# Apollon: Towards a Citizen Science Methodology for Urban Environmental Monitoring*

Antonella Longo[a,b,*], Marco Zappatore[a,b] and Mario A. Bochicchio[a]

[a] Dept. of Engineering Innovation - University of Salento - via Monteroni sn - 73100 Lecce (Italy)
[b] Hesplora srl, via Don Luigi Sturzo, 36 - 73100 Lecce (Italy)

ARTICLE INFO

*Keywords*:
Mobile Crowd Sensing
Citizen Science
Smart Cities
Internet of People

ABSTRACT

The collaborative power of ICT systems is a key enabler of social and technological advances providing multiple opportunities for public involvement in participatory activities, thanks to novel paradigms like citizen science and mobile crowd sensing. These paradigms, if applied according to specific methodologies, promise to increase the pervasive observation of urban environmental pollution either directly by human observers, or by means of crowd-sourcing data measurement tasks using sensors in smart phones or other mobile devices. We propose a platform, named Apollon, to enable scientists and others to take part in citizen science projects based on the exploitation of mobile devices. The platform has been implemented and validated in an educational context, in which students participate in urban environmental monitoring activities. In the paper, we describe the platform and the approach developed to produce successful experiments.

## 1. Introduction

The World Health Organization (WHO) states that the two main causes of pollution in urban areas are related to air quality and noise [57, 40]. In fact, according to [57] 91% of people living in cities worldwide experience poor air quality and 20.000 deaths per day (nearly 7M per year!) are due to exposure to both outdoor and indoor air pollution. According to [40] long-term exposure to environmental noise is estimated to cause 12.000 premature deaths and to contribute to 48.000 new cases of ischemic heart disease per year in the European territory.

This is why the WHO launched the "BreatheLife 2030" global campaign[1] with the ambitious goal of halving this figure by 2030. To achieve this change, two out of the three key actions defined by the campaign, namely: "Citywide solutions" and "Actions for individuals", can greatly benefit from the adoption of smart-city technologies and citizen science approaches.

In order to address these actions, the involvement of selected groups of citizen is pivotal. Fortunately, recent advances in ICT technologies have made mobile devices and cloud computing two pervasive enabling factors fostering the required systemic change. Moreover the increasing centrality of people in the Internet has moved the paradigm towards the so-called "Internet of People" [11, 10], where *"personal devices have become their proxies in the cyber world, in addition to acting as a fundamental tool to sense the physical world"*. This new paradigm, from the end–user's perspective, singles out some challenges with regard to the value and reliability of data collected and shared within the network of people. Moreover, it offers new opportunities to reinforce the human control of personal data opposite to the trend to concentrate personal data in a few proprietary platforms, or the chance to process and share data among nearby people rather than sending everything to cloud platforms.

In this scenario, professional scientists and researchers are more and more interested in the role that non-professional practitioners can play in new research projects, and how they can exploit profitably their contribution. The involvement of the public in scientific activities is not novel: in the last two decades, a considerable shift towards open science has appeared in several contexts [49]. Nowadays, technological advances make it possible for the public to address scientific investigations not only as data captors but also as independent non-professionals, capable of conducting research activities on their own or with a significant autonomy within global scientific projects, thus becoming part of the "Citizen Science" realm [37]. This trend is also supported by a changed mindset among scientists, who now realize the benefits achievable by engaging volunteers from the public in terms of cost-reduction, increased project application range, improved public awareness on significant topics and wider dissemination of research outcomes. However, several challenges have yet to be coped with (e.g., data provenance qualification, data validation, volunteer engagement and motivation strategies, privacy preservation, anonymization, etc.) [12, 3], the number of citizen science projects is increasing. According to a recent survey conducted in European cities [24], such activities now span across multiple domains, ranging from life sciences to engineering, and from cultural heritage to social sciences.

* This work was partially supported by the research project "APOLLON - environmentAl POLLution aNalyzer", within the "Bando INNONETWORK 2017" funded by Regione Puglia (Italy) in the framework of the "FESR - Fondo Europeo di Sviluppo Regionale" - "POR Puglia FESR FSE 2014-2020 - Asse Prioritario 1 - Ricerca, sviluppo tecnologico, innovazione - Azione 1.6".

* Corresponding author
antonella.longo@unisalento.it (A. Longo); marcosalvatore.zappatore@unisalento.it (M. Zappatore); marcozappatore@hesplora.it (M. Zappatore); mario.bochicchio@unisalento.it (M.A. Bochicchio)
ORCID(s): 0000-0002-6902-0160 (A. Longo); 0000-0002-8277-9390 (M. Zappatore); 0000-0002-9122-6317 (M.A. Bochicchio)

[1] https://breathelife2030.org/

Public involvement in citizen science activities, specifically tailored to the educational perspective and capable of addressing global - local (i.e. *glocal*) problems, requires a suitable methodology and proper technological solutions, such as leveraging of mobiles, miniaturized sensors, cloud computing resources and social networks. Mobile devices, more specifically, play a pivotal role in such a scenario, thanks to the Mobile Crowd Sensing (MCS) paradigm. Therefore, our investigation aims at exploiting the two emerging technological trends of citizen science and MCS in promoting civic engagement in science with significant educational outcomes.

