketing. From the city perspective, such data should be used for public utility, thus it is crucial to define the boundaries of the data exploitation.

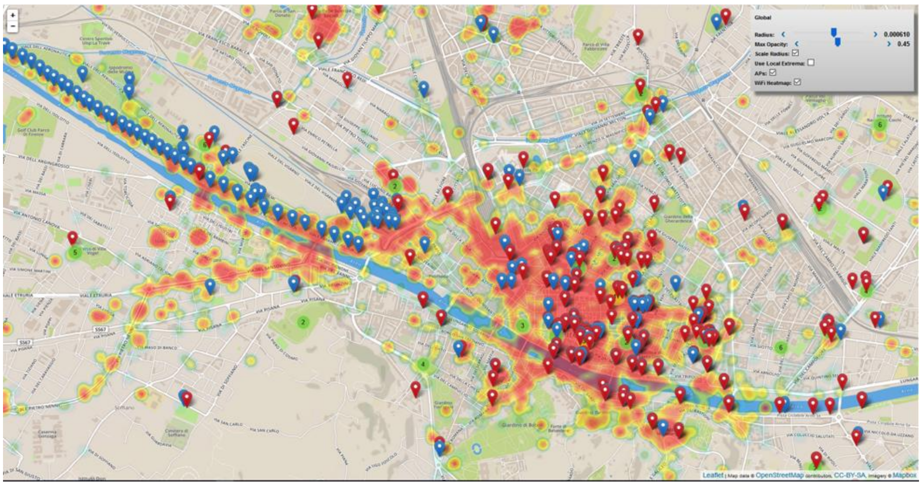


Figure 19 Heat map (gently provided by http://www.disit.org) Heat-map comparing city users' most frequented places vs the position of the 1500 Wi-Fi APs of the whole network (using a colour gradient scale to discriminate between different densities of measures)

A key aspect to understand human behaviour it the derivation of the so called Origin Destination matrix. This data allows to know how the people in a given point (monitored by an AP) are moving (in which percentage are moving) to other closer point in the city. A similar approach can be derived from movements of people by using mobile App data.

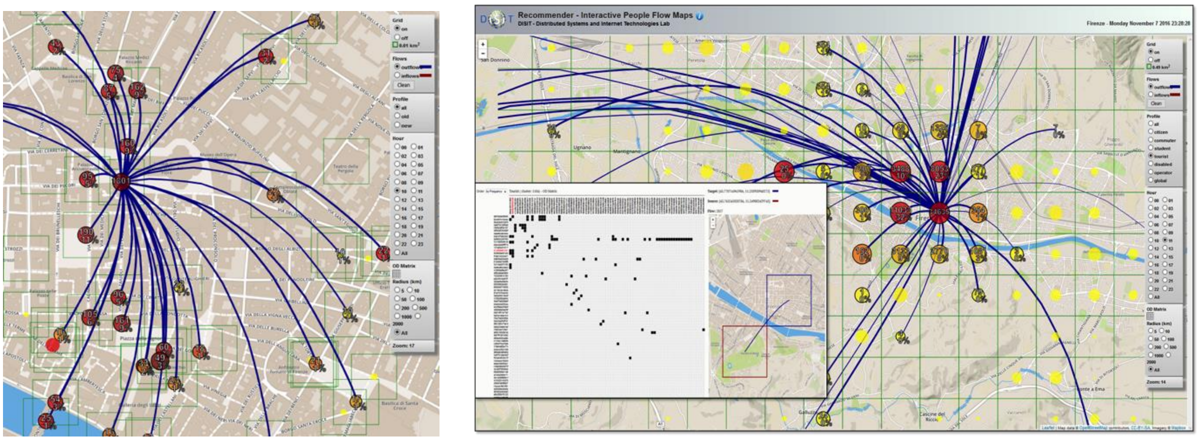


Figure 20 people spider OD matrixes gently provided by Http://www.disit.org (Wifi on the left and Mobile App on right)

Human behaviour can be analysed also generating classical Origin-Destination (OD) Matrix as a dataset for further analysis. In Figure 41 an example of OD Matrix is represented where on the y –axis is reported the Lat/Log of the Origin and on x-axis the Lat/long of the destination. This kind of dataset is important to understand the people habits in the city and how the city is „used". In this way, it is possible to anticipate and mitigate a number of potential risks that periodically emerge from the city usage.

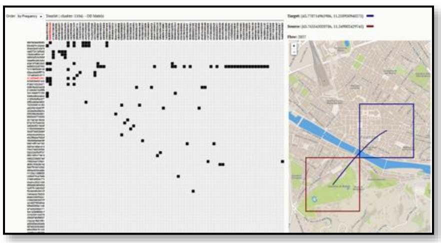


Figure 21 OD Matrix example

Another source of information that can be collected from mobile App or for vehicles by using OBD, are the typical trajectories in the city as depicted in the following figure. They are typically estimated for clustering the millions of trajectories collected in the temporal window in the area. Traced by using OBD on private vehicles are typically collected by insurance agencies. In Figure 22 are shown the trajectories calculated in Florence.

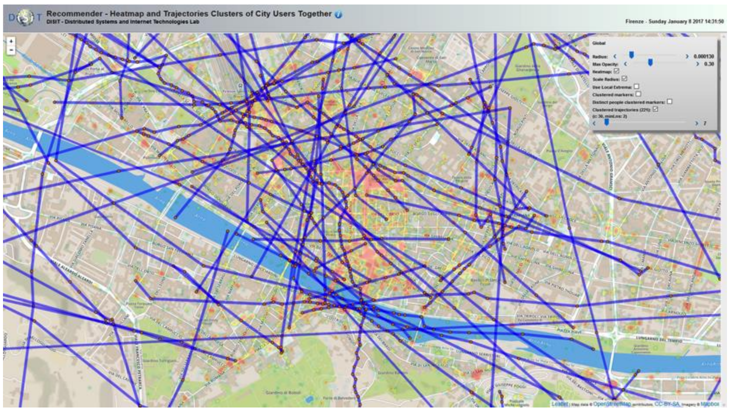


Figure 22 Trajectories representation

5 NEW DATASETS CREATED

For RESOLUTE, a number of data and data flows have been added with respect to what was in place in the smart city infrastructure (http://www.km4city.org). In particular, new data flows connected are:

- Wi-Fi Data (provided by City of Florence) collection and analysis (provided by DISIT lab), and corresponding data analytics for heat maps, trajectories, origin destination map, clustering of human behaviour, etc.; An extension of the Wifi network is also foreseen provided from the CMfirenze and Thales on the tramline of Florence.
- Social media data in real time as Twitter Vigilance real time (Extension of the DISIT Twitter Vigilance tool developed by DISIT lab) with corresponding data analytics: NLP and SA;
- Real time number of available beds in the emergency room data (provided by the hospitals in Florence), implemented by DISIT lab.
- Specific sensors data, as e.g., underpasses, not yet implemented as data ingestion process;
- Specific areas of the city as: Standing areas for population, recovery buildings, meeting points of rescuers and resources, assistance areas for population with the capability of extracting POI and other issues in the area as transport system facilities, etc.
    - Aree di Assistenza della Popolazione http://opendata.comune.fi.it/statistica_territorio/dataset_0373.html
    - Strutture di Ricovero della Popolazione http://opendata.comune.fi.it/statistica_territorio/dataset_0306.html
    - Aree di Ammassamento Soccorritori e Risorse http://opendata.comune.fi.it/statistica_territorio/dataset_0308.html
    - Aree di attesa della popolazione http://opendata.comune.fi.it/statistica_territorio/dataset_0307.html
- Flooding susceptibility areas
- Generic weather database
- Twitter base dataset for EvacuationDSS

In particular, these datasets have been created to enhance respond the capacity of the RESOLUTE system, in particular for the evacuation solution. This chapter describes the data connected to the data layer as the result for the pilot analysis (Task 6.1) workshops (DE7.2, 7.3) and user requirement analysis (Task 5.1 and DE5.1 and D.5)

5.1 Flooding Susceptibility Areas in Florence dataset

5.1.1 Motivation

Intense rainfall over an extended period and an extended area of the catchment basin may induce the rivers to overflow their banks or levees. The water can cover enormous areas. Downstream areas may be affected, even when they didn't receive much rain themselves.. When this happens, the event is called "flood", while if it takes place in uninhabited areas it can be simply defined "inundation" and the related risks are thus lower. There are several kind of floods. A flash flood is a very rapid and extreme flow of high water into normally low-lying dry areas. It may be caused by heavy rain associated with a severe thunderstorm, hurricane, tropical storm, or melt water from ice or snow flowing over ice sheets or snowfields on saturated soil or dry soil that has poor absorption ability. Flash floods are distinguished from regular river floods by a timescale of less than 6 hours (it begins within 6 hours of the causative event). As a larger area gets covered, the speed will be reduced. The water spreads out as much as possible flowing to the lower lying areas before slowly rising. Pluvial flooding occurs in urbanized areas when an extremely heavy downpour of rain saturates the natural or man-made drainage systems and the excess water cannot be absorbed. As there is little open soil that can be used for water storage, nearly all the rainfall needs to be transported to surface water or the sewage system. High intensity precipitation events can cause flooding when the city sewage system and draining canals (on surface or culverted) do not have the necessary capacity to drain away the amounts of rain that are falling. Water may even enter the sewage system in one place and then get deposited somewhere else in the city on the streets. In the first case the infrastructure is full to capacity (totally collapsed functionality) while in the second case the system works partially and tends to its saturation with the persistence of the fluviometric event. The water level raises relatively slowly and the water level usually does not reach life endangering heights, except in restricted and particularly concave areas with respect to the walking surface (basements, garages, road and pedestrian underpasses). **The latter aspects are usually undervalued and it is an important risk factor that the this dataset aims to mitigate when used in the RESOLUTE system.** In fact, the current urban identity of Firenze is the result of an intricate urbanization that is characterized by a dense overlapping and interlocking of constructions works, buildings and infrastructural elements of different areal extension and territorial relevance (such as railway lines, water distribution systems and other services etc.) which were embedded with each other over the centuries. The first great urban expansion that projects the ancient city confined by walls towards a larger and more modern urban dimension can certainly be linked to the period of capital of Italy (1865-1871) (Poggi, 1882; Mazzeo, 2011). After the Second World War, the urban growth had its new impulse, relying for decades on unscrupulous soil consumption (Black, 2009). In such context, a large number of grade separation junctions were realized with cuts and fills modifying the natural altimetric progression of the ground. All this evolution, if on the one hand introduced a better and faster connection between distant points, on the other hand the high number of “infrastructural corridors” contributed to increase the number of risky areas in the city that should be considered for better calculate the evacuation strategies

5.1.2 Hydraulic maps framework and new perspective

The level of attention for the river flood risk is rather high in the consciences of administrators, as well as in the mind of population since the hydraulic problems connected to the river overflooding have always accompanied the history of the city with sometimes-disastrous consequences (Morelli et al. 2014). Such menace comes primarily from its main watercourse, the Arno river, as consequence of its high values of flow: yearly averaged flow rate of about 50 m^3s^{-1} and maximum flow rate of 3540 m^3s^{-1} recorded on 4 November 1966 just upstream the metropolitan area (Campolo, Soldati, & Andreussi, 2003; Morelli et al. 2014). However, the problem of floods can also occur from the secondary drainage network within the urban fabric, not only for the flow level normally expected during a significant rain event, but especially for the artificial character that have assumed over time (narrowings, planimetric deviations, transfer in culverts, covering of entire sections and so on). For example, in August 1984, shortly after its completion, the Mensola creek broke the tubing in which it was contained and spilled over. For this kind of events detailed maps of vulnerability, hazard and risk have been realized by the competent authority involved in the assessment of flood risks (i.e., the Arno River Basin Authority).

Moreover, for some time now the entire metropolitan area suffers other natural occurrences linked to surface water circulation, which have similar consequences to the river flooding, but different dynamic of development. Such condition can be related to extremely heavy rains on restricted areas in a very short time (thunderstorm, downburst, etc.) that can cause a sudden appearance of excess water on the ground and consequently produce a so-called pluvial flooding in sunken areas (Figure 23, Figure 24, Figure 25). This kind of events occur each year with almost regular basis and with increasingly significant impact on the safety of citizens and tourists, but unfortunately, maps of Vulnerability, Hazard and Risk have not been produced inducing the local administration in managing case by case the contingent event. This lack is due to the high unpredictability of the rainfall events (mode of occurrence, quantification and distribution at the neighbourhood scale) using all the measuring instruments and forecasting models currently available. Also the constantly updated knowledge of the urban system capacity to intercept and effectively remove the amounts of rain through its individual or multiple components is still an open challenge. Consequently, it is not possible to define with certainty the probability that a potentially destructive event occurs in a given period of time and in a given area and maps of Hazard cannot be realized. Being unable to estimate the occurrence of a natural phenomenon of a given intensity even the Vulnerability maps, which are strictly dependent on the intensity factor, cannot be expressed. By the lack of these two basic elements, also the Risk cannot be estimated and described in a graphic map as in other natural events. Waiting for new equipment and methods able to provide more reliable forecast information on extreme rainfalls, we can concentrate on studying in detail the shape of the urban ground in its smallest aspects (microanalysis of the urban arrangement) in order to understand its predisposition to be submerged by floods. This is possible thanks to high detailed survey existing for the city of Florence.

Figure 23 Flooding of the road underpass in *via del Gignoro*, August 1, 2015 (street level totally below the surrounding ground level).

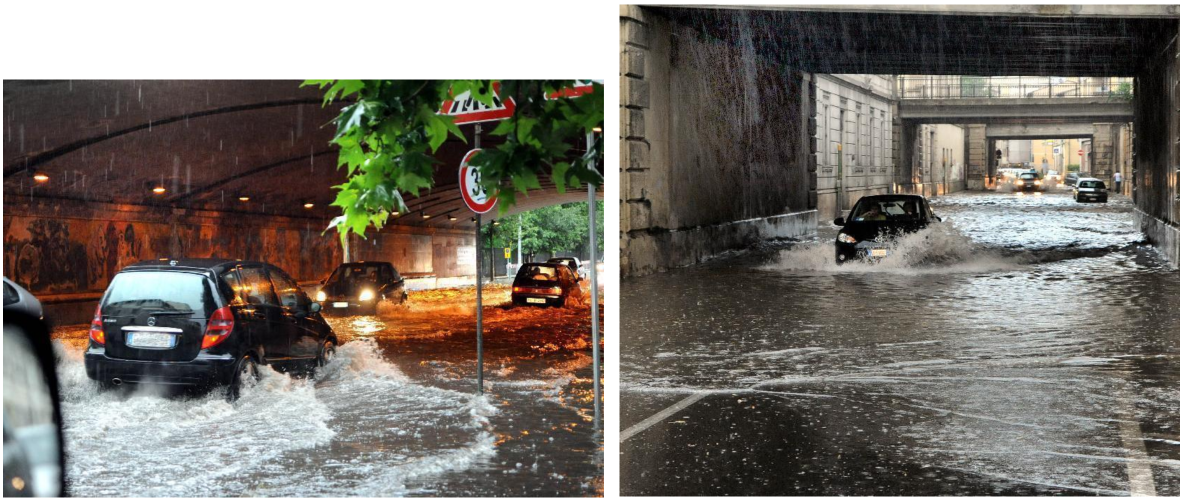


Figure 24 flooding of road underpass in *via Circondaria* (left) and in *via del Romito* (right), May 2, 2013 (street levels in planar continuity with the surrounding ground level).

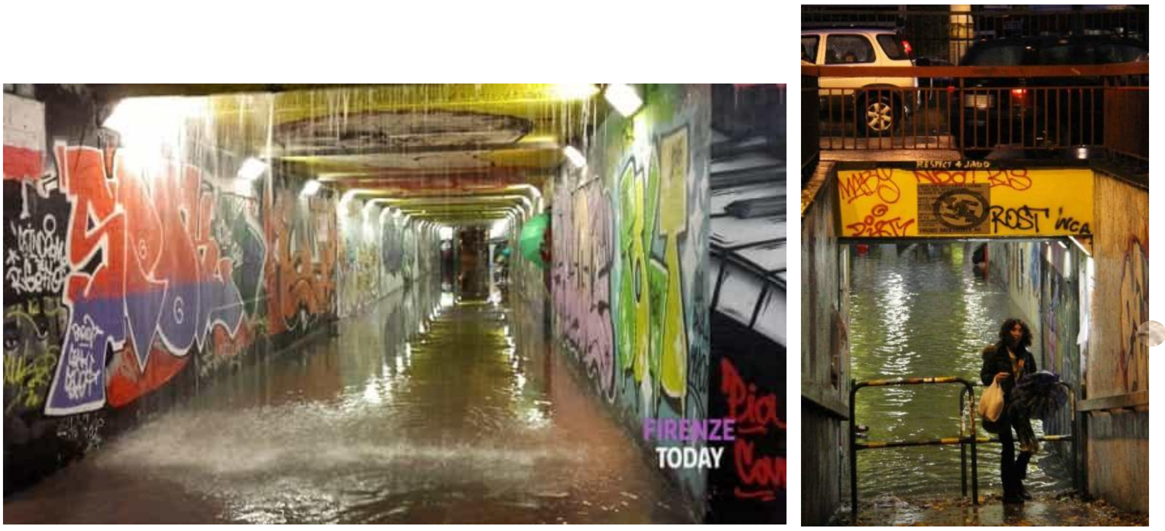


Figure 25 Flooding of the pedestrian underpass below the railway track and adjacent road junctions, October 2, 2016 (street level totally below the surrounding ground level).

