Dynamic spatio-temporal health impact assessments using geolocated population-based data: the PULSE project

Domenico Vito [1], Riccardo Bellazzi [1], Vittorio Casella [2], Cristiana Larizza [1], Andrea Pogliaghi [3], Daniele Pala[2]

Università di Pavia LISRC Lab Pavia-Italy [1] ,Università di Pavia Dipartimento di Ingegneria Civile ed Architettura - Lab Geomatica Pavia-Italy [2],Genegis Milan-Italy [3]

dvito.pulse@gmail.com

Abstract. Despite the silent effects sometimes hidden to the major audience, air pollution is becoming one of the most impactful threat to global health.
Standing to the report of the European Environmental Agencies the number of deaths from cardiovascular disease that can be attributed only in Europe alone,is about 790.000 a year and each of these deaths affects an average reduction in life expectancy of more than two years: air pollution is addressed to be the cause of premature death in 41 European nations
This outcomes are enforced by the estimates of WHO, finding that air pollution is responsible for 120 extra deaths per year per 100,000 of the population.
Cities are the places where these deaths are concentrated most, as the consequences of bad air quality are more severe and localized. In order to correctly address intervention and prevention thus is essential to assest the risk and the impacts of air pollution spatially and temporally inside the urban spaces.
PULSE (Participatory Urban Living for Sustainable Environment) is a pioneer EU-financed project that aims to develop a set of models and technologies to predict and manage public health problems in cities and promote health.It aims to develop and test dynamic spatio-temporal health impact assessments using geolocated population-based data.
The project is currently active in eight pilot cities, Barcelona, Birmingham, New York, Paris, Singapore, Pavia, Keelung and Taiwan, following a participatory approach where citizen provide data through personal devices and the PulsAIR app, that are integrated with information from heterogeneous sources: open city data, health systems, urban sensors and satellites.
PULSE aims to design and build a large-scale data management system enabling real time analytics of flows of personal data
The objective is to reduce the environmental and behavioral risk of chronic disease incidence to allow timely and evidence-driven management of epidemiological episodes linked in particular to two pathologies; asthma and type 2 diabetes in adult populations. developing a policy-making across the domains of health, environment, transport, planning in the PULSE test bed cities.
The work will present the main frameworks of the project and the most relevant components of the decision support platform, such as satellite data processing, deployment of sensors, management of acquired spatial data, WebGIS and Dashboard tools to provide visualization of the correlations between epidemiologic and spatiotemporal data and models.act.

Keywords: Air Pollution, Health, Data, Participation.

1. Introduction

Despite progress in recent years, air pollution continues to be a serious environmental and health problem.
Air pollution consists of harmful or poisonous substances in outdoor or indoor air: this of course threatening for normal subject but also for people affected by respiratory and cardiovascular diseases like asthma and diabetes.
According to the WHO impacts deaths from air pollution reach 4.2 million annually due to air pollution (WHO, 2016), and one ouf of every nine deaths was the result of air pollution-related conditions.
The problem of air pollution currently affects all regions, settings, socioeconomics groups and of every age.
Standing to EEA, air pollution causes 790.000 new deaths per year and it reduces of 2 year the life expectancy and premature deaths in 41 EU countries.
The European Environmental Agency has estimated that 13 % of the EU-28 urban population was exposed to PM10 levels above the daily limit value and approximately 42 % was exposed to concentrations exceeding the stricter WHO AQG value for PM10 in 2016 (EEA, 2019).
Due to this urgency to intervene indeed air pollution has been identified as a global health priority in the sustainable development agenda.
Particularly air quality issues are addressed by Goal 3 related to health, Goal 7 related to energy and goal 11 "Sustainable cities and communities".
On the purpose of address sustainability at a city level is inserted the european project PULSE.
PULSE (Participatory Urban Living for Sustainable Environment) is an EU-financed project that aims to develop a set of models and technologies to predict and manage public health problems in cities and promote health. It follows a participatory approach where citizen provide data through personal devices, that are integrated with information from heterogeneous sources: open city data, health systems, urban sensors and satellites. The project deals with various issues concerning air quality, lifestyle and personal behavior and it aims to investigate the correlations between the exposure to atmospheric pollutants, the citizen habits and the health of the citizen themselves, focusing on Asthma and T2 diabetes.

2. The PULSE project: system architecture

PULSE is a participative project focused on well-being in communities. The final goal is to build extensible models and technologies to predict, mitigate and manage public health problems, and promote population health, in cities.
The eight pilots of the project– Barcelona, Birmingham, New York, Paris, Singapore, Pavia, Keelung and Taiwan– can be defined as "Smart Cities". "Smart Cities and Communities" embrace integrated IT infrastructure and solutions, and citizen services, across city sectors, including health.
To accomplish the transformation of public health systems, and stimulate the development of intersectorial policy in cities, PULSE leverages large amounts of data from city governments, health systems, and citizens.
Beyond the collection of existing data, PULSE undertakes the following:

- implement a novel environmental/health surveillance system on air quality within specific neighborhoods and model risk of exposure to polluted air for citizens, especially those with asthma; develop novel insights on the relationship between risk for the onset of T2D and environmental and behavioral factors;
- collect comprehensive data on individual and community well-being;
- model public health risk and resilience and develop tools and technologies to intervene and change behaviour, translating Big Data to Policy.

Currently the project pilots are running in 8 cities; the Consortium is implementing dashboards on support to Public Health Organization to visualize all the gathered data, the outputs of the health risk models and the simulation tools to assess neighbourhoods air quality conditions and the health status of the citizens.
The core of the PULSE project is the data management architecture: it is designed in order to collect and elaborate a multivariate set of data.
Figure 1 frames the basic concept and the main elements of PULSE architecture: it is made by several modules that are able to collect data from heterogeneous sources.
The different modules of the PULSE infrastructure enable the management of participative data from the PulsAIR App, geolocated Satellite (MODIS, Sentinel, Landstast) and also from static AIR sensors that can be from existing official AQ stations but also from from a dedicated dense networks of lowcost sensors, installed on test bed cities by the PULSE project.
The collected data are used to generate maps of environmental air quality,calculate personal exposure and to feed health risk models that assess the risk of asthma and diabetes T2 occurrence due both to air quality condition and subject health status. The PulsAIR app in facts is designed both to connect to a wearable eonnected to FitBit, Garmin Asus health tracker devices and to interact and to collect data from citizens by questionnaires and gamification techniques related to subjective health status, physiology and daily activity and current location.