5.1.3 Underpasses characterization

As demonstrated by chronicles of last years the most negative impacts of pluvial flooding on the human safety is related to the accumulation of water within the underpasses. Therefore a detailed mapping of these infrastructures has been performed in order to obtain reliable data for the creation of a terrain digital model suitable to the aims of the work and, at the same time, to provide strategic information for traffic management in real-time conditions and routine maintenance scheduling. Starting from a database of the municipality of Florence with minimal information inside (e.g., broad and punctual location of underpasses) we performed a field survey aimed to check the already mapped elements and also to map those not considered until now. Since our intention was to find, catalogue and describe with areas all underpasses in correspondence of grade separation junctions (also if not confined in the underground), a laser distance meter (Leica DISTO™D5) was used to measure them reporting immediately these geometric information in cartographic materials through fixed points, such as entrances structures, recognized in the available orthophotos (Figure 26,Figure 27). Simultaneously many other aspects were recorded in a summary table, accompanied, when possible, with critical information coming from various available sources (for specific details see the following section).

Figure 26 Field survey examples

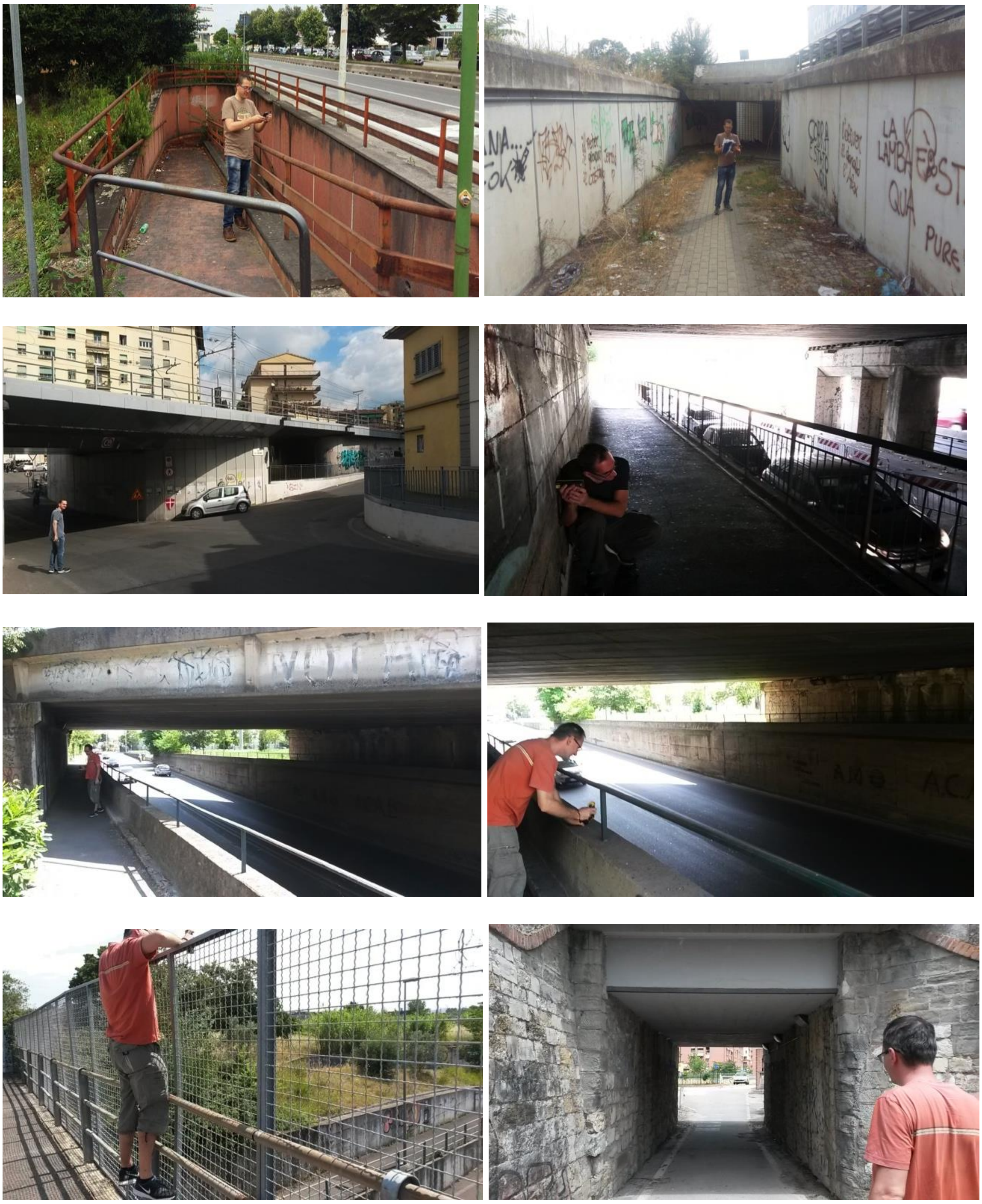


Figure 27 Field survey examples

5.1.4 Employed data

Many data have been used for the urban landscape characterization and the consequent pluvial flood modelling. These can be categorized into three groups: i) Base Official Data; ii) Data created on project purpose; iii) Derived data. The first group represents the existing official data that are available for consultation in different public administrations, the second has been created for this purpose, while the third includes derived data obtained by special processing of records coming from point i) and ii). Here a brief description of all these and a summary table are provided.

5.1.4.1 Base Official Data

The base data for this project are almost all Open Data provided from the municipality of Florence (http://opendata.comune.fi.it; http://datigis.comune.fi.it).

5.1.4.1.1 Digital Surface Model (DSM)

Data obtained merging official data acquired in a Lidar campaign (Light Detection And Ranging) funded by the City of Florence. Digital Surface Model is 3D Digital model that represents the surface of the city obtained by interpolating the data acquired by the laser scanner. It includes buildings, bridges, trees, cars and other objects that have reflected the laser beam.

5.1.4.1.2 Digital Terrain Model (DTM)

Data obtained merging official data acquired in a Lidar campaign (Light Detection And Ranging) funded by the City of Florence. Digital Terrain Model is 3D Digital model that represents the surface of the city without artificial elevations, like bridges, buildings, trees, cars or other objects. It theoretically represents the bare ground and is obtained by using special algorithms that try to “clean” these objects from laser data.

5.1.4.1.3 Orthophoto

Aerial photographs geometrically corrected representing the territory of Florence. Thes data are owned by AGEA (Italian agency for disbursement in agriculture) and distributed by the Tuscan region. They are used for reference only or for local corrections if necessary.

5.1.4.1.4 Technical map

Numerical cartography obtained merging maps from Regional Technical Map. This map, also known as GeoDB 10K, is a type of topographic map produced by regions of Italy to represent their territory. To every vector of the map a standard code is associated, to identify the source of the digitization (about 200 codes divided in 10 major categories: traffic network, buildings, hydrography, infrastructure, dividers and sustaining elements, geomorphological elements, vegetation, orography, administrative limits and toponyms).

5.1.4.1.5 City connective network

Numerical cartography acquired by the City of Florence in a celerimetric survey of the whole municipality. Celerimetric information is obtained with no direct measurement of distances, but through planimetric and elevation detection of the ground using tachymeter, theodolite and level staff. Data is divided in 9 categories: vehicular area, cycle area, pedestrian area, bridges, river banks, built up areas, retaining walls, trees and public parks. Where information relate the same type, this data is preferably due greater accuracy and more regular updates, in comparison to technical maps.

5.1.4.1.6 Road area network

Polygon elements that represent the Florentine public street areas, organized in road sections and junctions.

5.1.4.2 Data created on project purpose

5.1.4.2.1 Underpasses database

Polygon elements that accurately describe the planimetric shape of covered areas in correspondence of grade separation junctions. Due to their planimetric position in these special road/railway intersections, they can be called urban underpasses even if the running surface is in planar continuity with the adjacent areas. Their main features have been collected in a related attribute table.

5.1.4.2.2 Floods database

Vector representation of the floods that affected the City of Florence in the last 10 years. The base data was collected consulting fire department bulletins. This data was integrated by gathering all news about rainfall induced floods in the city. Every flood is classified by the day when it happens. For each reporting entry in the news bulletin, the flood area was digitized exclusively using the areas detectable in photos or videos found in the news. Where possible, additional information were included, coming from photos or videos related to the event, which were published in social networks.

5.1.4.3 Derived data

5.1.4.3.1 Digital Hydraulic Model (DHM)

Digital Hydrological Model is 3D Digital model that theoretically represents the surface of the city where the rainfall can flow freely. It can be assumed that most of the flow passes by streets and intersections. The main areas of water accumulation are inside buildings below ground level (basements, cellars, etc. or on the ground floor with rising water). These areas, once flooded, do not constitute a significant obstacle to water flow because each area is isolated from the others by walls and water transition through these areas is estimated to be much lower than the flow along the roads. The Digital Hydrological Model is the surface model used for flow-path analysis under the hydraulic conditions described above (Morelli et Al. 2014). Basically it represents the terrain with all and only the elements that can heavily block the water runoff, such as river banks, ground elevated roads and rails, buildings basements, and so on, cleaned by small or movable items, such trees or cars, but also cleaned by elements like bridges and overpasses that cannot block the runoff.

5.1.4.3.2 Slope Model

Slope is a number that describes the steepness of a surface calculated as the maximum rate of change between each cell and its neighbours. In general, the lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. The slope data was calculated as degree of slope.

5.1.4.3.3 Flow Direction Model

Theoretical direction of flow from every cell of the model. Flow Direction Model is used to deriving hydrologic characteristics of a surface. The direction is coded using the following schema: 1 = east; 2 = south-east, 4 = south, 8 = south-west, 16 = west, 32 = north-west, 64 = north, 128 = north-east.

5.1.4.3.4 Flow Accumulation Model

The Flow Accumulation is based on the number of cells flowing into each cell of the surface model. Output cells with a high flow accumulation are areas of concentrated flow and can be used to identify stream channels.

5.1.4.3.5 Flow Length Model

The Flow Length Model represents distance along the flow path for each cell. In a watershed, the flow length is the distance from any point in the watershed to the watershed outlet. Its primary use is to calculate the length of the longest flow path within a given basin. This measure is also used to calculate the time of concentration of a basin.

5.1.4.3.6 Depressionless Model

Depressionless Model is a Terrain Model free of "sinks". Basically it is the model of the surface where all possible lake and ponds are filled and water is free to runoff in whole model.

5.1.4.3.7 Depression Model

Depression Model is 3D Digital model that represents the "sinks" of the surface. Basically is the model of the depressions that can become lakes/ponds if the surface was crossed seamless by water. The value of every cell is negative because is the depth relative the surface of these hypothetical lakes/ponds.

5.1.4.3.8 Depressions Density

Density of depression cells in a neighbourhood around every cell of the model. It indicates how many local sinks zones are present for area units.

5.1.4.3.9 Permeability Map

Permeability Map represents the capability of the terrain to permit the flow of water through it. A simplified permeability map where every cell of the map is classified using the following codes: 0 = not permeable (paved areas, buildings, etc.); 1 = lightly permeable (mixed area, areas without information); 2 = heavily permeable (ground, treed areas, grassy areas, etc.).

5.1.4.3.10 Underpasses Map

Underpasses Map represents the presence and type of underpass in the city territory. Every cell of the map is classified using the following codes: 0 = no underpass; 1 = In planar continuity with the surface-level road; 2 = Partially below the surface-level road; 3 = Totally below the surface-level road.

5.1.4.3.11 Floods Density

Density of floods cells in a neighbourhood around every cell of the model. It indicates how many floods were registered for area units. Useful also for more intuitive visualization of the floods on map.

Table 6 Summary table of the employed data.

Datum	Nominal scale	Resolution	Type	Acquisition time	Source
DSM Digital Surface Model	1:10000	1 m/pixel	floating point 32-bit	November 2007	Official
DTM Digital Terrain Model	1:10000	1 m/pixel	floating point 32-bit	November 2007	Official
Orthophoto	1:10000	50 cm/pixel	RGB Color 32-bit	Flights ranging from June 2013 to July 2013	Official
Technical map	1:10000	-	Vector	Last update in 2011	Official
City connective network	1:500	-	Vector	2015	Official
Road area network	1:50000	-	Vector	Continually updated 2016	Official
Underpasses database	1:10000	-	Vector	2015-16	Created
Floods database	1:25000	-	Vector	From 2006 to 2016	Created
DHM Digital Hydraulic Model	1:10000	1 m/pixel	floating point 32-bit	-	Derived
Slope Model	1:10000	1 m/pixel	floating point 32-bit	-	Derived
Flow Direction Model	1:10000	1 m/pixel	Integer 8-bit	-	Derived

Datum	Nominal scale	Resolution	Type	Acquisition time	Source
Flow Length Model	1:10000	1 m/pixel	floating point 32-bit	-	Derived
Flow Accumulation Model	1:10000	1 m/pixel	floating point 32-bit	-	Derived
Depression Model	1:10000	1 m/pixel	floating point 32-bit	-	Derived
Depressions Density	1:100000	10 m/pixel	floating point 32-bit	-	Derived
Permeability Map	1:10000	1 m/pixel	Integer 8-bit	-	Derived
Underpasses Map	1:10000	1 m/pixel	Integer 8-bit	-	Derived
Floods Density	1:100000	10 m/pixel	Integer 8-bit	-	Derived

5.1.5 Technical description of data used in the modelling

Every raster elaboration was done using the municipal area of Florence as boundary, filling with neutral values (i.e. 0 or no data) if necessary. Note that raster data colours in images have not a linear relation with their values, but a stretching was applied to increase the visual contrast of the display and emphasize the data characteristics.

5.1.5.1 Underpasses database

The attribute table of the database was filled in with original information derived from several field survey (Figure 28) and technical analysis of official data in relation to the considered flood risk. Such data compilation and graphic reproduction allows us to have a kind of data immediately adequate for the modification of the original digital terrain model. The collected material can be summed with the following components:

- **Id** = identification number of the underpass.
- **Area** = area covered by the grade separation area. This includes pedestrian areas, transit vehicles and related pavements (if existing).
- **Type** = indicates the kind of structure in relation to main user: pedestrian and vehicular (this may contain also pavements for pedestrian circulation).
- **Vehic Type =** indicates the type of vehicle (only for vehicular type).
- **MCL** (Most Constrained Layout) = more rigidly fixed structure on the territory for which the possible structural modification works seem more complex from a technical point of view.
- **Base_level** = vertical position of the street level of an underpass (covered area) in relation to the external road network in areas around the entrances. Three conditions are considered: i) totally below the surface-level road, ii) partially below the surface-level road, iii) in planar continuity with the surface-level road. When there is height difference the two levels are joined by the ascent and descent ramps.
- **Heights_m** = height differences between the street elevation in the underpass at the entrances of the covered area and the elevation of the external road network. The covered is assumed at a constant height in any case. The position of the values (in meters) for each entrance, if at a different height, is indicated by means cardinal points (N, S, E, W) and other specific symbology (e.g., C = at centre)
- **notes =** general information collected from the field survey and other sources (e.g., reports of flooding).
- **Pavements** = it indicates the presence of pavements in underpasses for vehicles. The side in which they develop is indicated

Figure 28 Underpasses map focused on a sample area

5.1.5.2 Floods database

Base information is taken from two sources: official fire department bulletins and flood news database managed by the Earth Sciences Department of the University of Florence in its MIG portal (Multirisk Information Gateway).

The UNIFI Department of Earth Sciences collects continuously news about hydrogeological events in Italy scanning online news bulletins every 15 minutes. Using data mining algorithms on the text of the news, the identified news are stored in a geographical database. For every news is reported type, municipality, time and theoretically intensity of the phenomenon. This system is active from 2009. We selected all floods news stored in MIG database occurred in the municipality of Florence.

For floods before 2009, fire fighters bulletins have been consulted. The fire department provides a list of all calling on the territory, reporting date, place and notes about operation. The operation inherent floods have been selected and only if placed on an explicit place (cross road, underpass, a road or square), discarding operations placed on a too vague place, such as a generic zone.

For every flood event collected (from MIG or Fire department) a specific position has been assigned. This was done by searching the internet for photos, videos or other localized evidence of floods. The major sources for this have been the news about the floods. A secondary source, especially for recent events, is the social networks. From news bulletins text, too vague position information, such as a generic indication of a road (especially if this is a long road) or a generic place (avenues of Florence, west zone, the hills, Arno river, and so on) has been discarded, however it has been taken into consideration only if a clear place is indicated (i.e. an underpass, a crossroad, a parking).

Every photo, video or specific indication gives an indication about a flood position. Each flood was georeferenced with points. Flooded roads were georeferenced with a point every 25m. Flooded areas (square, parking) were georeferenced with a point every 25m in a surface distribution. Theoretical assumptions on the extent of flooding, for example from a photo about a part of a street, have been made according to the knowledge of the place. Minimum 3 points have been allocated for each flood, indicating, for example, the two boundary along a road and the ideal centre of the flood.

For every point a field (**date**) with the day when the relative flood happened has been populated. Consulting the collected information, during the last 10 years the rainfall induced flood events in the city of Florence can be assigned to 15 major events dated: June 7, 2007; August 8, 2007; January 13, 2008; June 26, 2009; June 5, 2011; June 8, 2011; November 27, 2012; June 3, 2013; July 10, 2013; July 17, 2013; October 21, 2013; October 24, 2013; 29-30 June, 2014; August 1, 2015; September 16, 2016.

5.1.5.3 Digital Hydrological Model (DHM)

The source data to obtain DHM (see Figure 31) are principally DSM, that provides height data about blocking elements, and DTM, that provides height data in free runoff areas (Morelli et al. 2014). Due the automatic process that generates the models, many imperfections can afflict the model, so additional information about the ground must be integrated into the model (see Figure 29, see Figure 30).

Many natural blocking elements can be obtained directly from the DTM (river banks, ground elevated roads and rails), but buildings information are present only in the DSM. To extract this, the buildings area is needed. For this reason, polygons of the building have been obtained, using the technical map. After merging all the technical maps overlapping the municipal area of Florence, only polygons were extracted (technical maps come from CAD systems, so they contain point, line and polygon primitives). Polygons were only filtered with code from 200 to 216 (classification of buildings in Italian technical maps at scale 1:10000). To correct represent courtyards, these polygons were treated with geoprocessing algorithms, subtracting them from their surrounding polygons. At the end a vector dataset was obtained, were every polygon represents continuous built up area.

Then the basic DHM can be obtained merging DSM and DTM information using buildings as mask. Building polygons were converted to a raster 1m/pixel with value=1 inside polygons and value=0 outside. This mask raster was used as filter to merge DSM (where value=1) and DTM (where value=0).

This basic DHM can be used for base elaboration but has many imperfections in a complex city like Florence. The base DHM has problems especially where there are underpasses that permit water runoff despite height data derived from DSM or DTM.

Road underpasses can be added to the model using a mask obtained from the road network. There are more than one ways to obtain this data: extracting from technical map, using vehicular areas from city connective network data or polygons from road area network data. In this elaboration the last one among these was used. Despite the nominal scale declared by the open data portal, road area network data is suitable for use in scale 1:10000 and were data ready to use for this elaboration. Similarly to what was seen above about buildings, using polygons of the roads to create a mask raster (value=1 inside roads, value=0 outside), more data was added from DTM in the HDM filtering a merge with DTM (where value=1) and HDM (where value=0).

A second analogous adding to HDM has been done using underpasses polygons collected in the underpasses database survey. Using polygons of the underpasses to create a mask raster (value=1 inside underpasses, value=0 outside) more data was added from DTM in the HDM filtering a merge with DTM (where value=1) and HDM (where value=0).

The operations described above remove a lot of buildings (mainly involving bridges above roads or walkways, phenomenon particularly widespread in Florence downtown), but is not sufficient to have a perfect DHM. This is due to some imperfections in the DTM that provides data in free runoff area. In many cases the algorithm that created DTM did not consider small underpasses. In other cases, underpasses are under the ground and DTM cannot represent them. To correct these imperfections of the model further data has been applied to DHM.

Every underpass area was expanded to reach the first visible cells of the DTM that represent the entry point of the underpass. A height equal to the lowest value of the overlapping cells was assigned at the polygons. Then these values ware assigned to the HDM in all cells overlapped by the relative polygon. This operation corrects all of the above cited imperfections and permits us having a model where water can runoff not blocked in little walkways or other not detected “holes”. In this way, the model reports also the “trenches” due to pedestrian underpasses below square, stations and banks.

Florence is a complex city and some artefacts are particular. Thus, a punctual control, especially in downtown, was performed to check the correctness of the model. For example, the whole old bridges area has been edited because of the unique architecture of this artefact (bridges on Arno river with buildings on the sides also with arcades).

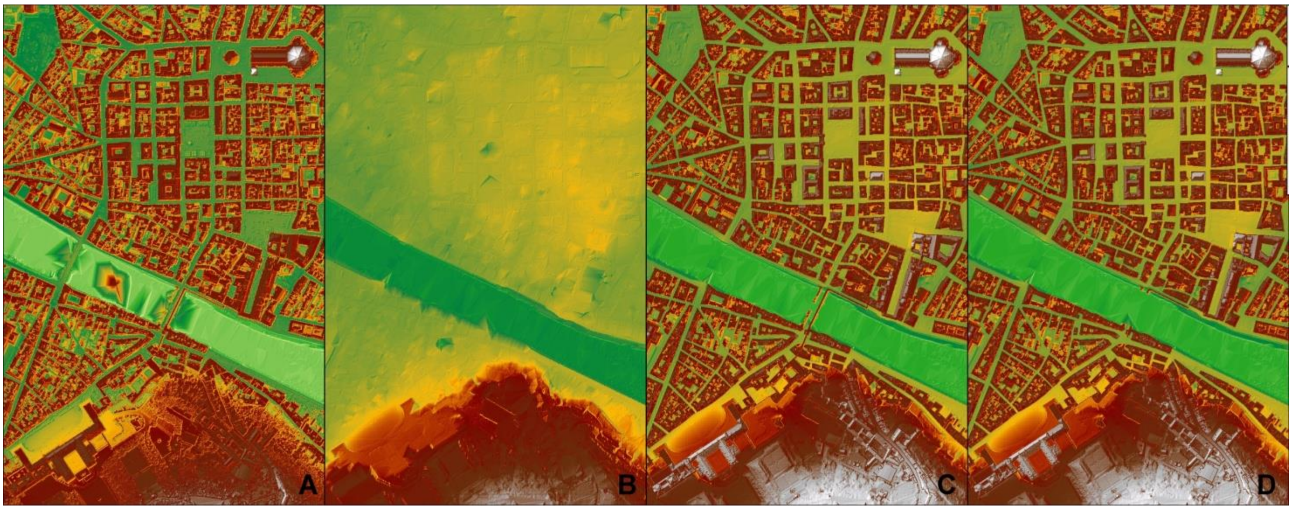


Figure 29 processing steps focused on a sample area: A) DSM, B) DTM, C) Basic DHM, D) final DHM

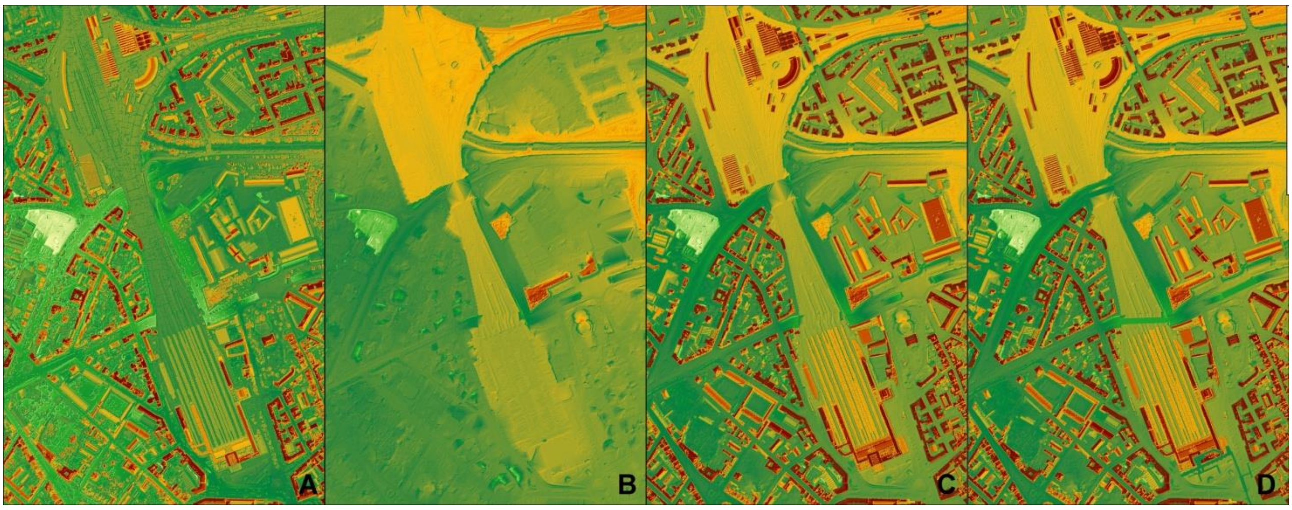


Figure 30 processing steps focused on a sample area: A) DSM, B) DTM, C) Basic DHM, D) final DHM

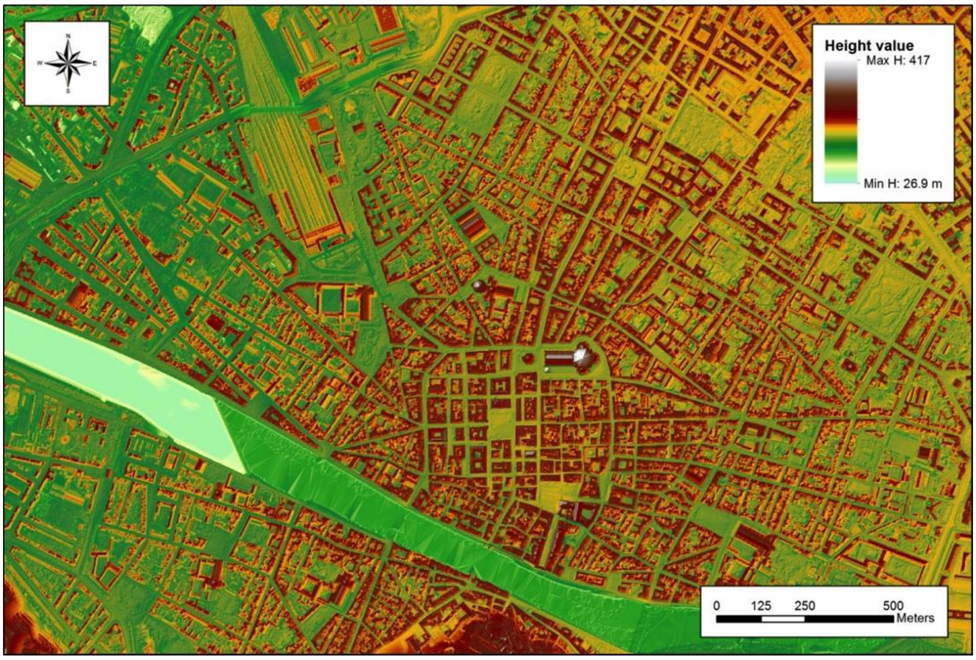


Figure 31 Digital Hydrological Model map focused on a sample area

5.1.5.4 Slope Model

Slope Model (Figure 32) has derived from Digital Hydrological Model. For each DHM cell, the maximum rate of change in value from that cell to its neighbouring ones was calculated. Basically, the maximum change in elevation over the distance between the cell and its eight neighbours identifies the steepest downhill descent from the cell. Conceptually, around the processing or centre cell a plane to the z-values of a 3 x 3 cell neighbourhood was fitted. The slope value of this plane is calculated using the average maximum technique. The direction the plane faces is the aspect for the processing cell. In principle, the lower the slope value is, the flatter the terrain; respectively, the higher the slope value, the steeper the terrain. The value of each cell represents its steepness in degrees.

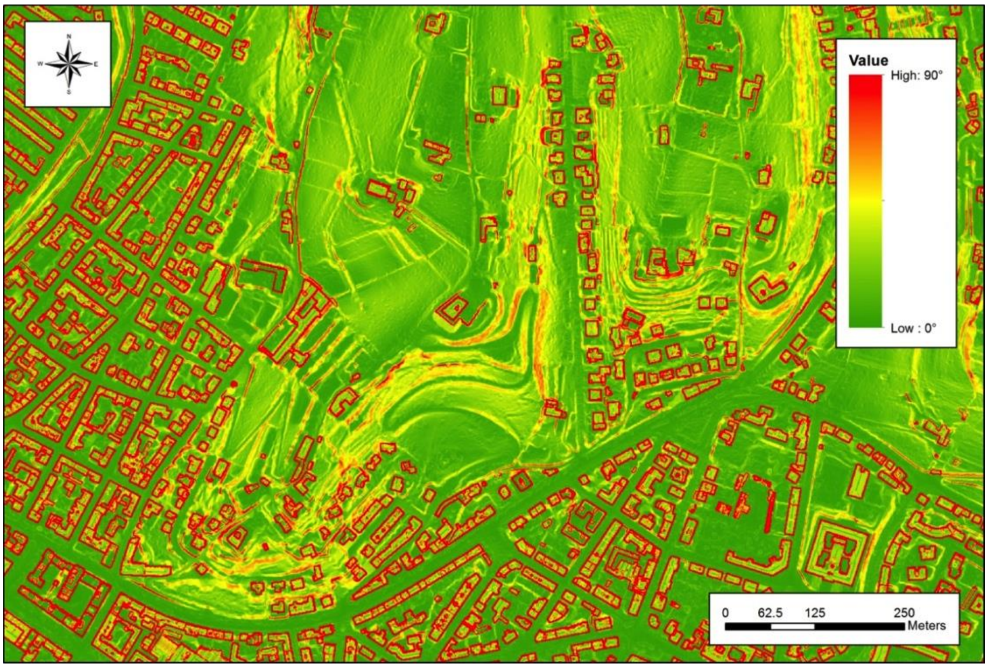
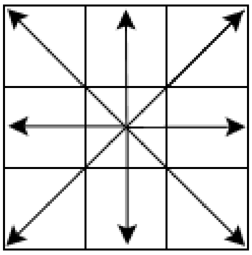


Figure 32 slope model map focused on a sample area

5.1.5.5 Flow Direction Model

Flow Direction Model has derived both from Digital Hydrological Model (DHM) and from its Filled Model (see Depressionless Model below). So two Flow Direction Models were produced like every other model derived from it. The direction of flow is determined by the direction of steepest descent, or maximum drop, from each cell. When a direction of steepest descent is found, the output cell is coded with the value representing that direction. If all neighbours are higher than the processing cell, it will be considered noise, be filled to the lowest value of its neighbours, and have a flow direction toward this cell. If this is not possible (i.e. in DHM Model), a not coded value was assigned. There are eight valid output directions relating to the eight adjacent cells into which flow could travel. This approach is commonly referred to as an eight-direction (D8) flow model and follows an approach presented in Jenson and Domingue (1988) (Figure 33).



32	64	128
16		1
8	4	2

Figure 33 eight-direction flow coding

5.1.5.6 Flow Accumulation Model

Flow Accumulation Model has derived from Flow Direction Model. Every cell value is the accumulated flow as the accumulated weight of all cells flowing into each downslope cell in the output raster. A weight of 1 was applied to each cell, and the value of cells in the output raster is the number of cells that flow into each cell. Cells with a high flow accumulation are areas of concentrated flow and may be used to identify stream channels. Cells with a flow accumulation of 0 are local topographic highs and may be used to identify ridges.

5.1.5.7 Flow Length Model

Flow Length Model has derived from Flow Direction Model. Each cell value represents the sum of the incremental distances from centre of the cell along the flow path from the selected cell to the outlet cell. The flow length assigned to the outlet cell is zero. The Flow Direction Model can be calculated downstream or upstream. For the needs of the project both were obtained.

When it rains, a drop of water landing somewhere in the basin must first travel some distance before reaching the outlet. Assuming constant flow velocities, the cell with the greatest flow length to the outlet represents the hydrologically most remote cell. Its flow length divided by the flow velocity represents a representative lag time for the basin. The lag time quantifies how long before the entire basin is contributing to flow at the outlet and is a representative scale of the basin.