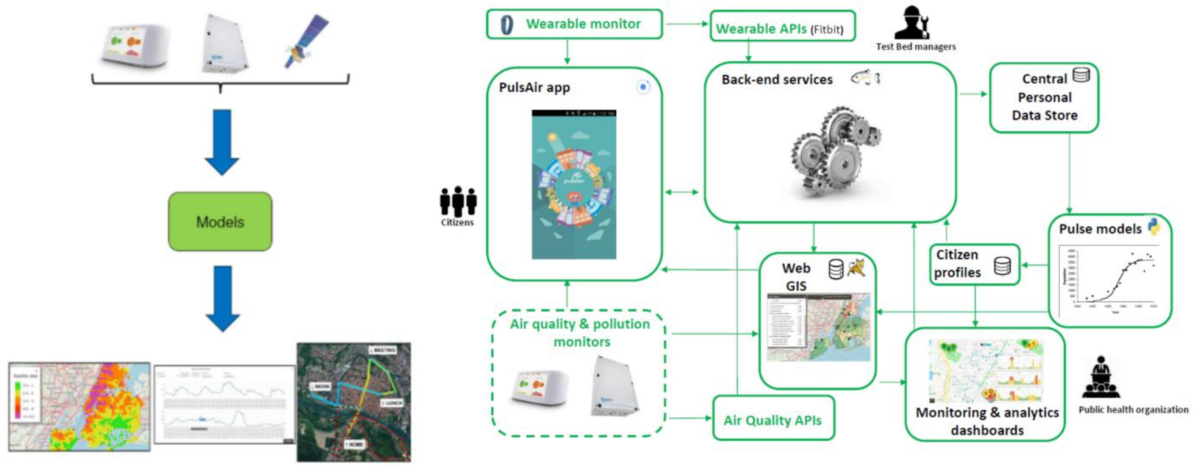
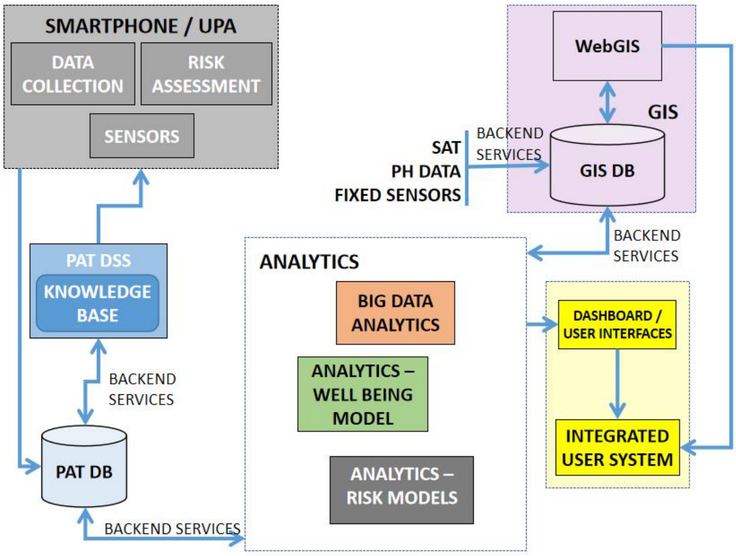
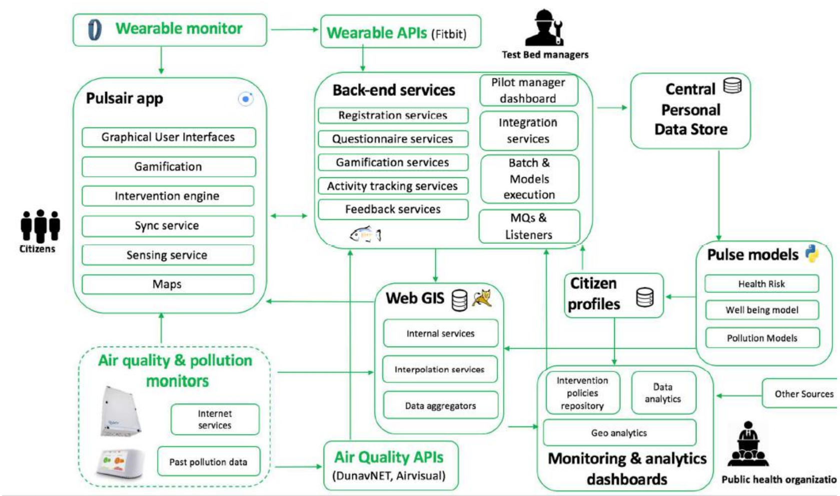


Figure 1. A. PULSE basic concept B. Pulse Architecture: Logic structrure

To manage the heterogeneous set of data the inner composition of the data architecture is structured in different modules and repositories.
Particularly the data related to health and behaviour coming directly from the users through PulsAIR are stored into a central personal data storage, and the outputs of the health model are backuped into a citizen profiles repository.
The environmental data, coming from the AQ stations and the diffuse network sensors, as well as the output environmental maps are stored into the WebGIS geodatabase (Figure 2.A).
The cross talking among the different repositories, the conversion between the different data within common measure units and timelapses is managed by a series of Backend Services hosted in a remote server.
The backend services also include all the necessary routines for user-data interactions provided by the PulsAIR app and the Fit-bit devices.



A. Database Structure



B. Full Overview of the interactions

Figure 2. PULSE architecture: service structure

2.1. PulsAIR App

PulsAIR is a novel participatory citizen science-based mobile application used to empower the citizens in their perceiving of urban environmental and health status.
The app has been developed to foster a healthy lifestyle and to make people more aware about the air pollution in the city.
PulsAIR is available both for iOS and Androis has been designed following a user-centred approach based on the goal-oriented design (GOD) methodology, in which end users and stakeholders guide the process, and ultimately validate the final product (Ottaviano et al 2019) .
The main purpose of the app is of getting a positive behaviour change regarding healthy and green habits.
Figure 3 resumes the main interfaces of the PulsAIR app

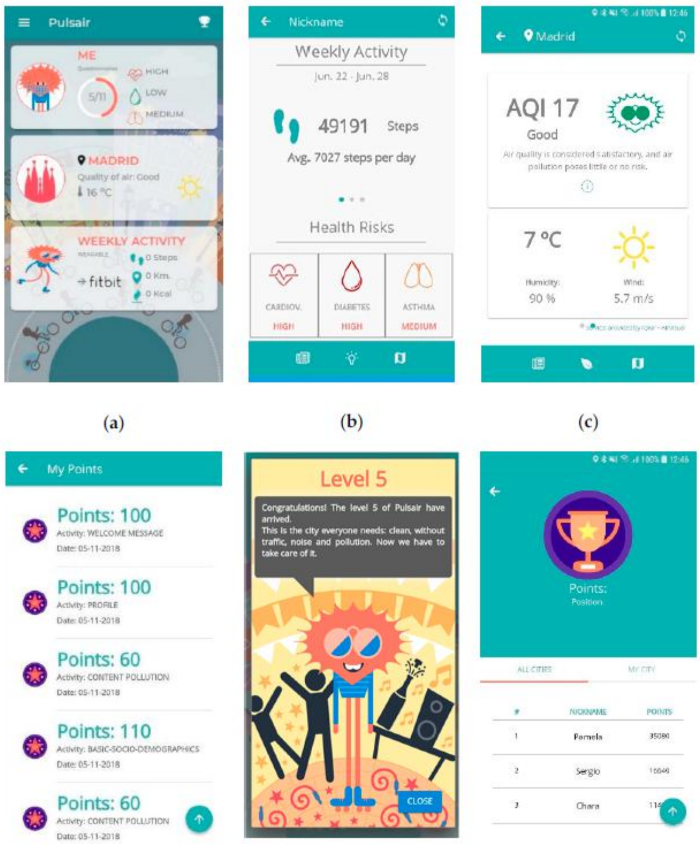


Figure 3. PulsAIR mobile application interface screenshots: (a) Home Menu; (b) "Me" module; (c) "My City" module; (d) "My Points" section; (e) Level up message; (f) Leaderboard

As schematized in Figure 4, PulsAIR can be connected to FitBit, Garmin and Asus health tracker devices: with this system the citizens can provide data on subjective health status though guided questionnaires and physiological and daily habits data for the wearable sensor.
The app can return back to the users health risks (of asthma and type 2 Diabetes) and it suggests specific behavioural changes for an improve on health that are delivered though a supportive feedback.
The app can also show user the exposure to the air pollutants by combining the information of the position (the GPS data) with the data coming from the environmental sensors that measure air pollutants (PM 2.5 and CO etc..)

Figure 4. PulsAIR connection with FitBit Sensors and interaction with Air Quality Sensor Network

To foster the active engagement of the individuals in the air quality health policies, the PulsAIR has been developed integrating game-design elements, to allows an entertaining and informative engagement experience around several aspects such as encourage people to do things they consider otherwise boring, rise user engagement, enhance attention, and increase motivation.
Particularly it exploit a rewarding system on the performance of healthy habits and information on air quality as described in Figure 5.A and 5.B.

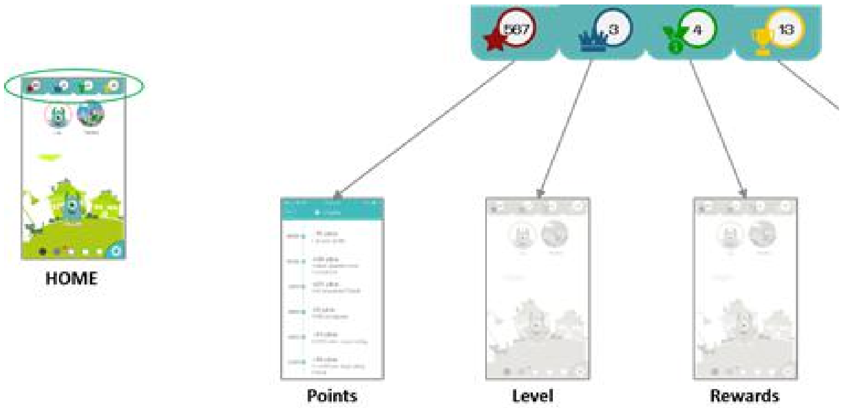


Figure 5.A PulsAIR App Game Views

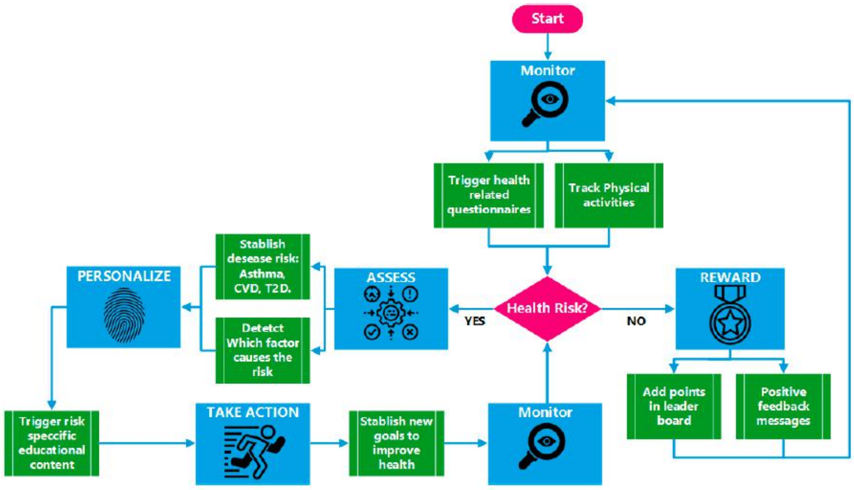


Figure 5.B PulsAIR workflow, CVD cardiovascular diseases (Ottaviano et al 2019)

Inside the PULSE project, PulsAIR App is the first user interface for data collection from users to create a data back-end ecosystem.
The data back-end ecosystem has been implemented on support to the interoperability of the data flows
The ecosystem allows a continuous interaction with the mobile application, enabling a participative and real time data gathering and integration among different sources.
The data back-end ecosystem can be considered as framed on four overlapping data contextual layers that are mirrored into the back-end infrastructure for data processing ad visualization and fed through the mobile application, enabling real-time information delivery (Figure 6)

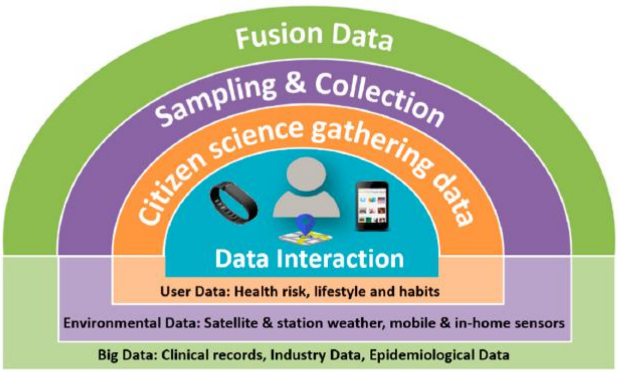


Figure 6. Data backend ecosystem framework (Ottaviano et al 2019)

Coupled with the data ecosystem on the backend it can be seen as a a “toolkit” that collects and process information that is rarely integrated as health data from citizens and environmental data on pollution stored in the same repository.
This framework gives a citizen-centric set of applications and data services that can contribute to address in a remote manner, changing the way that environmental research, monitoring, and policy-making are carried out.

2.2. Dense network of air quality sensors

The PULSE air quality sensor's system is composed of multiple type of sensors and sensor's datasets, historic and real-time data, for all pollutants (PM 2.5, PM10, NO_2, O_3, SO_2).
The network aims on monitor the variable trends in emission within urban areas with an high resolution, to appropriately address the temporal and spatial scales where usually pollutants are spread To have a capillary coverage of the sampling the project integrates mobile sensors and mobile network of sensors in order to establish a significantly enhanced monitoring system.
Two types of sensors are used across pilots that are the AQ10x of DunavNetn (20+, deployed in all pilots) and PurpleAir PA-II sensor (30+ in Pavia – started acquisition in 2018): they allowed to have an high frequency measurement providing a sample every minute.
Those two types of sensors are easily deployable in order to build both a mobile measurements station or a low-cost network of sensors, changing the paradigm to citizen participative urban network.
Before all the test bed acquisition the sensors has been tested and calibrated in the city of Barcelona for 2 weeks.