From the Flow Length calculated upstream it is possible to calculate the Time of concentration of the basin. Time of concentration is a concept used in hydrology to measure the response of a watershed to a rain event. It is defined as the time needed for water to flow from the most remote point in a watershed to the watershed outlet (C.T. Haan, B.J. Barfield, J.C. Hayes, 1994). It is useful in predicting flow rates that would result from hypothetical storms, which are based on statistically derived return periods. Several functions to calculate this time were proposed in literature but this calculus is beyond the scope of this project.

5.1.5.8 Depressionless Model

Depressionless Model, also known as Filled Model, was derived from Digital Hydrological Model.

All sinks in the DHM was removed filling them. If a cell is surrounded by cells with higher value it is a sink cell. A sink cell is filled with the minimum value of the surrounding cells. As sinks are filled, others can be created at the boundaries of the filled areas, which are removed in the next iteration. Sink areas (many adjacent cells with same values surrounded by cells with higher value) are filled in the same manner. This process is repeated until the model has no sink cells or areas.

The Depressionless Model normally is the desired input to the flow direction calculation and derived data. The presence of sinks may result in a short flow that ends at the first local sink. It is important to understand the morphology of the area well enough, in order to know what features may truly be sinks on the surface of the earth and which are merely errors in the data.

5.1.5.9 Depression Model

Depression Model (Figure 34) has derived from Digital Hydrological Model and his Filled Model. The Depression Model was obtained by subtracting the cell values of the Filled Model from the original DHM. In this way every cell with 0 values represents surface of DHM where there are no lakes or ponds, zones where water runoff is free. Cells with value less than zero represent zones where water theoretically stagnates before it can run anywhere else.

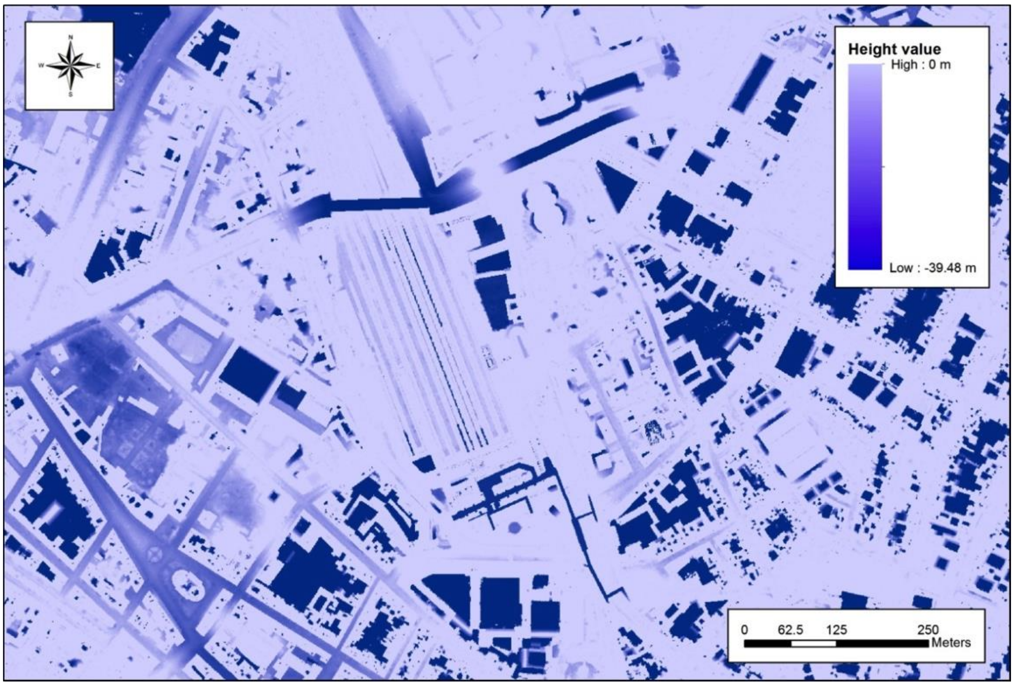


Figure 34 Depression Model map focused on a sample area

5.1.5.10 Depressions Density

Depressions Density has derived from Digital Hydrological Model. Depressions Density represents the density of sinks cells for area units. If a cell is surrounded by cells with higher value it is a sink cell. Contrary to the Depression Model that represents the depressions as a whole, these data try to measure the distribution of the single sink cells on the surface.

After identified each single sink cell of DHM, they were converted into points using the centre of the cell. Then kernel density was used to calculate the density raster 10m/pixel relative to every square meters. Kernel Density calculates the density of point features around each output raster cell. Conceptually, a smoothly curved surface is fitted over each point. The surface value is highest at the location of the point and diminishes with increasing distance from the point, reaching zero at the 100 meter distance from the point. The density at each output raster cell is calculated by adding the values of all the kernel surfaces where they overlay the raster cell centre. The kernel function is based on the quartic kernel function described by Silverman (1986, p. 76, equation 4.5). Every sink point is counted once.

5.1.5.11 Permeability Map

To quickly calculate the permeability map (Figure 35) information obtained from city connective network data and technical map data was used. Once the city was subdivided in polygons marked by attributes “not permeable”, “lightly permeable” or “heavily permeable”, these polygons were converted in a raster 1m/pixel classified respectively 0,1,2.

All polygons from connecting network of category “vehicular area”, “cycle area” or “pedestrian area” were considered not permeable due their coverage in asphalt or in beaten earth (rare). Also buildings were considered not permeable. The same buildings polygons obtained as blocking elements for DHM creation were also used here.

The built up areas in connecting network data represent the whole built up area seen by the roads (the celerimetric survey was done only on street). So buildings polygons were subtracted from these to obtain internal courtyards polygons. Being this courtyard mixed gardens, parking and roads, they were consider as lightly permeable.

All other areas were considered heavily permeable. They almost entirely consist of tree lined areas, parks, garden, wooded areas, cultivated or herbaceous areas.

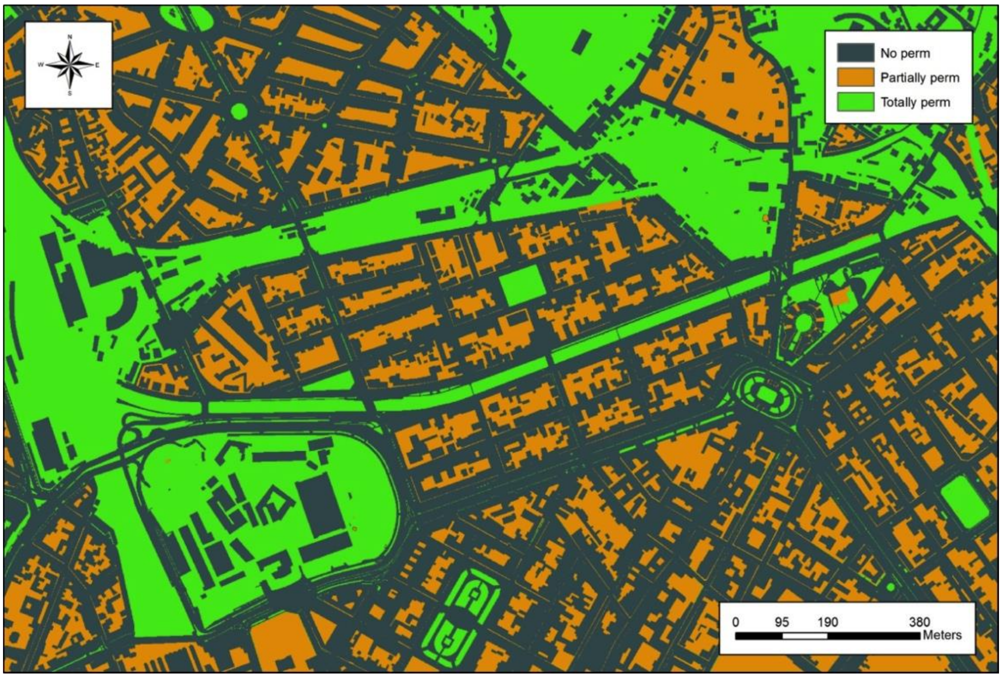


Figure 35 Permeability map focused on a sample area

Some zones required particular attention. For example, Campo di Marte (a sportive area with gardens, big parking, streets, stadiums, swimming pools and auditorium) was edited digitalizing asphalted areas and buildings as not permeable area and the rest as mixed. Furthermore, city of Florence has many private parks in the urbanized areas, also in downtown (Torrigiani garden, Corsi garden, Gherardesca garden, Valfonda garden). These areas would be considered highly permeable in vegetated areas and not permeable in the walkways. For the purposes of our project, because these areas have never been flooded in the past, it was chosen to consider them globally as lightly permeable.

5.1.5.12 Underpasses Map

Underpasses Map (Figure 36) has derived directly from Underpasses Database. The polygons digitalized in the survey were converted in a raster 1m/pixel coded by base level field.

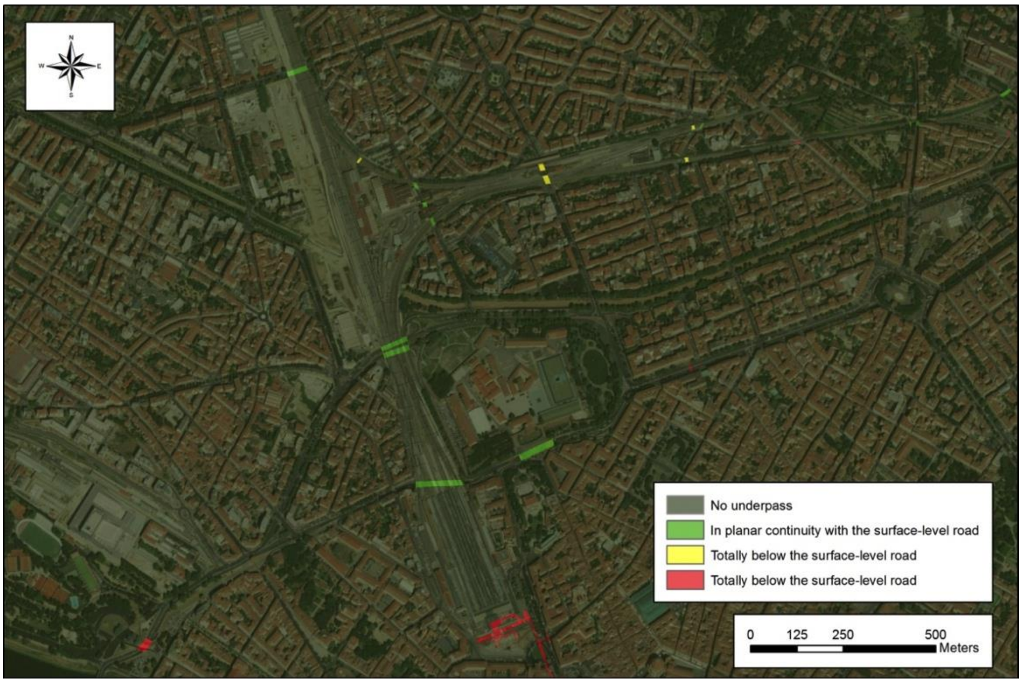


Figure 36 underpass map focused on a sample area

5.1.5.13 Floods Density

To have a more intuitive map of the registered flood a density map has derived (Figure 37) from floods database. Using the point positions stored in the database, kernel density was used to calculate the density raster 10m/pixel relative to every square meter. The kernel density spreads the known quantity of the population for each point out from the point location. The resulting surfaces surrounding each point in kernel density are based on a quadratic formula with the highest value at the centre of the surface (the point location) and tapering to zero at the search radius distance (100 meters). For each output cell, the total number of the accumulated intersections of the individual spread surfaces is calculated.

The resulting raster was classified in 5 classes using the Jenks optimization method (Jenks & George 1967). This method is a data clustering method designed to determine the best arrangement of values into different classes. This is done by seeking to minimize each class's average deviation from the class mean, while maximizing each class's deviation from the means of the other groups. The 5 classes vary from 0 (no flood registered in the area) to 4 (almost heavy rainfall the area is flooded).

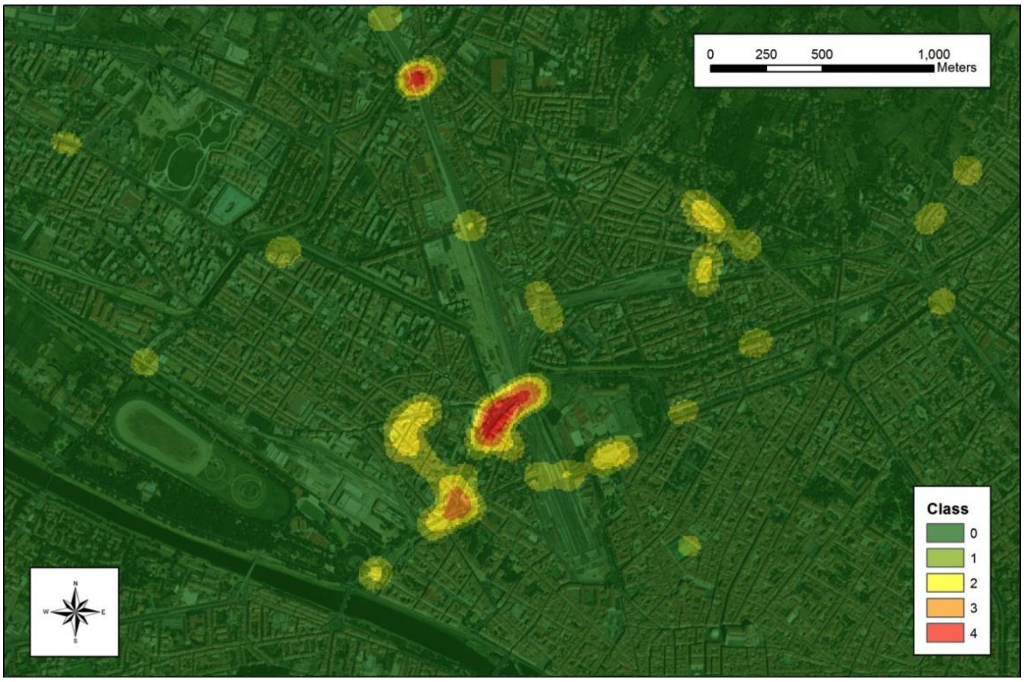


Figure 37 Floods density map focused on a sample area

5.1.6 Flooding modeling discussion and conclusion

Using Regression analysis is possible to try to evaluate which variable, represented in our datasets, most affects the creation of floods during a heavy rainfall in Florence. Regression is a multivariate analysis that establishes relationships between numerous input variables and presents the relationships in a succinct manner, usually as a number or a series of numbers. Basically a regression analysis can be used when sites where the phenomenon exists are already known and it needs to be determined why the phenomenon occurred in these locations (logistics regression).

A regression analysis has two parts:

1. The **dependent variable**, which is the phenomenon whose level or presence is to be predicted or explained for each location in a study site. In our case the dependent variable is the Flood Density.
2. The **independent variables**, which are the known attributes of the locations that influence the level or presence of the dependent variable. In our case the independent variables are the derived datasets.

To produce a table with the dependent variables and the independent ones, first of all, points from Flood Density raster were extracted, one point for every cell centre with value equal to its classification density (0-4) that is the dependent variable. Due the large amount of points, more than 1 million, the number of point with value 0 (no floods, about 99% of the total) were reduced to 0.5% of the starting number randomly, and the number of point with value > 0 to only ones that are on ground (our floods database come from photo or video on ground). At the end a total of 9336 test points resulted (4092 with value>0 and 5244 with value=0). At every point the corresponding cell value of other raster datasets was assigned that represent the independent variables.

Finally, using the attribute table of the test points, a global Ordinary Least Squares (OLS) linear regression to model the dependent variable (coded know floods density) was performed, in terms of its relationships to the set of the explanatory variables.

It was concluded that the flood phenomenon is statistically strongly related with:

- Permeability (inverse proportion): floods occur where terrain is not permeable;

- Presence of underpasses: more under ground level is the underpass and more probable is the flood;

The phenomenon is also related to:

- Height value of DHM (inverse proportion): no knows floods were in hills, but all registered in lowland;

- Slope value of DHM (inverse proportion): floods tend to occur at low gradients;

Furthermore, there is a weak relation with

- Flow length of DHM (upstream and downstream): using both value indicates that floods depends roughly from the dimension of the sub-basins that the complexity of a city creates in its territory. In particular, this indicates that floods are related to the Time of concentration of the sub-basins that, as shown previously, is related to the flow length.

The other independent variables seem not to influence the flood phenomenon. A special observation must be made for the two datasets deriving from sinks and depressions. Probably their irrelevance to the phenomenon is due to the fact that collected flood events are both stagnant and runoff floods.

Finally, it is pointed out that it would be interesting to introduce in this examination also a dataset on sewers and drains, but their distribution, relation and dimension are not really known and the information available are kept confidential by the managing institutions in order to prevent intentional attacks in a sensitive target. Thus, modelling in time and space of the fluid circulation is very difficult because of the high number of variables involved in so little known environment. For all these reasons, the performed analysis considers the worst situation in which the subterranean system is completely saturated (i.e., too much water for the existing capacity) or largely not working for different reasons (accesses blockages, natural or anthropogenic obstructions inside, breaking and so on).

5.2 City weather database

One of the most common factors that may set a city in an emergency is the weather. For instance, as already have been described, intense weather phenomena such as heavily rains can cause flood events of several types (see chapter 3.2). Thus, in such cases, the knowledge of the weather condition may be of vital significance regarding the evacuation phase. In particular, by combining the knowledge of the weather conditions with the knowledge of the areas that is most susceptible to floods we can provide a more safe evacuation plan for the involved people. This dataset is going to be used in both Florence and Athens pilot.

In this direction, a tool for collecting weather data was implemented, in order to be provided them to the Evacuation DSS (responsible for the evacuation plans). The processing of the weather data, by the Evacuation DSS, results in the production of a weather risk factor, which in conjunction with the pluvial susceptibility maps indicates the high risk areas/roads. Thus, the new evacuation plan considers the areas of high risk in order to provide more safe evacuation plans. The tool that we use in order to create this database was the MongoDB, an open-source document database and leading NoSQL database. MongoDB is a cross-platform, document oriented database that provides, high performance, high availability, and easy scalability. MongoDB works on concept of collection and document. The implementation of this database uses one collection for all the files and separate these files based on an assignment token that represents the connection of the file with the respective city. We use JSON files that represents the weather condition in the City of Florence and the City of Athens. This dataset consists of a number of different metrics aggregated in 3 hours period for every day. Some of the measurements that are collected for these two cities are the temperature, cloud cover, humidity, wind speed, visibility and many other that is presented in the Section 3.3.8.

The City Weather database includes the following weather measurements, which are available through a RESTful API.

Table 7. Measurements of Weather

Value	unit	aggregation	Description
eventDate	String	-	The date and time of the measurement observation in UTC
tempC	Integer	3 hours average	The temperature in degrees Celsius
feelsLikeC	Integer	3 hours average	The temperature that humans understand in degrees Celsius
dewPointC	Integer	3 hours average	Dew point temperature in degrees Celsius
heatIndexC	Integer	3 hours average	Heat index temperature in degrees Celsius
windChillC	Integer	3 hours average	Wind chill temperature in degrees Celsius
cloudcover	Integer	3 hours average	The percentage of cloud cover in the city
visibility	Integer	3 hours average	Visibility in kilometres
pressure	Integer	3 hours average	Atmospheric pressure in millibars
humidity	Float	3 hours average	Humidity in percentage
precipMM	Integer	3 hours average	Precipitation in millimetres
windSpeedKmph	Integer	3 hours average	Wind speed in kilometers per hour
windGustKmph	Integer	3 hours average	Wind gust in kilometers per hour
windDirDegr	Integer	3 hours average	Wind direction in degrees

In order to access this data, two RESTful services are available, for the GET and POST data. The assignment token is the identification code of each city until now two assignment token exists one for Florence and one for Athens:

Table 8. City Assignment for Weather Data

City	Token
Athens	2a999219-44ec-43f1-8fdc-fbc54c7bbaf4
Florence	b7d61a51-c645-4340-86c0-f31a51922200

1. **Title:** Get City Measurements
   URL: /assignments/{token}/measurements/
   Method: GET
   URL Params: Required: token=[string] OR Optional: page=[int], pageSize = [int], startDate=[string], endDate = [string]
   Response Codes: Success (200 OK), Unauthorized (401), Forbidden (402), Not Found (404)

2. **Title:** Post City Measurements
   URL: /assignments/{token}/measurements/
   Method: POST
   URL Params: Required: token=[string]
   Data Params: {"eventDate": "","updateState": false,"metadata": "object","measurements": [{"key":0}]}
   Response Codes: Success (200 OK), Unauthorized (401), Forbidden (402), Not Found (404)

The above data is used for weather severity monitoring and associated hazards inference using neural clouds. Using this historical data of the same location of interest, we research the statistical characteristics of the distribution of each phenomenon and design a neural network capable of determining the deviance in a singular or multiple features of the predicted instances. The system architecture is designed in order to accommodate to both unlabeled (historical) or labeled (processed by field experts) data. Moreover, we define specific scenarios that attempt to capture real world hazards that relate directly to weather conditions and their severity. Expertise in the weather monitoring field can additionally finely tune how each scenario is affected by each measurement and their combinations. Additionally, we extrapolate the extremity in each scenario case based on predicted data for the upcoming day, and return a relevant warning.

5.3 Real time Social Media Sensing on Florence

In RESOLUTE has been developed a feature for the Twitter Vigilance tool in order to push the monitoring capability even closer to when the events happens. The Real Time processing of the twitter act as an early detection and warning of possible events in the city and in combination with other timely alarms may drive to an escalation / de-escalation strategy and decisions.

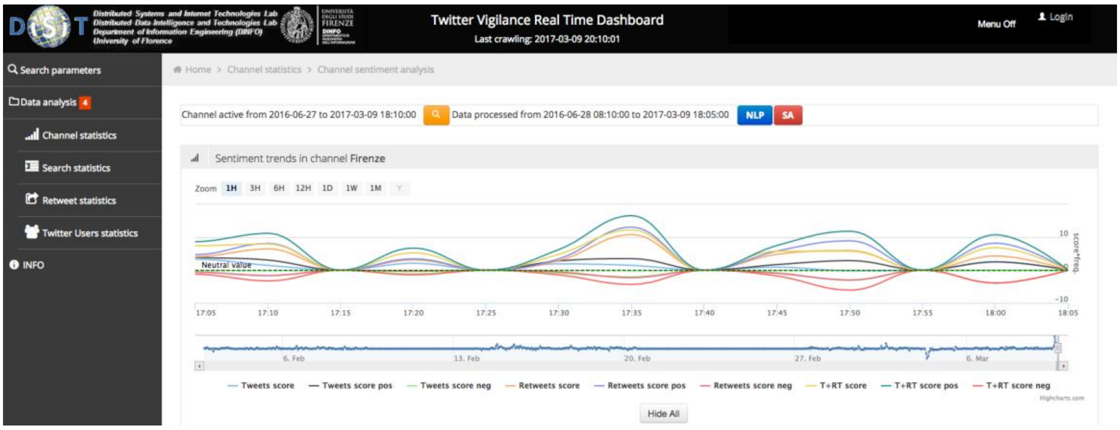
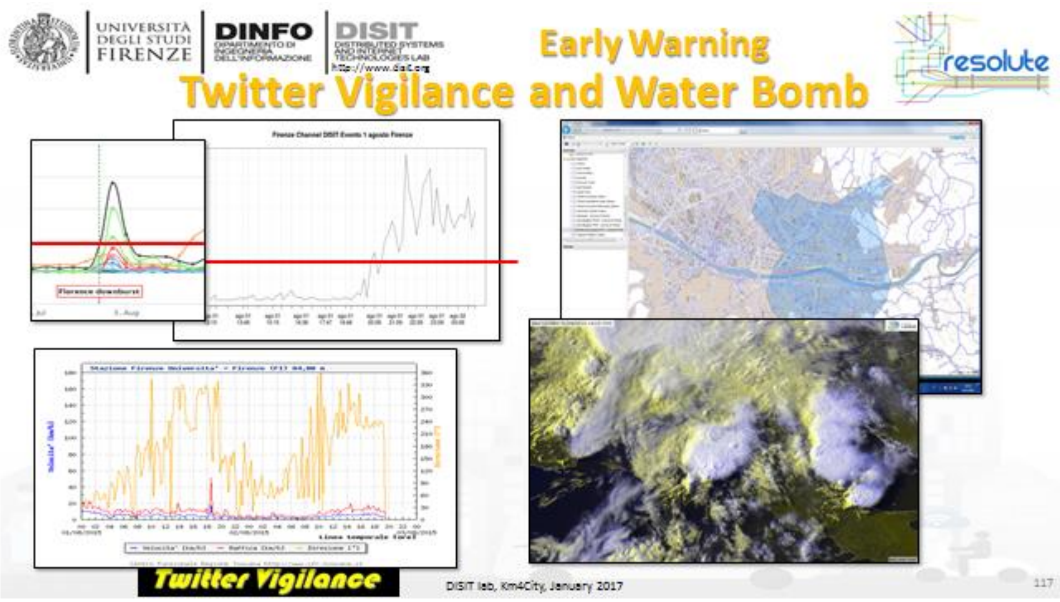


Figure 38 Real Time Twitter Vigilance

In Figure 38, the screenshot of the real time sentiment analysis of the channel "Florence" composed by several tags and keywords as been presented. The delay between data collection and processing is 5 minutes. This sampling rate is the minimum necessary to process a significant number of twitter data that are continuously collected from the API live stream.

The real time analysis of Twitter data can be used as early warning about critical events. Such occurred in Florence for the event of first August 1, which was a water burst in a specific area of Florence as depicted in the following figure. The inception of viral number of tweets was capable of detecting a critical case much earlies than Civil protection and the municipality. Radar and other technologies are not capable to make prediction near early detection.



5.4 Human behaviour tracking via Wi-Fi in Florence

The human behaviour analysis is a critical point of the RESOLUTE. This dataset is made by 2 main sources: a) public Wi-Fi data that is composed by the Access Point registry distributed in the city and the Wi-Fi log data and b) the *Florence where, what?* mobile app released by DISIT.

In particular all the AP need to be georeferenced and identified univocally with an ID within the infrastructure. In the Table 9 has been provided an extract of the AP registry managed by Florence City Council (the mac address provided are simulated) and ingested in the data layer, where are included the following fields:

- Id: unique identifier of the AP in the registry
- Mac address: unique id of the device
- Lat/long: position of the device in the city
- Address: address where the device is situated
- City: City of reference
- Label: a short description of the location to better identify the place (e.g. parking Europa)
- Cluster_lat/cluster_long: this field represents the center of the cell of the grid created to analyse the people flow as shown in

Table 9 AP registry

id	mac	lat	long	address	City	label	Cluster_lat	Cluster_long
1	00-E0-18-98-56-35	43.75583	11.306211	V.le Europa	Firenze	Parcheggio Europa	43.755640199442	11.306244249962
2	00-E0-18-98-56-35	43.760928	11.241871	P.za della Calza	Firenze	Parcheggio Calza / Oltrarno	43.760713175264	11.241753897602
3	00-E0-18-98-56-35	43.780979	11.246678	Fortezza da Basso	Firenze	Parcheggio Fortezza Fiera	43.781000777484	11.246862044323
4	00-E0-18-98-56-35	43.778853	11.237609	P.za Porta al Prato	Firenze	Parcheggio Porta al Prato / Leopolda	43.778695714719	11.23792278756

The AP registry is extended with a the coordinate of the calculated grid used for flow analysis as shown in Figure 39.

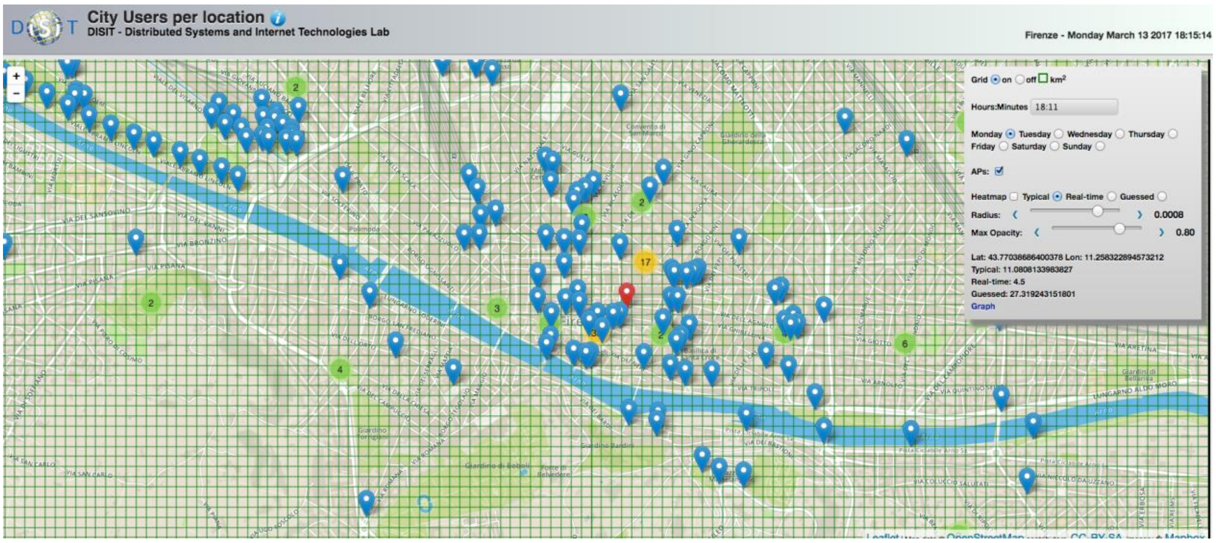


Figure 39 Grid representation for AP-based flow analysis

The city of 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. In Figure 40, a heat map derived from the log data of the WI-FI Aps is represented.

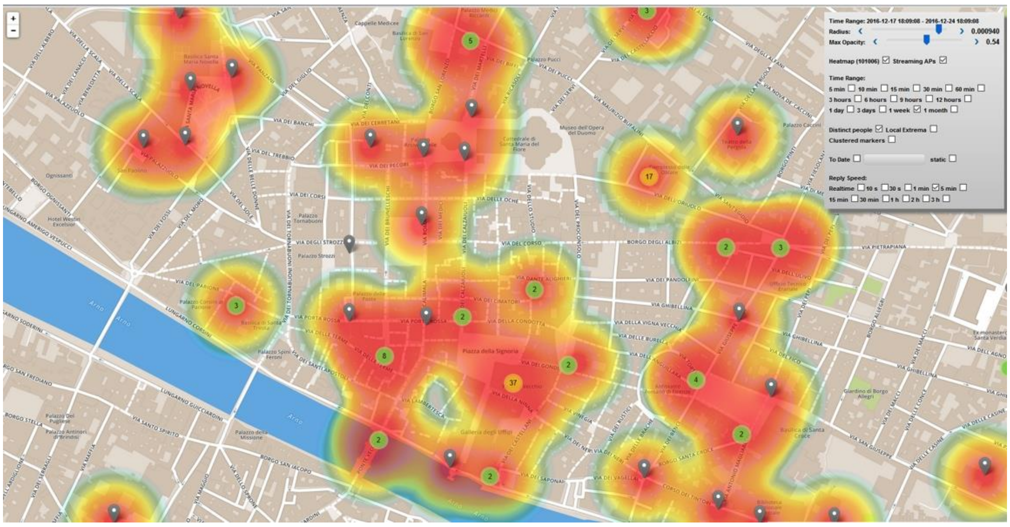
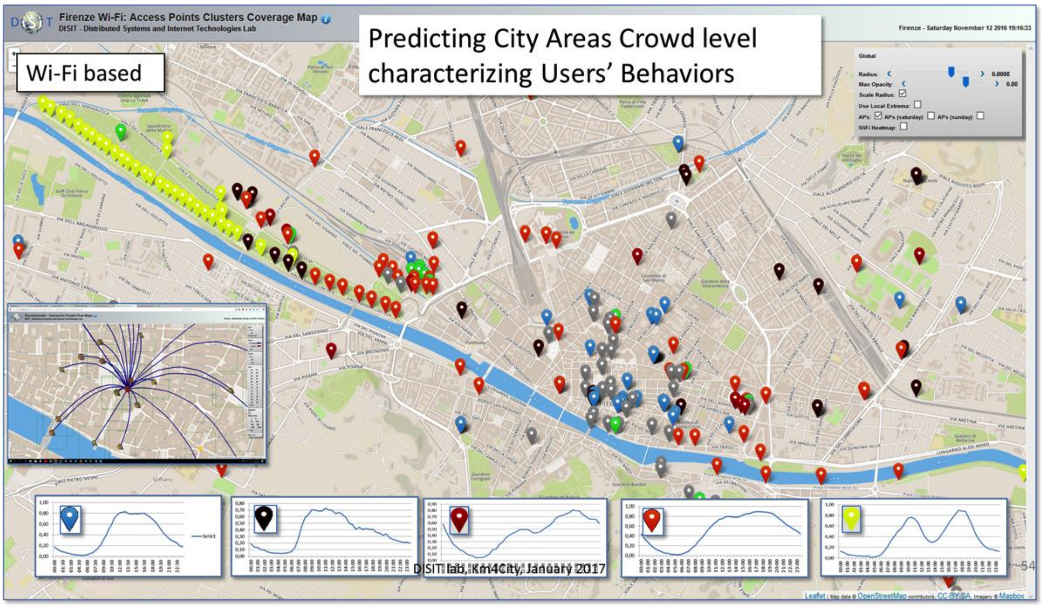


Figure 40 Segment of the heat map reporting the hottest places detected by using selected Firenze WI-FI APs in Florence downtown

Clustering of human behaviour

On the basis of the Wi-Fi data collected over time, it is possible to understand and characterize each are of the city for the activities of city users with respect to the Wifi. Thus a clustering can be performed as reported in the following figure. The clustering allows to perform prediction and early warning detection on each single AP. AP with the similar behaviour presents the same colour.