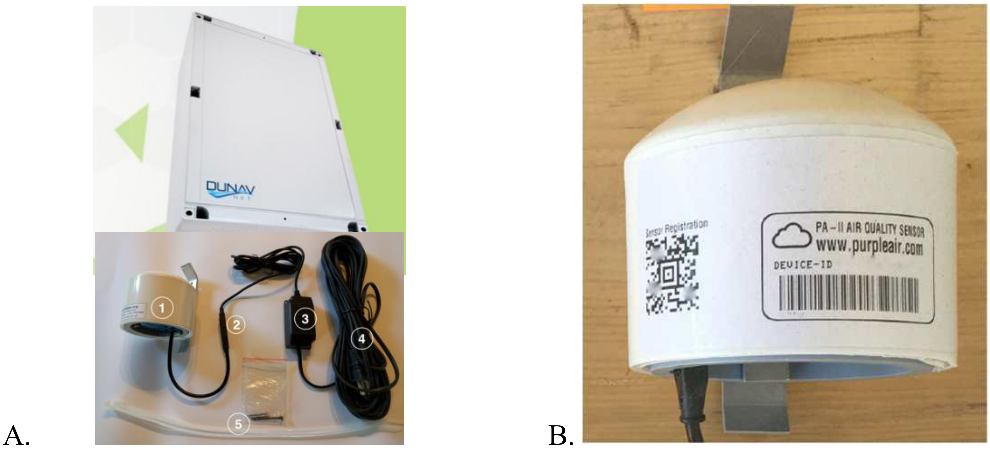


Figure 7. Network sensors: A. AQ10x of DunavNet sensor B. PurpleAir PA-II sensor

2.3. The WebGis Module

The pulse WebGIS module gathers and geo-refers the heterogeneous sets of data gathered in PULSE.
The repositories of PULSE particularly collect socio-economic data (level of education, poverty rates, unemployment and violence rate), environmental data (concentration of fine dust PM 2.5 and PM10, concentration of NOx), demographic data (access to healthier food stores and dinners, ortotransport and mobility, use of public transport), health and subject behavioral data (hospitalizations for asthma and T2D, prevalence, mortality, alcohol, smoking, physical inactivity).
The WebGIS module allow to view the collected data using maps that help to graphically show relationships between data and models that would otherwise remain hidden without any geographical representation.
The module is basically composed by 8 WebGIS , one for each pilot site, working on 3 levels (Figure 8).The "Data Ingestion layer" level gathers the data from the different sources (sensors, apps, satellite) and allows also to integrate further data by uploading in .csv formats
The "Processing layer" processes and creates the geostatistics and the maps of interest and the "visualization layer" displays in maps that can be temporally and spatially explored for each pilot site.

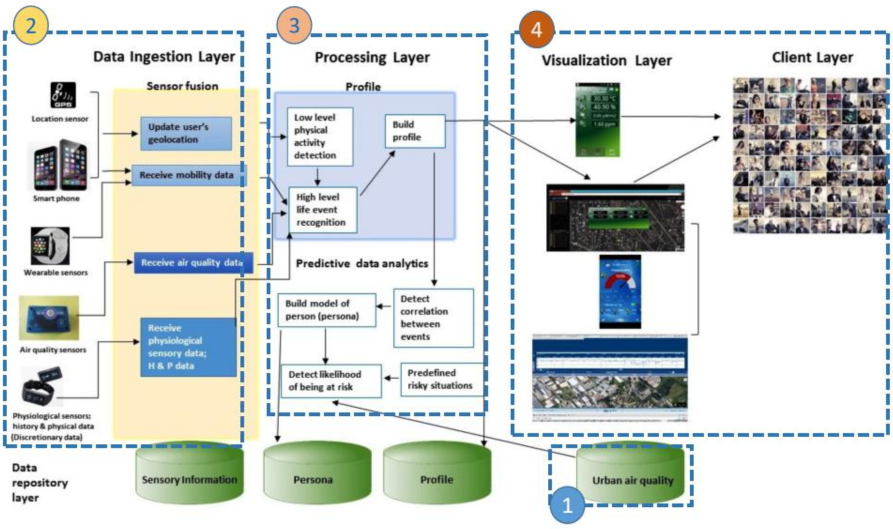


Figure 8. WebGIS module three layer structure

The WebGIS provides the geospatial services used by other Pulse components.
It offers basic and advanced features by exploiting the data and maps (Figure 9).

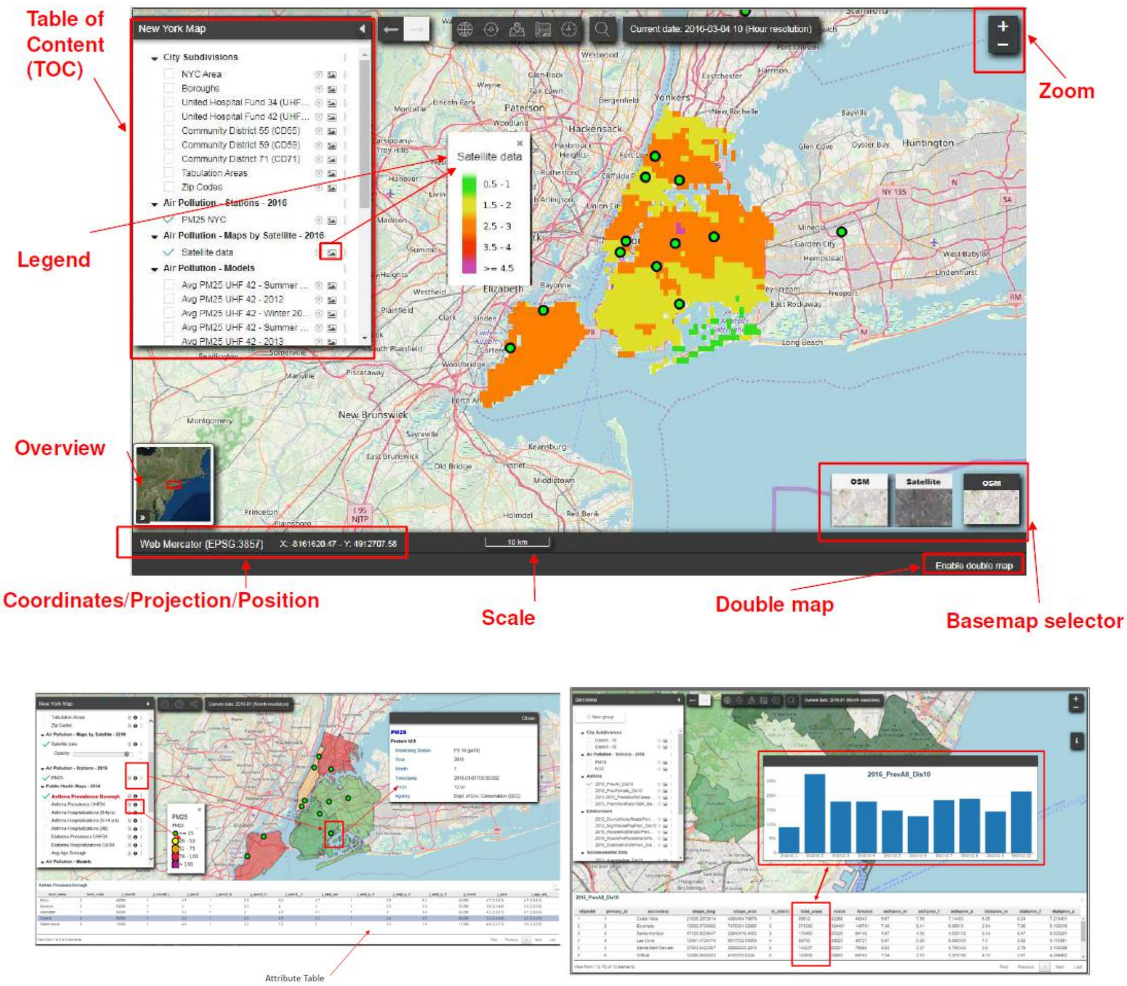


Figure 9. WebGIS module three layer structure

The standard functionality include zoom, pan extent, geocoding, current position, history navigation, the possibility to execute featured query and to show the table of attributes of the geodatabase, and finally to download of the data in .cvs format.
The WebGIS module is composed by two parts *the Viewer* and *the Configurator* (Figure 10)

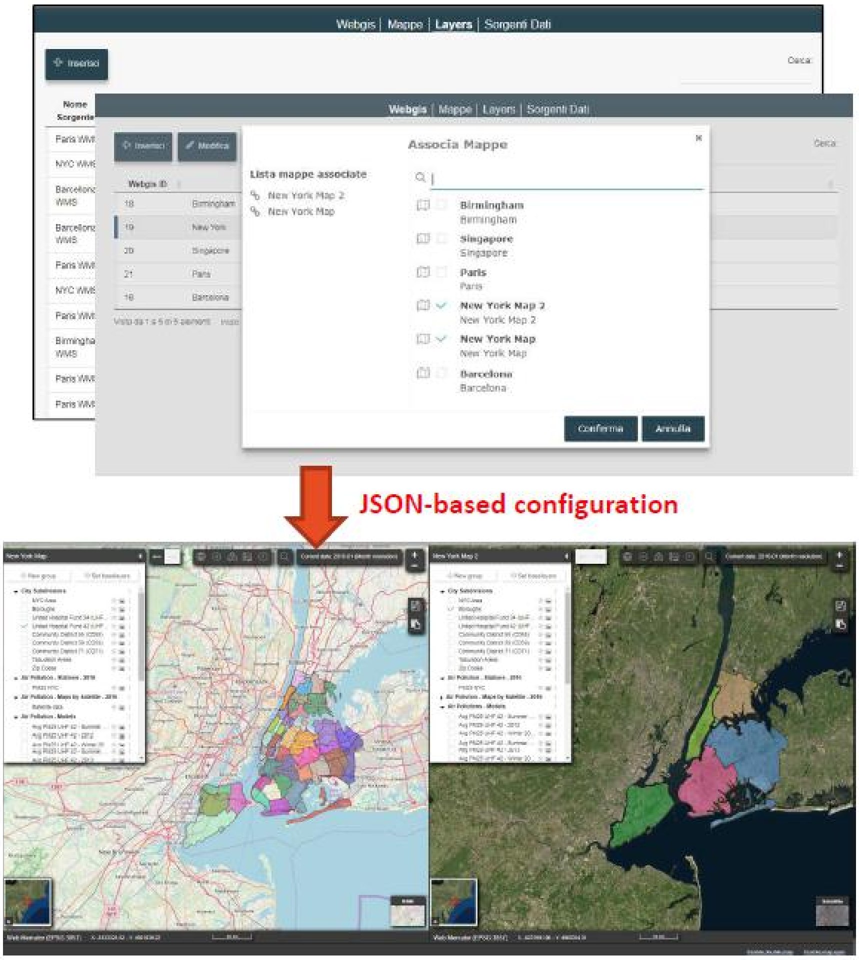


Figure 10. Architecture of the WebGIS module

The Viewer is the basic visualization interface of the module that shows the data as requested
The Configurator tool simplifies the creation of the WebGIS for the pilot cities, and the layer registry: it supports also the automatic data ingestion from external files, both combining with the existent layer from the WebGIS DB and GeoServer.
The Configurator is also a map composer, through the functionalities of layer choice and widgets to set the data and visualization permissions.
The Configurator is the underline element for the advanced functionalities of the WebGIS.

2.4. *Advanced Application of WebGIS module*

Beside the basic ones the WebGIS module has also some advanced functions for a deeper exploitation of the data.

First of all the WebGIS module allows to put two maps side-by.side (Figure in order to visually make confrontations or visual interpretation of pattern and correlations and to compare data (e.g asthma hospitalizations and air quality) or same data at different times (Figure 11). The side-by.-side maps can support also geostatistical analysis to make more explicit the relationship between some parameters (for instance correlation between poverty and hospitalizations for asthma).

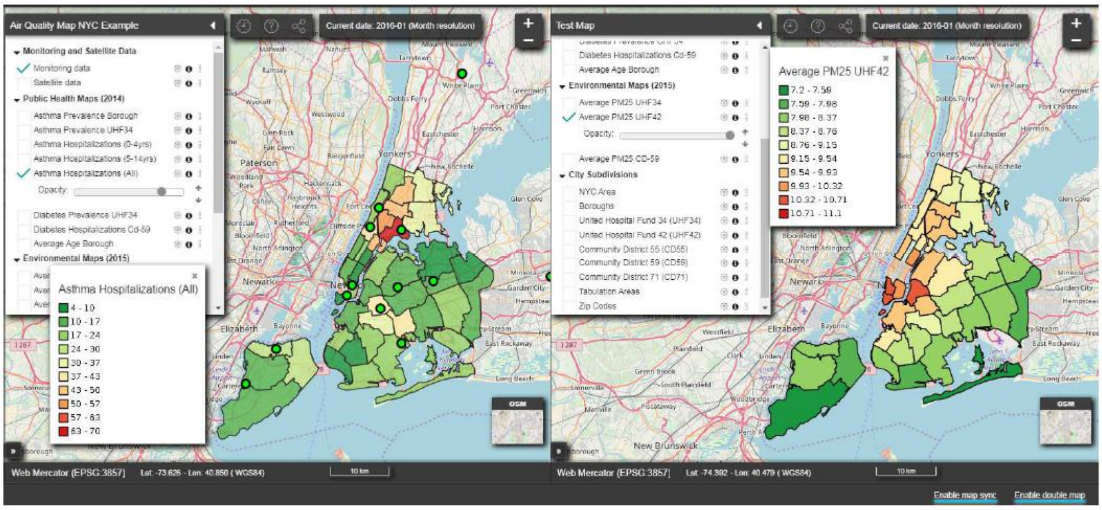


Figure 11. Two maps side-by.side feature

In addition to side-by-side maps, that can be useful to visually asses potential correlation, data can be analyzed also via geostatistical functions. These analysis allow to make more explicit the relationships between variables of different clusters. Figure 12 show for instance the geostatistical correlation between poverty and hospitalizations for asthma.

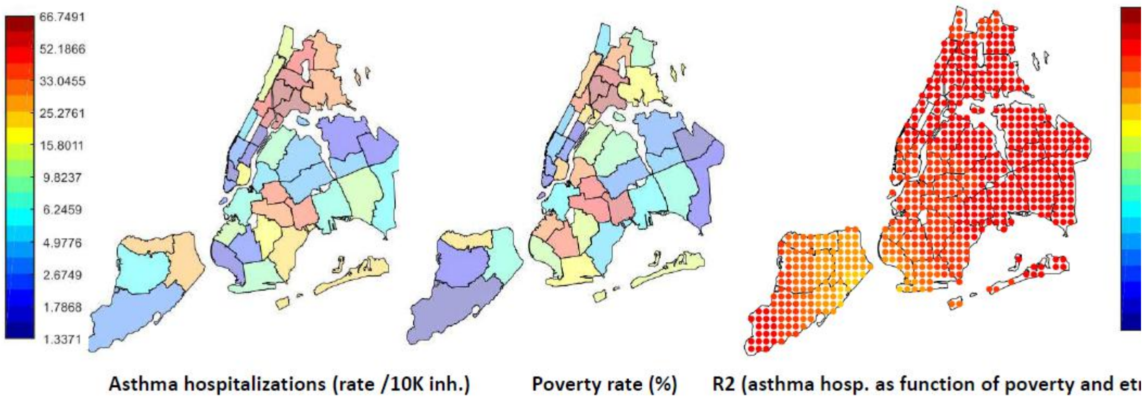


Figure 12. Geostatical tool application between asthma hospitalization and poverty rate

Another advanced feature of the WebGIS is the *dynamic thematization* that is a data-driven on-the fly definition and application of a style to a layer. It si possible to style an existing layer using external parameters (via API or CSV): Figure 13 show an example of a district thematization applying different colors to the districts according to an external parameter A.