Origin Destination Matrix

The Origin Destination Matrix is a valuable information to be taken into account during the preparation and planning phases of resilience management. In fact knowing which are the areas of the city interested by mass people movement, may lead dedicated strategies to enable alternative paths to required city destinations.

This dataset has been created 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. In Figure 41, the OD matrix with its the geographical representation is shown.

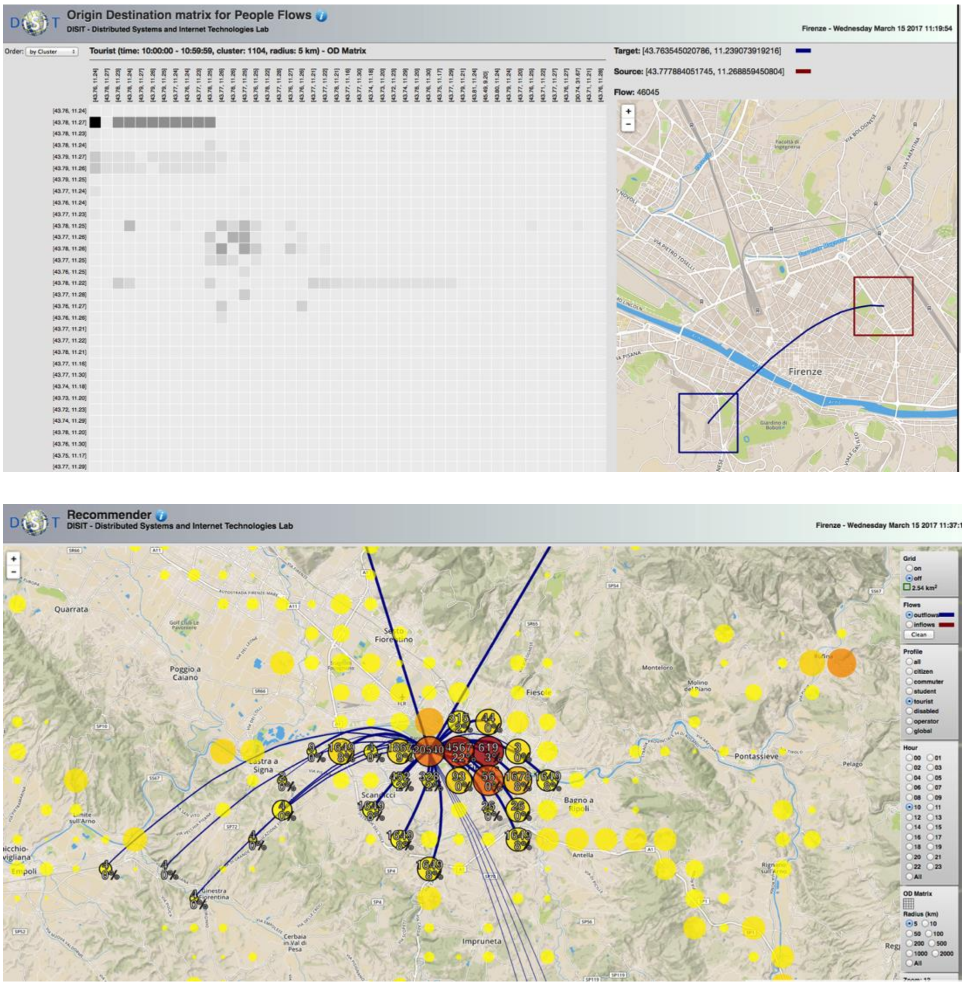


Figure 41 OD Matrix in Florence (tradition and spider versions from mobile Apps, also available from WiFi data)

Trajectories

Another very important dataset useful for resilience management in preparation and planning phase is the trajectories dataset. This dataset is able to cluster the directions of the people according to the OD matrix (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. In the Figure 42 is represented the visual representation if this dataset

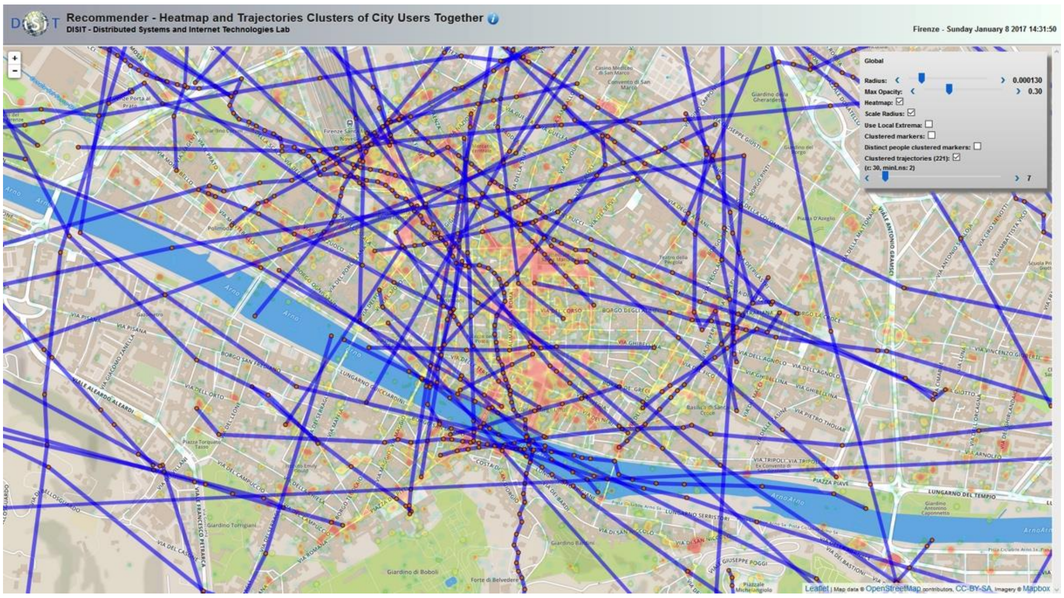


Figure 42 Trajectories representation

5.5 Triage monitoring

The real time data coming from the monitored hospitals is integrated in the ontology and thanks to such a integration, the status can be associated to georeferenced entity (Figure 43).

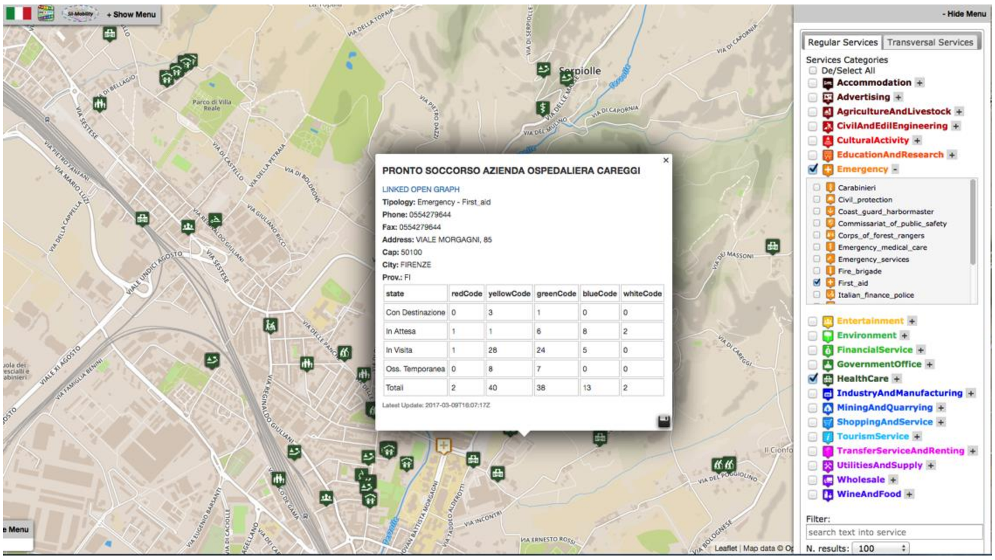


Figure 43 Real time triage associated to the georeferenced hospital

The triage dataset is created collecting data generated by several hospitals in the city of Florence. This is a reliable and real time dataset. It is reliable since the source is generated by the information systems installed in the emergency rooms in the hospitals. The data is updated every time a patient arrived is evaluated and a “color” is assigned.

The hospitals monitored in Florence are:

- Mugello, Hospital
- San Giovanni di Dio Hospital
- Careggi Hospital
- Santa Maria Annunziata Hospital

- Santa Maria Nuova Hospital

Others are gong to be connected during the project.

5.6 Sensing Social Network on Athens

5.6.1 Social Networks in Emergencies

The wide usage of social networks from distributed users around the world and the daily increased number of new communications as well as the constant production of new messages, make social networks the appropriate field to build powerful real-time event detection methods, as well as to track humans' behaviour using social media.

In recent years, several efforts have been made in order to sense social networks during emergencies. These efforts can be mainly categorised into the following lines: a) real-time detection of events, b) human behaviours/sentiment assessment and c) crisis management and. Most of the relative studies try to leverage Twitter among of the rest social networks due to Twitter's way of work (e.g. character-limited post, hashtags, followers-followings relationship).

5.6.2 Twitter Dataset creation

Similarly for the exploitation of social networks' dynamics, three Twitter-based crawlers have been implemented for collecting different datasets of Twitter. These datasets are going to be used by the Evacuation DSS in Florence and Athens pilots for giving the evacuation responsible the opportunity to have a view of the most discussed topics in Twitter, at real time, or for giving him/her the opportunity to monitor tweets containing certain keywords within a specified area. The evacuation responsible will be able to analyse further the dataset by an interactive way, in order to be able to detect early abnormal/emergent events.

In this direction three Twitter-based crawlers have been implemented:

- Twitter crawler for collecting trending topics for specified cities.
- Twitter crawler for collecting tweets containing certain keywords around specified areas.
- Twitter crawler for collecting the Followers and the Followings for a number of users.

Using the aforementioned crawlers, the following datasets of Twitter have been collected. Below we describe two of the collected datasets, and two testing examples of them, fulfilled the following requirements

- Main Datasets
    1. Main Dataset 1
        - the location it was posted referred to one of the following cities {Amsterdam, Ankara, Athens, Baghdad, Beijing, Berlin, Bogota, Brasilia, Brussels, Budapest, Buenos Aires, Damascus, Dublin, Edinburgh, Firenze, Hong Kong, Kyiv, London, Madrid, Mexico City, Moscow, New York, Ottawa, Paris, Rome, San Francisco, Singapore, Stockholm, Sydney, Tehran, Tokyo, Toronto, Vienna, Warsaw, Bucharest, Panama City, Zagreb}. The list with the cities can be updated (e.g. add/remove a city) according to user's preferences.
        - at least one of the {Firenze, Athens} cities was mentioned within its text
        - All available metadata are retrieved (e.g. original/retweeted post, author, tweet's text, date and time of tweet's creation, etc.)
    2. Main Dataset 2
        - the location it was posted referred to one of the following cities {Athens, Firenze}

- keywords relative to emergency conditions was mentioned within its text
- All available metadata are retrieved (e.g. original/retweeted post, author, tweet's text, date and time of tweet's creation, etc.)

- Testing Datasets
  1. Testing Dataset 1
     - it was posted between the 10th and the 16th of November 2015
     - the location it was posted referred to one of the following cities {Amsterdam, Ankara, Athens, Baghdad, Beijing, Berlin, Bogota, Brasilia, Brussels, Budapest, Buenos Aires, Damascus, Dublin, Edinburgh, Hong Kong, Kyiv, London, Madrid, Mexico City, Moscow, New York, Ottawa, Paris, Rome, San Francisco, Singapore, Stockholm, Sydney, Tehran, Tokyo, Toronto, Vienna, Warsaw, Bucharest, Panama City, Zagreb}
     - at least one of the above cities was mentioned within its text
     - All available metadata were retrieved.
     - It should be highlighted that the period the collected data referred to was not randomly selected. On the contrary, it was the week that included the day of the unfortunate terroristic attacks against Paris on the 13th of November 2015.
  2. Testing Dataset 2
     - it was posted between the 28th of January 2017 and the 6th of February 2016
     - the location it was posted referred to one of the following cities {Athens, Brussels, London, New York, Paris}
     - the posted Tweets were referred to one of the following keyword {Brexit, grexit, terrorist}

After retrieving the data a post processing user-interactive procedure can be applied in order to extract meaningful and useful information. The post processing procedure is based on a visual analytics approach which uses as a core method the k-partite graphs and a force-directed algorithm to create a visual graph. Graphs are widely used structures for representing networks consisting of "nodes" and nodes' interconnections which called "edges ". Regarding the graph representation of the collected datasets, all the basic entities (e.g. users-authors, tweets, keywords, location of the retrieved tweets, etc.) can be represented as nodes, while their interconnections can be represented as edges. The positioning of the objects over the graph is based on the relevance of their attributes, such as nearby nodes present highly similar behaviour, while distant points should have nothing in common. Considering the **Errore. L'origine riferimento non è stata trovata.**, the following example is presented that displays the formatted graph using the k-partite and a force directed method. (see Figure 44), displays how much, the users of Paris, Athens, Brussels, New York and London talked about the following subjects: brexit, grexit, terrorist between 1st and 3rd of February 2017. It can be seen that within this period the users of Paris talked more for the brexit regarding the users of the rest cities.

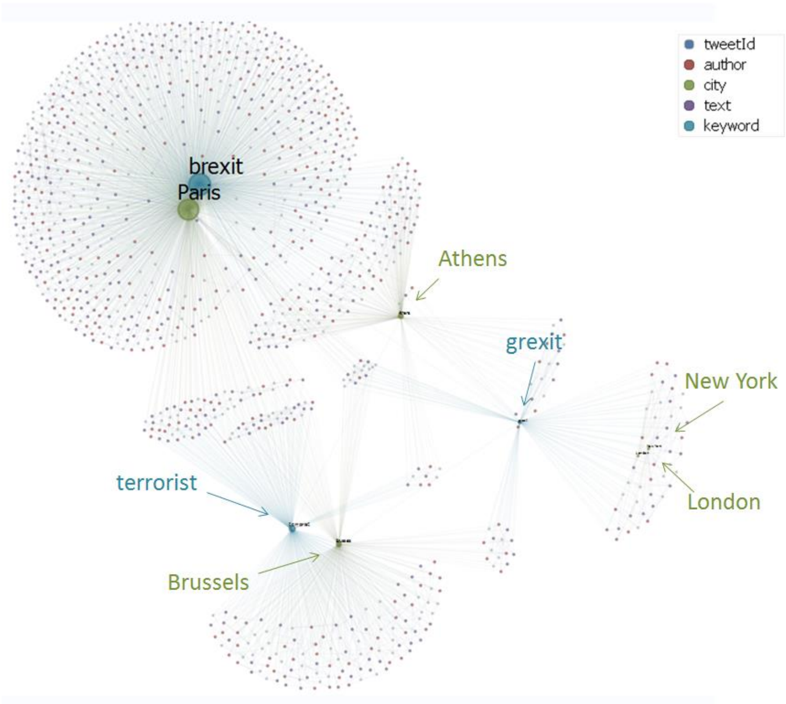


Figure 44 k-partite Graph visualization of all the tweets of Errore. L'origine riferimento non è stata trovata. between 1st and 3rd of February 2017. Users of Paris seem to have talked much more for "brexit" than the users of the rest cities.

After the k-partite graph positioning further analysis can be applied over the datasets. For instance, root cause analysis of suspicious/malicious/abnormal phenomena can be performed and/or even anomaly detection. For example, in order to examine the way of detecting an anomaly/emergent using the aforementioned approach we will consider the **Errore. L'origine riferimento non è stata trovata.**. As it has been already described the Tweets of the **Errore. L'origine riferimento non è stata trovata.** were retrieved between 10th and the 16th of November 2015. Among these dates and particularly on the 13th of November a terrorist attack was took place in Paris. By monitoring the histogram describing the amount of the retweeted Tweets during the week 1-16 of November 2015 (Figure 45), it can be easily observed that at the 13th of November an outburst of the amount of the retweeted Tweets occurred. Besides that this is already an indicator of an anomaly event, this urge us to examine the event in more detail.