By the advanced features of WebGIS is possible to navigate data spatially but also temporally.

The *temporal navigator* manage to explore data series (e.g.: satellite data / air quality monitoring measurements) by time, allowing to analyze the evolution or variation of a phenomena in different in.

The navigator support different temporal resolutions: the native is hourly, but there are also aggregated as hourly, daily, monthly. It supports also different statistical operators as temporal and spatial average.

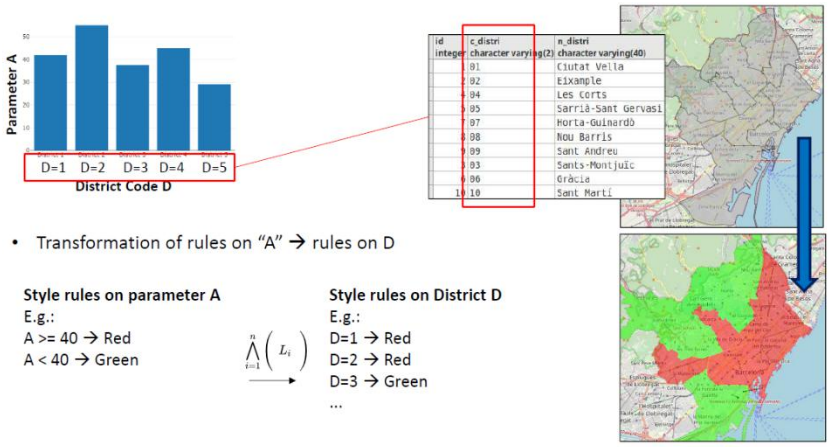
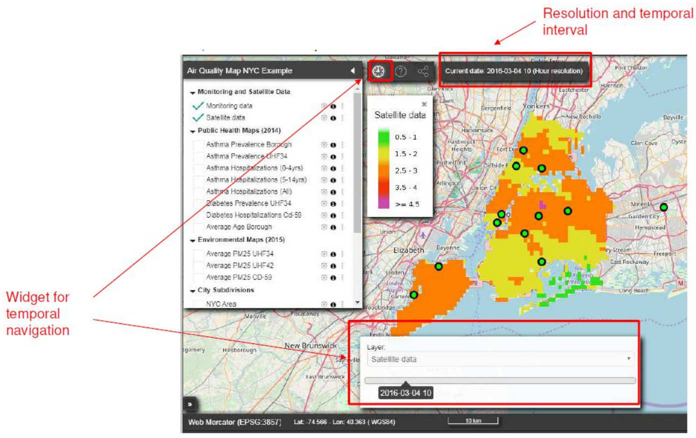
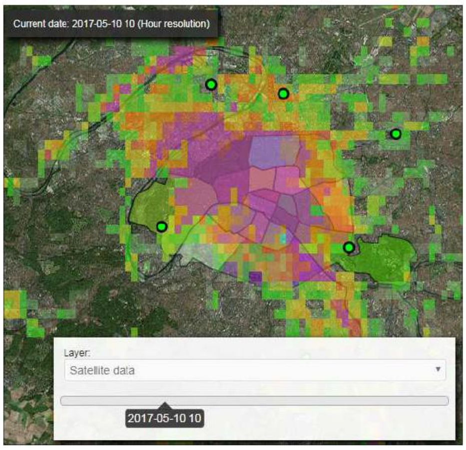


Figure 13. Example of a dynamic thematization of a "district style" based on an external parameter A



A.

Air quality satellite data (hour resolution)



Monitoring station data (NOX, monthly/daily/hourly average values)

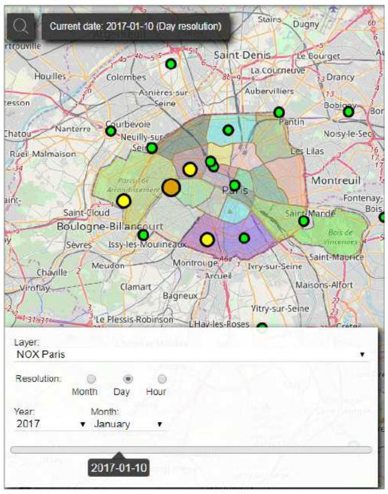


Figure 14. Screenshots of the WebGIS temporal navigator with A.basic interface and B. time lapse statistical data manipulation

2.5. The monitor & analytics dashboards

The PULSE project is provided bye a dedicated for pilot management to follow-up the upper level and aggregated indicators . It is equipped by visual monitoring and analytics in order to assess the quality and statistic for the collected data.

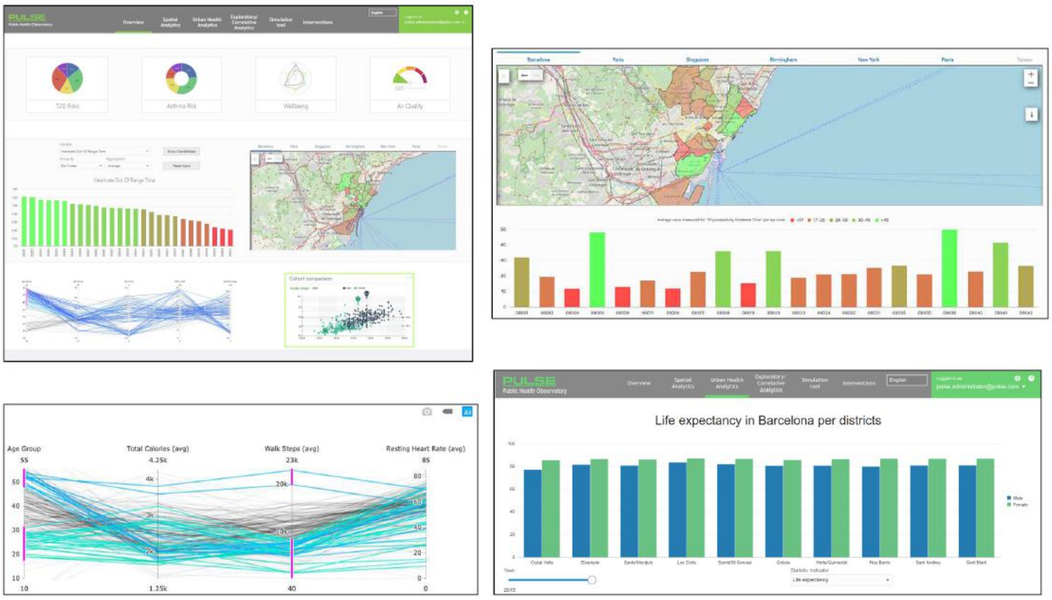


Figure 15. PULSE monitor and analytics dashboard

3. Example of spatio-temporal health impact assessments

On the following paragraph some examples of application of PULSE will be exposed showing how the combination of multivariate date can support the spatio-temporal assessment of the risk on health related to air quality.

Figure 16 reflect the schematic paradigm on how the PULSE platform works to reach this objective. Using existing open data and acquired data, and exploiting spatiotemporal models on health risks, asthma and diabetes T2D risk management data and environmental exposure can be assessed.

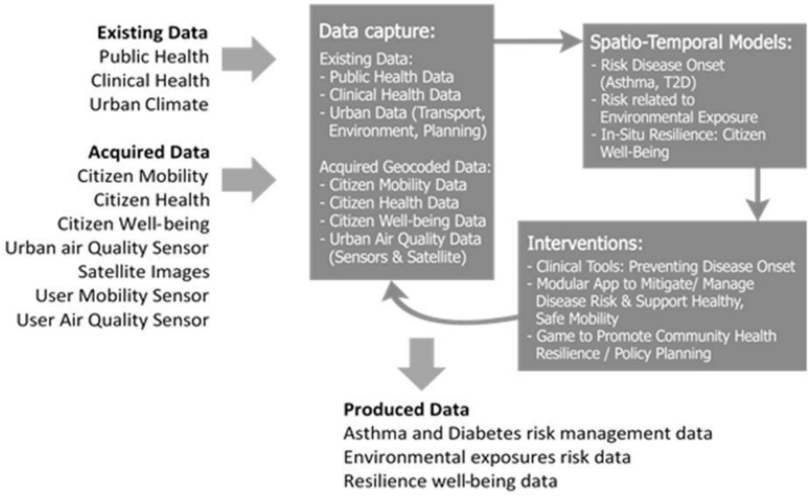


Figure 16. Spatio-temporal health impact assessments on PULSE; schematic paradigm

3.1. *The MODIS satellite data for air quality*

One of the source of external data integrated in PULSE has regarded the geographical distribution of air quality: particularly the data coming MODIS (Moderate Resolution Imaging Spectroradiometer) satellite sensor of NASA has been integrated to produce diffuse air quality maps.
The integration has required a suitable calibration and validation of parameters based on a relationship between the thermal band recorded by the space-borne sensor and the air pollutants. This technique permitted to reach a resolution of 500 meters, allowing the generation of one map per date at a fixed time corresponding to the transit time of the satellite.
In terms of numbers from the MODIS data, the test beds has been provided by more than 50 maps for NYC, more than 25 maps for Paris, and more than 20 for Barcelona.

3.2. *Land surface maps to assess Urban Heat Island effect*

Other types of satellite data used in PULSE have been the ones produced by Landsat-8 and Sentinel. These data have been exploited to assess the "Urban heat island" UHI effect (Manley 1958).
The UHI phenomenon is generally considered as caused by a reduction in latent heat flux and an increase in sensible heat in urban areas as vegetated and evaporating soil surfaces are replaced by relatively impervious low albedo paving and building materials that have an higher heat capacity and conductivity compared to bare soil and vegetation such as concrete, asphalt, metal.
This creates a difference in temperature between urban and surrounding non-urban areas (Manley 1958).
The UHI can be monitored through Land Surface Tempature maps obtained by the Landsat-8 and Sentinel satellite images (Imhoff et al. 2010). Land Surface Temperature (LST) is an important parameter to be assessed for air quality as an increase in LST determines an higher mobilization of air pollutants, greenhouse gasses and ground-level of ozone (smog) and particulate.
Figure 17 shows an example of LST map for the city of Pavia, with the LST calculation used in PULSE.
The used calculation procedure is able to obtain maps for a spatial resolution of 30m and a temporal resolution of 2 week by the Landsat 8 images and maps of 10m of spatial resolution for a 5 days of temporal resolution.

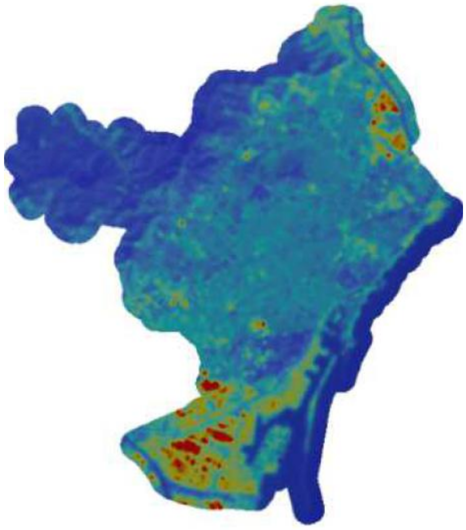


Figure 16. Example of Land Surface Temperature map for the city of Pavia

3.3. Personal Exposure evaluation

Personal exposure is a concept used by the epidemiologists to quantify the amount of pollution that each individual is exposed to, as a consequence of the living environment , habits etc.
Personal exposure has been evaluated in PULSE combining the information from the dense network of low-cost sensors and the data on habits coming from the PulsAIR app. Following the sampling rate of the sensors the data has been calculated.
Figure 17 shows a map for the personal exposure to PM10 with an hourly frequency.

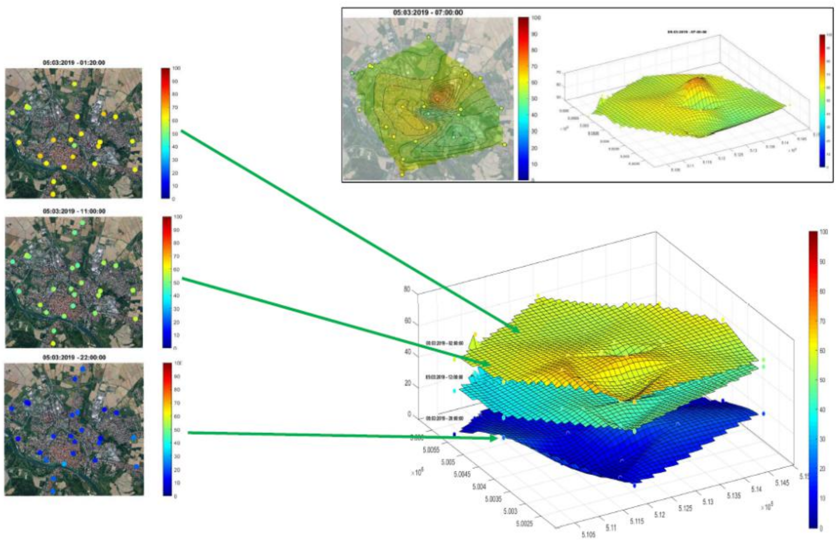


Figure 17. Personal Exposure Map to PM10

Exploiting the data from the PulsAIR app, FitBit and the personal exposure, an estimate of inhaled pollutant has been calculated in association to three classes of movement by the speed of body translation; standing, walking and running, considering the breaths per minute and the air volume per breath (Table 1).

Table 1. Page setup and measurements

Speed [km/h]		#breaths per min	Air volume in litres per breath
>=0	<=2	15	0.6
>2	<=6	28	1.8
>6		40	2.5

Personal exposure result could be also traced into exposure paths as in Figure 18 each dot in the movement line correspond to a time-lapse of 1 minute.