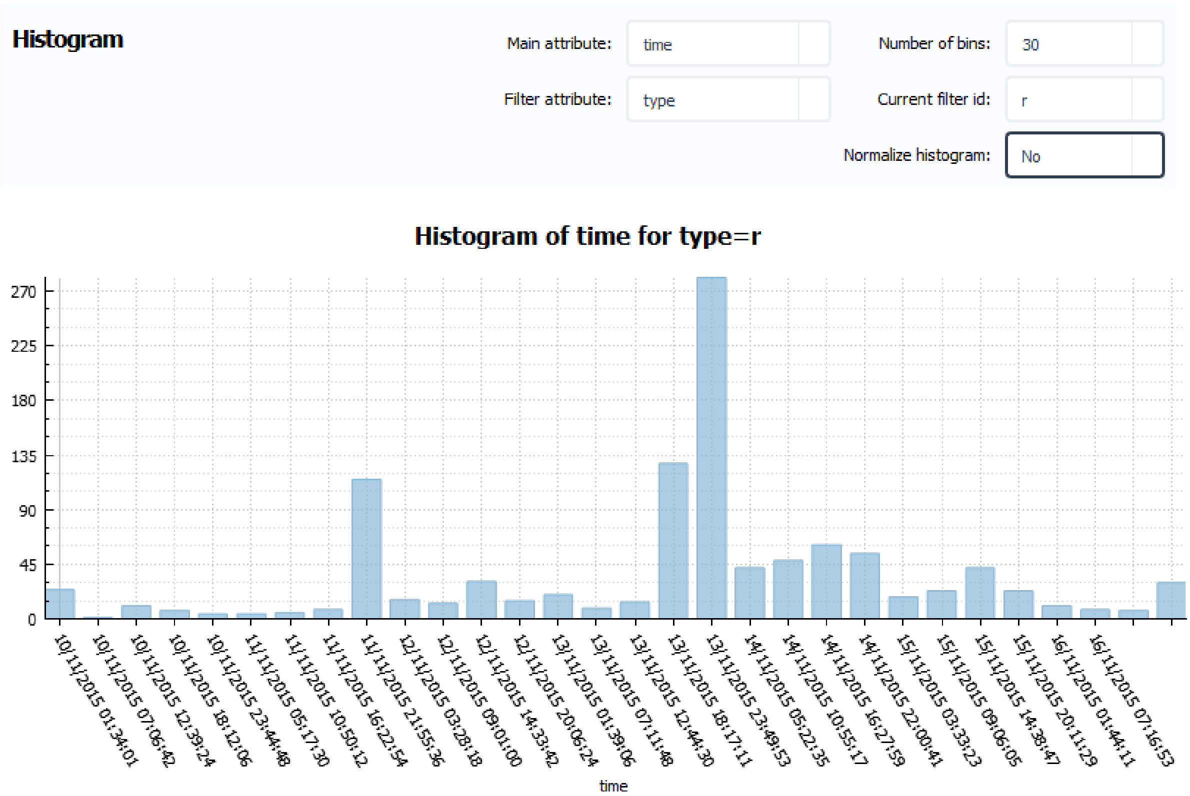


Figure 45 The histogram of posted re-tweets, of Dataset T1, over time during the week 10–16 of November 2015.

In order to analyze the event further we construct the graph representation of the Dataset. As (Figure 46), displays on the right side of the image there are two uni-partite graphs. The first graph represents the retrieved tweets containing the name of the cities within their text, while the second graph represents the city where retrieved tweets have been sent from. The graph at the center of the image combines the two uni-partite graphs to one large k-partite. By positioning the entities using the force directed algorithm visual clusters are formulated. The cities between which a large amount of information is exchanged are positioned in close distance. As we can see there are a lot of tweets from New York about Paris and also quite a few from Ottawa, Toronto & Singapore talking about New York. By selecting the corresponding nodes (i.e. tweets) around the aforementioned pairs and by producing the word clouds from them, we can notice that the most frequent words found in the tweet and re-tweeted messages are the words ‘attacks’, referring to terrorist attacks happened that day in Paris, and ‘job/hiring’, referring to some new vacancies and job opportunities in the area of New York.

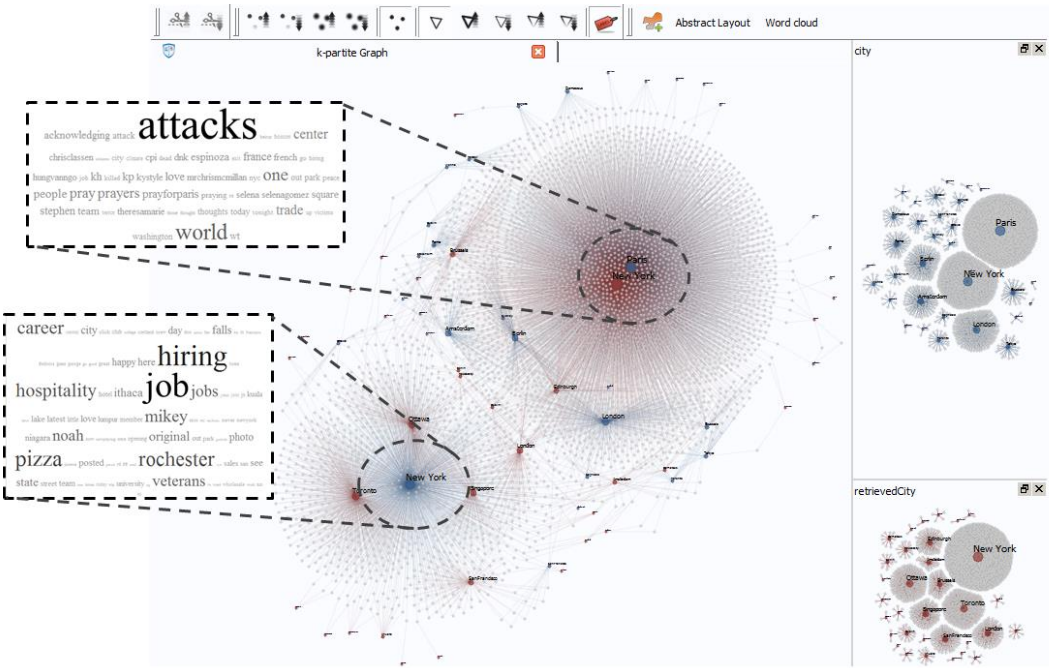


Figure 46: k-partite Graph visualization of all the tweets (i.e. nodes) that were posted between the 10th and the 16th of November 2015. The utilized features were the name of the city the tweet retrieved from and the name of the city contained within its

The final output of the Twitter Sensing Module will be the results of the word clouds -words and times of appearance within dataset for each word- annotated with the metadata of the selected actions that contributed to their production (e.g. dates, involved nodes, etc.) The outputs will be produced using the retrieved data from the real-time crawlers in conjunction with the procedure of data analysis, and will be stored locally to the Evacuation DSS for three months.

6 CONCLUSIONS

This deliverable provides a description of the method used to select and collect useful datasets for Florence and Athens pilots, defining a multi-step approach that requires the definition of a strategy, resilience needs of decision makers, dataset identification for addressing resilience needs, assessment of data availability and the definition of techniques to collect big data.

The strategy adopted is inspired by Bungay's directed opportunism (Bungay, 2010), a decision strategy where the flexibility in the control loop is considered a main driver to bridge knowledge, alignment and effect gaps. RESOLUTE adopts this perspective to understand the resilience needs of the operator and to identify the early relevant datasets able to address such needs (see Annex A and B). In particular, the datasets are mapped to the several classes of decisions identified from the T5.1 analysis. The method to collect and ingest heterogeneous datasets in the Data Layer of the platform is also presented.

Moreover a number of new specific datasets needed to extend the RESOLUTE pilots for ERMG validation, have been created: Flood susceptibility areas, (for Florence), twitter based dataset (for Florence and Athens), Weather Database (for Florence and Athens). These new datasets with the new data flows identified and connected to the data layer, integrate the current data availability present in the KM4CITY.

Finally the datasets identified and reported in Annex A will be further refined and new datasets might be added according to the Pilots defined in the Task 6.1.

To conclude it is worth to remark that working with data brings several issues that needs to be solved case by case. Relevant issues encountered in the activities and their related solutions are summarised in the Table 10 as a lessons learnt.

Table 10 Lesson Learnt

Issues encountered	Solution adopted
Data availability	• Working with organizations to verify the (explicit or implicit) existence of the information within the organization • Working with organizations to implement a methods on their side to allow the data layer crawler to access to their data
Security- related data (Dataset that cannot be moved outside of the owner premises for national or company security reasons as the energy network topology)	• Aggregated or distilled information are collected. • Definition of a offline inter-institutional protocol to be activated in case of emergency where only relevant data will be sent on demand, • Execution of the RESOLUTE applications on data aggregator premises
Personal data (e.g. people movement tracking)	• Disclaimer requested to the user (e.g. app users) • Anonymization techniques to ingested data are applied • Special cybersecurity strategy to protect the collected data in the data storage are applied • Execution of the RESOLUTE applications on data aggregator premises
Legal constrains in data re-use because of their relation to business (e.g. data generated by private transport operators)	• Dedicated agreements has been created between data producer and data aggregator to rule the data distribution, reuse and elaboration • Execution of the RESOLUTE applications on data aggregator premises
Data not available in a standardised file formats	• Ad hoc ETL processes has been produced to translate data into the adopted data model
Data value/content not provided in a well structured format	• Data/content inspection to understand the meaning of the data fields
Data provider APIs reliability	• A periodical check on the APIs operation is include din the data collection layer
Common Geographical reference of the data generated by different DSS (e.g. Exist several version of the Florence streets network adopted by public institutions (region, Florence mobility dept. , Citta' metro of Florence)	• Agreement at pilot (city) level (among the actors on which is most authoritative and updated version of the cartography to be adopted as a baseline.

7 REFERENCES

- [km4city] KM4City Sentient city Ontology. http://www.disit.org/km4city/schema http://www.disit.org/km4city/schema
- Bellini Pierfrancesco, Monica Benigni, Riccardo Billero, Paolo Nesi, and Nadia Rauch. 2014. Km4city ontology building vs data harvesting and cleaning for smart-city services. Journal of Visual Languages & Computing 25, 6 (2014), 827–839.
- Bellini, P., Cenni, D., Nesi, P., (2016) "AP Positioning for Estimating People Flow as Origin Destination Matrix for Smart Cities", The 22nd International Conference on Distributed Multimedia Systems, DMS 2016, Italy, 2016.
- Black, R. (2009). Market Gardening in Florence: Transgressing Urban and Rural Borders in the Struggle for Local Food. *Estudios del hombre*, (24), 203-212.
- Bungay Stephen – The Art of Action: How leaders close the gaps between plans, actiosn and resuls, 2010
- Campolo, M., Soldati, A., & Andreussi, P. (2003). Artificial neural network approach to flood forecasting in the river Arno. Hydrological Sciences Journal, 48(3), 381e398.
- Crisci, A.; Grasso, V; Morabito, M.; Nesi, P.; Pantaleo, G. (In corso di stampa). Public crowd-sensing of heat-waves by social media data. In: 6th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), Trieste, Italia, 12-16 settembre 2016
- Gaitanidou, E., Bellini, E., Ferreira, P., (2016a), European Resilience Management Guidelines, RESOLUTE Deliverable 3.5, May 2016. http://www.resolute-eu.org/index.php/deliverables
- Gaitanidou, E., Tsami, M., (2016b), ERMG Adaptation to UTS, RESOLUTE Deliverable 3.7, May 2016 http://www.resolute-eu.org/index.php/deliverables
- Grasso V., A. Crisci, P. Nesi, G. Pantaleo, I. Zaza and B. Gozzini, (2016) "Public crowd-sensing of heat-waves by social media data", 16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), 12–16 September 2016 | Trieste, Italy, CE2/AM3, Delivery and communication of impact based forecasts and risk based warnings.
- Grasso, Valentina; Crisci, Alfonso; Nesi, Paolo; Pantaleo, Gianni; Zaza, Imad; Gozzini, Bernardo (In corso di stampa). “Italian codified hashtags for weather warning on Twitter. Who is really using them?. In: 6th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), Trieste, 12-16 settembre 2016
- Haan, C. T., Barfield, B. J., & Hayes, J. C. (1994). Design hydrology and sedimentology for small catchments. Elsevier.
- Howard RA (1988) Decision analysis: practice and promise. *Management Science* 34(6): 679-695.
- Igor Linkov, Zachary A. Collier, Decision Making for Resilience within the Context of Network Centric Operations 19th ICCRTS
- Jacobs, (2009). Improved Understanding of Pluvial Flood Risk in Scotland, Report to SEPA.
- Jenks, George F. (1967). The Data Model Concept in Statistical Mapping, International Yearbook of Cartography 7: 186–190.
- Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric engineering and remote sensing, 54(11), 1593-1600.
- Mazzeo, G. (2011). Trasformazioni urbane nelle capitali post-unitarie: Torino, Firenze, Roma. *Tema. Journal of Land Use, Mobility and Environment*, *4*(1).
- Morelli, S., Battistini, A., & Catani, F. (2014). Rapid assessment of flood susceptibility in urbanized rivers using digital terrain data: Application to the Arno river case study (Firenze, northern Italy). Applied Geography, 54, 35-53.

- National Weather Service, (2010). Definitions of flood and flash flood. Available at http://www.srh.noaa.gov/mrx/hydro/floodde
- Poggi G. (1882). Sui lavori per l'ingrandimento di Firenze. Relazione (1864-1877), 374 pp., Barbera Ed., Firenze, Italy.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

8 ACRONYM TABLE

API	Application Programming Interface
C&C	Command & Control
CI	Critical Infrastructure
CRAMSS	Collaborative Risk Assessment and Management Support System
DBMS	Data Base Management System
DIM	Data Ingestion Manager
DoW	Description of Work
DR	Disaster Recovery
DSS	Decision Support System
ECA	Event Condition Actions
EFMEA	Extended Failure Modes and Effects Analysis
ERMG	European Resilience Management Guidelines
ESB	Enterprise Service Bus
ESMA	Emergency Support Mobile App
ETL	Extract Transform Load
FRAM	Functional Resonance Analysis Method
GPS	Global Positioning System
GSLB	Global Server Load Balancing
GTM	Global Traffic Management
GUI	Graphic User Interface
HMI	Human Machine Interface
HTTP	Hyper Text Transfer Protocol
IPR	Intellectual Property Rights
JSON	Java Script Object Notation
LAN	Local Area Network
LTZ	Limited Traffic Zones
MDR	Maximum Defect Rate
MTBF	Mean Time Between Failures
MTR	Minimum Time to Recovery
NLP	Neuro Linguistic Programming
OD	Origin Destination
PA	Public Authority
PTA	Public Transport Authority
RDF	Resource Description Framework
RPN	Risk Priority Number
SA	Sentiment Analysis
SotA	State of the Art
SPARQL	SPARQL Protocol and Query Language
SQL	Structured Query Language
TRE	Total Risk Estimate
UPT	Urban Public Transport
UTC	Urban Traffic Control

UTM	Urban Traffic Management
UTS	Urban Transportation System
VMS	Variable Message Sign
WAN	Wide Area Network
WKT	Well Known Text
WP	Work Package
XML	eXtended Mark-up Language

ANNEX A - EARLY DATASET SELECTION FOR FLORENCE PILOT

	Name	Category	Dynamic	Frequency update period	Format	Data Provider	Data URL	Licence/Condition of use	Dataset description
F1	Tuscany streets and related civic numbers	Infrastructure	S	ND	SHP	Open Data-Regione Toscana (Tuscany Region)	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0 IT).	Road Arches classified by technical type - Administration of the Metropolitan City of Florence.
F2	Tuscany street directions, type	Infrastructure	S	ND	SHP	Open Data-Regione Toscana (Tuscany Region)	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0 IT).	Road Arches classified by technical type - Administration of the Metropolitan City of Florence.
F3	Florence squares	Infrastructure	S	ND	SHP	Open Data-Regione Toscana (Tuscany Region)	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0 IT).	Road Arches classified by technical type - Administration of the Metropolitan City of Florence.
F4	Pedestrian areas - the city of Florence	Mobility	P	2/Year	SHP	Open Data-Comune di Firenze (City of Florence)	http://opendata.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F5	Districts of Florence	Infrastructure	S	ND	SHP	Open Data-Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	

F6	Public Gardens	Service	S	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F7	Traffic lights	Mobility	S	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F8	UTC traffic lights	Mobility	S	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F9	Tram traffic lights	Mobility	S	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F10	Limited traffic area,	Mobility	S	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	computerized gates in the centre of Florence
F11	Limited traffic area, emergency plans, routes and solutions	Mobility	D	ND	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	ND	Real time speedlights status
F12	Cycle paths	Mobility	S	NA	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	network
F13	Controlled Parking Zones	Mobility	S	NA	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	position
F14	Covered parking description	Mobility	S	1/Month	SHP	Open Data- Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	position

F15	Covered parking free place	Mobility	D	1/min	SHP	Open Data-Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	free spaces in real time
F16	park and ride	Mobility	S	1/Month	SHP	Open Data-Comune di Firenze (City of Florence)	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	position
F17	traffic flow sensors position	Sensor / Mobility	S	1/min	XML	MIIC/Citta' Metropolitana Firenze	ND	protected	Position
F18	traffic flow sensors	Sensor / Mobility	D	1/min	XML	MIIC/Citta' Metropolitana Firenze	ND	protected	Flows (average veolocity)
F19	people flow sensors	Sensor / Mobility	D	1/min	JSON/ DB	DISIT lab	NA	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	Real time position
F20	Ordinances	Mobility	P	1/Month	SHP	Open Data-Comune di Firenze (City of Florence)	http://datigis.comune.fi.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	city yard areas
F21	Cemetery	Service	S	1/Month	SHP	Open Data-Comune di Firenze (City of Florence)	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	position
F22	Airports	Mobility	S	1/Month	SHP	Open Data-Regione Toscana (Tuscany Region)	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0 IT).	Road Arches classified by technical type - Administration of the Metropolitan City of Florence.
F23	Ports	Mobility	S	1/Month	SHP	Open Data-Regione Toscana (Tuscany	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0	Road Arches classified by technical type - Administration of the Metropolitan City of Florence.