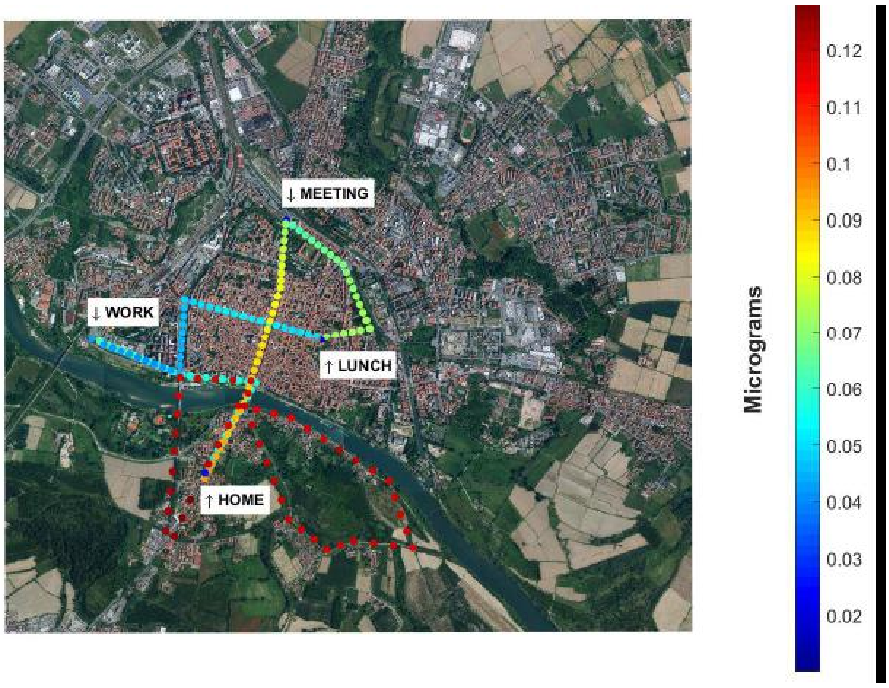


Figure 18. Personal Exposure Paths

4. The community of practices

The main aim of PULSE is to engage in a collaborative dialogue a range of stakeholders across the pilot studies to transform public health from a reactive to a predictive system focused on both risk and resilience.

On this aim PULSE works to raise awareness about those issues and empower stakeholders (Public Health Observatories, Community representatives, municipalities, health institutions, etc.) through innovative PULSE solutions.

Beside the development of the platform in each pilot site *communities of practices* , have been fostered in order to create the stakeholder ecosystems to transfer the collected data into knowledge and practices.
The communities of practices are supported by PULSE through a dedicated blog that post articles related to air quality and health impact issues and by the organization of capacity building events for practitioners on PULSE innovative system.

5. Discussion

PULSE offers an example of data driven decision making support system to orient health policies in response to air quality.
It foster the paradigm of citizen science and "putting the human in the loop" (Fraternali, 2012), that mean engage the citizens in the feedback loops related to the sensing of the environment.
The extremely detailed organization of the internal architecture, helps to drive the exploitation of the collected data in very different manners.
PULSE represent an innovative support on the collection of air quality data and on the support to urban health policies.
The installation of the sensors and of all technologies requires low cost hardware and software and furthermore most of the software is open source (Ottaviano et al 2019). Beside the low cost hardware a great potential of the project is the direct involvement of the citizens.
The participatory approach indeed brings several advantage, both allowing a more capillar sensing of the environment and a direct feedback communication with the citizens. Currently PULSE has reached more than 1611 in all pilot cities and is open to engage more users in order to increase the statistical significance of the results.The objective will be to involve at least 300 participants in each pilot, for a total of 2400 users involved.

6. Conclusions

The PULSE project has been successfully deployed in 8 pilot cities and a data architecture to collect data from environmental sensors, from a participatory app and from external sources.
The data gathering process has been based on several methodologies following both qualitative (questionnaires/interviews) and quantitative (sensors, open public and private data) approaches.
The PULSE architecture furnishes the possibility to combine the collected data in several ways in order to have a dynamic spatio-temporal assessment effects of air quality on asthma risks and type 2 diabetes: the PULSE data in fact feed and validate dedicated predictive models on asthma risk and type 2 diabetes.

Furthermore, the PULSE project offers the possibility to different stakeholders (PHOs, Community representatives, municipalities, health institutions, etc.) to join a common effort and collaborative work to raise awareness and address to a communitarian level the problem of urban air quality and its consequences on the health of the citizens.

7. References

EEA (European Environment Agency), 2017, European Union emission inventory report 1990-2015 under the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP), Copenhagen, 149 pages

EEA (European Environment Agency), 2018. Air quality management. https://www.eea.europa.eu/themes/air/explore-air-pollution-data#tab-air-quality- management accessed in 2019.

Fraternali, P., Castelletti, A., Soncini-Sessa, R., Vaca Ruiz, C., Rizzoli, A.E., 2012. Putting humans in the loop: social computing for water resources management, Environ Model Softw 37, 68–77

Imhoff, M. L., Zhang, P., Wolfe, R. E., & Bounoua, L. ,2010. Remote sensing of the urban heat island effect across biomes in the continental USA. Remote sensing of environment, 114(3), 504-513.

Manley, G. 1958. On the frequency of snowfall in metropolitan England. Quarterly Journal of the Royal Meteorological Society, 84, 70–72.

Ottaviano, M., Cabrera-Umpiérrez, M. F., & Waldmeyer, M. T. A., 2019. PULSE: Participatory Urban Living for Sustainable Environment. Proceeding of IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, pp. 62-63.

Ottaviano, M., Beltrán-Jaunsarás, M. E., Teriús-Padrón, J. G., García-Betances, R. I., González-Martínez, S., Cea, G., Arredondo Waldmeyer, M. T. , 2019. Empowering Citizens through Perceptual Sensing of Urban Environmental and Health Data Following a Participative Citizen Science Approach. Sensors, 19(13), 2940.

Raaschou-Nielsen, O. 2012.'Traffic air pollution and mortality from cardiovascular disease and all causes: a Danish cohort study', Environmental Health 11(60).

World Health Organization. (Ambient air pollution: a global assessment of exposure and burden of disease. World Health Organization), 2016. https://apps.who.int/iris/handle/10665/250141, burden of accessed in 2019

WHO (World Health Organization) Regional Office for Europe, 2012. Environmental Health Inequalities in Europe, 2012. http://www.euro.who.int/__data/assets/pdf_file/0010/157969/e96194.pdf accessed in 2019.

World Health Organization (World Health Organization) and UN Habitat, 2016. Global report on urban health: equitable, healthier cities for urban development., http://www.who.int/kobe_centre/publications/urban-global-report/en accessed on 2019.

World Health Organization (World Health Organization) and UN Habitat, 2016. Global report on urban health: equitable, healthier cities for urban development., http://www.who.int/kobe_centre/publications/urban-global-report/en accessed on 2019.

Acknowledgement

PULSE project has been founded by the European Union’s Horizon 2020 research and innovation programme, and it is documented in the grant agreement No 727816. Specifically. PULSE has been founded under the call H2020-EU-3.1.5. in the topic SCI-PM-18-2016-Big Data supporting Public Health policies. More information on http://www.project-pulse.eu.