						Region)		IT).	
F24	Public light poles	Infrastructure/ Mobility/Servic e	S	2/Year	SHP	Open Data-Comune di Firenze (City of Florence)	http://dati.tosca na.it	CC BY 3.0 IT https://creativecom mons.org/licenses/b y/3.0/it/	position
F25	Local public transport	Mobility	P/D	1/Month	google standa rd (gtfs)	Open Data-Comune di Firenze (City of Florence)	http://dati.tosca na.it	CC BY 3.0 IT https://creativecom mons.org/licenses/b y/3.0/it/	timetables, routes, bus stop realted to 16 different TUSCANY TPL operators: 13 bus 1 Ferry 1Tram 1Train
F26	Local Public Transport	Mobility	D	1/min	gtfs	MIIC - Regione Toscana (Tuscany Region)	http://www501.r egione.toscana.i t/osservatoriotra sporti/	ND	1 bus operator: 5 lines (ATAF Firenze)
F27	Local Public Transport	Mobility	D	1/min	?	THALES/GES T	http://www501.r egione.toscana.i t/osservatoriotra sporti/	ND	TRAM operator (GEST)
F28	Local Public Transport	Mobility	D	1/min	gtfs	MIIC - Regione Toscana (Tuscany Region)	http://www501.r egione.toscana.i t/osservatoriotra sporti/	ND	all the operators
F29	Taxi Areas (taxi stops)	Mobility	S	NA	SHP	Open Data-Comune di Firenze (City of Florence)	http://dati.tosca na.it	CC BY 3.0 IT https://creativecom mons.org/licenses/b y/3.0/it/	position

F30	Electric vehicles' charging columns	Mobility/Service	S	NA	SHP	Open Data-Comune di Firenze (City of Florence)	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	position
F31	Electric vehicles' charging columns	Mobility/Service	D	5/hour	ND	Open Data-Comune di Firenze	ND	ND	
F32	variable message signs position	Signalling	S	1/Month	SHP	City of Florence	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F33	variable message signs status	Signalling	D	1/min	ND	City of Florence	http://dati.toscana.it	ND	
F34	wifi	Structure	P	1/hour	SHP/ DB/JSON	City of Florence / DISIT	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	Wifi access point positions
F35	beacon	Structure	P	1/hour	JSON	DISIT		CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	Beacon installed in the City
F36	Information related to Civil Protection (via Social Channels: Twitter)	Social	D	1/min	API	DISIT - Twitter Vigilance		CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	
F37	Events in the city of Florence	Event	P	2/Day	XML/JSON	Open Data-Comune di Firenze (City of Florence)	http://dati.toscana.it	CC BY 3.0 IT https://creativecommons.org/licenses/by/3.0/it/	volume, capacity, schedules, type of people
F38	Emergency hospitals, emergency rooms, location, capacity	Healthcare	S	2/Year	API	Open Data-Regione Toscana (Tuscany Region)	http://dati.toscana.it/	CC BY 3.0 IT Creative Commons - Attribuzione 3.0 Italia (CC BY 3.0 IT).	

F39	Triage accesses at the emergency room	Healthcare	D	1/min	API	Azienda Sanitaria locale	Private API	protected	
F40	Triage capacity of each emergency room	Healthcare	S	NA		Azienda Sanitaria locale	Private API	Protected	
F41	Earthquake	Environment	D	2/hour		TO Be established INGV/ http://www.emsc-csem.org	http://www.ingv.it/it http://cnt.rm.ingv.it/	ND	
F42	Civil Protection Alert Messages	Social	D	2/hour		Civil Protection	http://protezionecivile.com		
F43	Presence at school	Service	P	1/Day	SHP	City of Florence	Private API	protected	
F44	Places of shelter	Structure	P	1/Year	SHP	Florence Civil Protection			
F45	Idrogeological Risk map	Environment	S	NA	SHP	Basin Authority	http://www.adbarno.it/opendata/	Italian Open Data Licence http://www.formez.it/iodl/	
F46	Bilancino basin level 2016	Environment	P	1/week	CSV	Basin Authority	http://www.adbarno.it/opendata/	Italian Open Data Licence http://www.formez.it/iodl/	
F47	Underpasses position (DST)	Infrastructure	S	1/Year	SHP	UNIFI-DST	ND	ND	
F48	Underpasses status	infrastructure	RT	NA	SHP	City of Florence	ND	ND	
F49	Fuel suppliers position	Service	P	NA	SHP	MISE- Italian Minstry of economic developemnt	ND	ND	
F50	Fuel supplier description	Service	P	NA	JSON	MISE- Italian Minstry of economic developemnt	ND		

F51	KM4City app user position	Human Behaviour	RT	1/min	HTTP POST	DISIT	protected		
F52	Green areas	Structure	P	1/week	SHP	City of Florence	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F53	Community gardens	Structure	P	1/week	SHP	City of Florence	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F54	Neighbourhoods	Structure	S	NA	SHP	City of Florence	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F55	Admission areas	Structure	S	NA	SHP	Florence Civil Protection	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F56	Waiting areas	Structure	S	NA	SHP	Florence Civil Protection	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F57	Gathering areas	Structure	S	NA	SHP	Florence Civil Protection	http://datigis.comune.fi.it	https://creativecommons.org/licenses/by/3.0/it/	
F58	Civil protection volunteers availability number	Service	D	Real time	API	Florence Civil Protection	ND	ND	
F59	Meteo forecast	Environment	D	Real time	API	LAMMA	protected	ND	
F60	Anemometer sensor position	Environment	S	NA	API	LAMMA	protected	ND	
F61	Anemometer measurement	Environment	D	Real time	API	LAMMA	protected	ND	
F62	Rainfall sensor position	Environment	S	NA		LAMMA/Autorita' di Bacino	http://www.adbarno.it/pagine_sito_opendata/gds_md_scheda_ridotta.php?id_ds=186	http://www.formez.it/iodl/	
F63	Rainfall measurement	Environment	D	Real time		LAMMA/Regione Toscana	http://www.cfr.toscana.it/index.php?IDS=42&IDSS=277	http://www.formez.it/iodl/	

F64	Hydrometry sensors position	Environment	S	NA		Basin Authority	http://www.adbarno.it/pagine_sito_opendata/gds_md_scheda_ridotta.php?id_ds=1090	http://www.formez.it/iodl/	
F65	Hydrometry measurement	Environment	D	Real time		Basin Authority	http://www.cfr.toscana.it/index.php?IDS=42&IDSS=276	http://www.formez.it/iodl/	
F66	Hygrometry sensors position	Environment	S	NA		LAMMA/Regione Toscana			
F67	Hygrometry measurement	Environment	D	Real time		LAMMA/Regione Toscana			
F68	Cultural heritage at risks of Flooding	Environment	S		SHP		http://www.adbarno.it/pagine_sito_opendata/gds_md_scheda_ridotta.php?id_ds=321	http://www.formez.it/iodl/	
F69	River network	Environment	S		SHP				
F70	Potential flooding area of Arno river	Environment	S	NA	SHP	Basin Authority	http://www.adbarno.it/pagine_sito_opendata/gds_md_scheda_ridotta.php?id_ds=321	http://www.formez.it/iodl/	
F71	flooding susceptibility areas (DST)	Environment	S	NA	SHP	University of Florence			
F72	Tramway map	Mobility	S	NA	SHP	Citta' Metro Firenze	TBD	private	Tramway map topology, Tramway stops location
F73	Tramway schedule	Mobility	S	NA	TBD	Citta' Metro Firenze	TBD	private	Train ID, Train schedule details

F74	Tramway tram position	Mobility	RT	Real time	API	Citta' Metro Firenze	TBD	private	Train ID, Real-time Position
F75	Tramway tram scheduling	Mobility	RT	Real Time	API	Citta' Metro Firenze	NA	private	Train ID, Service ID, Run ID, Scheduled time of arrival, Foreseen time of arrival
F76	Wifi connected devices	Structure	RT	Real Time	API	Tramway WiFi System (Area: Florence)	NA		
F77	Assistance areas for population	Structure	S	NA	SHP	Florence city council	http://opendata.comune.fi.it/statistica_territorio/dataset_0373.html	https://creativecommons.org/licenses/by/3.0/it/	
F78	Buildings	Structure	S	NA	SHP	Florence city council	http://opendata.comune.fi.it/statistica_territorio/dataset_0306.html	https://creativecommons.org/licenses/by/3.0/it/	
F79	Meeting points for rescuers and resources	Structural	S	NA	SHP	Florence city council	http://opendata.comune.fi.it/statistica_territorio/dataset_0308.html	https://creativecommons.org/licenses/by/3.0/it/	
F80	Waiting areas for population	Structural	S	NA	SHP	Florence city council	http://opendata.comune.fi.it/statistica_territorio/dataset_0307.html	https://creativecommons.org/licenses/by/3.0/it/	
F81	Wifi - connection	Behavior				Florece city council		protected	Wifi connection

F82	Twitter data (NLP/SA)	Social	D	Real time	JSON	DISIT	Twitter APIs	ND	Twitters produced by users
F83	Mobility events	Event	D	Real time	WFS	Citta Metropolitana	http://pubblicazioni.cittametropolitana.fi.it/geoserver/wms/wms?service=WFS&version=1.0.0&request=GetFeature&typeName=topp:infomob_eventi_di_traffico&maxFeatures=50&outputFormat=text%2Fxml%3B%20subtype%3Dgml%2F2.1.2	Open Data	Critical events in the city as incidents, roadway reduction, etc.

ANNEX B - EARLY DATASET SELECTION FOR ATHENS PILOT

COD	Name	Category	Dynamic	Frequency update period	Format	Data Provider	Data URL	Licence/Condition of use	Dataset description
A1	OASA	Mobility	S/P/RT	Depends on the dataset choosen	CKAN API	OASA (Area: Athens)	http://geodata.gov.gr/en/	Creative Commons Attribution 3.0	
A2	OASA	Mobility	RT	min	APIs	Trasports for Athens - O.A.S.A. - Real time information for busses and Trolleys (Area: Athens)	http://oasa-telematics-api.readthedocs.io/en/latest/	ND	webGetLines webGetLinesWithMLInfo webGetRoutes webGetStops webRouteDetails webRoutesForStop webGetRoutesDetailsAndStops getStopArrivals getBusLocation getScheduleDaysMasterline getLinesAndRoutesForMl getRoutesForLine getMLName getLineName getRouteName getStopNameAndXY getSchedLines
A3	Ministry of Development and Competitiveness	Energy	P	Day	SMS to 54141	Ministry of Development and Competitiveness (Area: Greece)	http://www.fuelprices.gr/	Free	Fuel prices Text
A4	Ministry of Health	Healthcare	P	Day	Word, PDF	Ministry of Health (Area: Attica region)	http://www.moh.gov.gr/articles/citizen/efhmerie	Free	Hospitals

							s-nosokomeiwn/68-efhmeries-nosokomeiwn-attikhs		
A5	OpenStreetMap	Infrastructure	S	Uknown	OSM.xml	OpenStreetMap Foundation (OSMF) (Area: World)	https://www.openstreetmap.org/#	Open Data Commons Open Database Licence (ODbL)	Street Network OSM.xml
A6	Attikes Diadromes	Mobility	RT	min	XML	Attikes Diadromes (en.aodos.gr) (Area: Attiki Odos & Imittos Ring Road)	ftp	Upon request	VDS Raw Data about Speed (Km/h) and Traffic Flow (veh/hour) Fields: VDS ID, DateTime, Traffic Flow, Speed, VDS status
	Weather Data	Environment	RT (limit of 500 requests per day)	NA	CSV	CERTH via WORLD WEATHER ONLINE; URL: http://developer.worldweatheronlD82ine.com/ (Area: worldwide)	N/A	https://developer.worldweatheronline.com/api/free-api-terms.aspx	Weather data: • Historical data from 1st July 2008 until present day • Current weather plus up to 14 day weather (1, 3, 6, 12, 24 hourly interval) • Monthly climate average data Fields: **DateTime:** Time of the observation in UTC, **TempC:** The temperature in degrees Celsius, **feelsLikeC:** Feels like temperature in degrees Celsius, **dewPointC:** Dew point temperature in degrees Celsius, **heatIndexC:** Heat index temperature in degrees Celsius, **windChillC:** Wind chill temperature in degrees Celsius, **cloudcover%:** Cloud cover amount in percentage (%), **visibilityKm:** Visibility in kilometres, **pressureMB:** Atmospheric pressure in millibars, **humidity%:** Humidity in percentage, **precipMM:** Precipitation in millimetres, **windSpeedKmph:** Wind speed in kilometers per hour, **windGustKmph:** Wind gust in kilometers per hour, **windDirDegr:** Wind direction in degrees

	Bing Incidents	Mobility	RT	N/A	CSV	CERTH via Bing Maps; URL: http://www.bing.com/maps/ (Area: worldwide)	N/A	https://www.microsoft.com/maps/product/terms.html	Information about traffic incidents and issues, such as construction sites and traffic congestion. Fileds: **estimatedTotal:** Number of incidents in the requested area; **startPoint:** The latitude and longitude coordinates where you encounter the incident; **endPoint:** The coordinates of the end of a traffic incident, such as the end of a construction zone; description: A description of the incident. **strartDateTime:** The time the incident occurred; **endDateTime:** The time that the traffic incident will end.; **incidentId:** A unique ID for the incident; **lastModified:** The time the incident information was last updated; **roadClosed:** A value of true indicates that there is a road closure; **severity:** Specifies the level of importance of incident (LowImpact, Minor, Moderate, Serious); **type:** Specifies the type of incident (Accident, Congestion, DisabledVehicle, MassTransit, Miscellaneous, OtherNews, PlannedEvent, RoadHazard, Construction, Alert, Weather); **verified:** A value of true indicates that the incident has been visually verified or otherwise officially confirmed by a source like the local police department;
A7	Rss Reader	Social	RT	N/A	CSV	CERTH via RSS Feeds (Area: worldwide)	N/A	N/A	Rss aggregator **link:** Link of the article **title:** Title of the article **dateTime:** Date and time of the release **sourceLink:** Link of the source page **source:** Name of the source page **text:** The description of the article
A8	Twitter Parser	Social	RT	N/A	CSV	CERTH via Twitter API (Area: worldwide)	N/A	https://dev.twitter.com/overview/terms	The Twitter parser provides information about: **1.** Tweets containing a specific word around a specified area (ex Tweets containing the word "Arno" from users living at City of Florence) **2.** Followings, Followers of a specific user Fields: **1. tweet_id:** Tweet ID

									city: City of user **time:** Publish time of the tweet **userID(int):** User IDwho published the tweet. **screen_name:** Screen name of the user who published the tweet. **type(int):** types: 0. Original tweet, 1. retweet **text:** Text of the tweet **2. userID:** User ID of a specific user(UserA) **followings:** ID of a user who are followed by UserA **followers:** ID of a user who follows the UserA