In this paper we describe the lesson learned during the “Apollon Project” (formerly the "City SoundScape Project"), which aimed to support the "*BreatheLife 2030*" campaign by monitoring environmental pollution in two Italian regions and was poised to be expanded on a wider scale. Since 2015 our research has involved more than 500 students, teachers and families, and it has raised their awareness about the direct impact of daily activities on the environment. Specific attention has been devoted to citizen science-related aspects because of their relevance to the correct interpretation of the overall results. Unlike the former City SoundScape (which was based on noise data collected by citizens [63] and provided city administrators with a planning tool for sustainable urban mobility), Apollon is a multi-parameter platform for noise, air and UV-A/B pollution monitoring and integrate heterogeneous data sources (institutional, open data, social networks and crowdsourcing) to provide services to Communities of Interest (CoIs) and Public Administrations. It is developed to support collaborative, services on a large scale and involves schools and families in environmental monitoring, for the collection of trustworthy data. In this work three main research questions are investigated:

1. How to make urban environmental monitoring truly pervasive, so that sensing data from institutional monitoring stations can be complemented by crowdsensing.
2. How to validate the robustness of our citizen science approach against the presence of wrong measurements and outliers in collected datasets.
3. How to achieve satisfactory data quality (i.e., data biases must be known or estimated in advance and data accuracy must be acceptable) by properly training and involving citizens in the data collection phase.

With the aim of describing the citizen science approach developed in the project for involving people in data collection activities and providing good quality data, we structured the paper as follows: Section 2 describes the background of urban monitoring and the recent trends in citizen science and mobile crowd sensing. The Apollon initiative is presented in Section 3. Section 4 reports the guidelines and the methods developed for effective citizen science initiatives using MCS. Section 5 reports the achieved results and their significance. Conclusions and further developments are drawn in Section 6.

# 2. Background and related works

## 2.1. Urban environmental monitoring

Traditional urban pollution monitoring campaigns are performed using interconnected Air Quality Stations (AQS) and Noise Monitoring stations. These correspond to normative monitoring systems and are used in the application of policy-making, environmental laws enforcement and health and disease studies, among other applications. However, a relevant problem with these monitoring systems is the costs associated with the instruments involved, over $200,000 USD per unit for an AQS and some tens of thousands USD per unit for noise stations. Therefore, the number of monitoring stations is limited, especially in large cities [6], where equipment requires qualified professionals to perform maintenance and since it is located at passer-by level, can easily be damaged, tampered with, or subjected to robbery or vandalism. Conversely, environmental conditions and urban morphology are relevant aspects affecting the distribution of both air pollutants and acoustic waves, making it hard to model an entire time space distribution at microscale. This fact generates the necessity of improving widespread monitoring networks. The advent of low-cost technologies has allowed the development of different applications over the recent years. Low-cost sensors still have questionable quality because of their low accuracy and unstable response, which is the reason for the initial skepticism of the official metrology. In recent years reports from the European Environmental Agency (EEA) [15] and the World Meteorological Organization (WMO) [4] have declared that low-cost sensors are not currently a substitute for reference instruments, but they have pointed out that large low-cost network implementation, combined with statistical analysis and machine learning, could complement the quality of the current official data and provide new pathways to obtain accurate, real-time information about the temporal and spatial variability of pollutants. Therefore, citizen engagement should be encouraged and decision-making should be empowered through data-driven insight. In order to improve data quality, iterative calibration treatment based on numerical simulation like in [5],[13] is recommended, because although some devices are relatively reliable, low-costs sensors can, for example, be sensitive to weather conditions or lack the capacity to measure very high or very low pollutant concentrations. This is the case with Airbeam (HabitatMap), a pollution monitoring device used for collaborative mapping in New York, which is very sensitive to wind, as stated in [36]. The US Environmental Protection Agency suggests a range of applications in order of significance based on the quality of the acquired data by different devices [58], as summarized in Table 1:

- **Education and information:** this kind of application exploits the sensors as teaching tools. In this case, measurements have to be considered in a qualitative way, to make relative comparisons between measurements taken in different conditions (different places, different periods, etc.); in fact, the uncertainty that affects the measurements does not allow for retrieving reliable data and carries a 50 % accuracy error.

- **Identification and characterization of hotspots:** this kind of application relies on fixed stations and/or mobile sensors to map and establish pollution sources. The sensors will determine where the concentration of pollutants is very high, in order to find the source. In this case, the precision and bias error can reach +/-30%.

- **Supplementary monitoring network:** this kind of application is used to support the official monitoring network. Additional low-cost sensors are positioned in the territory of the official network in order to map uncovered areas and supply useful measurements, complementary to the official ones. Of course, these sensors do not guarantee the same quality as those from official stations, but they are still useful to recognize potential pollution sources. In this case, the precision and bias error should be +/-30 -50%.

- **Monitoring of personal exposure:** this kind of application is used to analyze the effects of air pollution on people's health; hence, it is more focused on its consequences, rather than on pollution itself. The EEA measures the Air Quality Index, which represents the personal exposure to air pollution. Currently it is based on hourly collection of data coming from the official monitoring network, with the result that it is very often not available or not calculated. In this case, the precision and bias error should be +/-30% or better.

- **Official legal monitoring:** this kind of application is regulated by national rules, therefore it is the one that guarantees maximum quality and formality. It is, indeed, the application that, more than the others previously described, can reliably state whether a specific area is safe or not, according to thresholds values. In Italy, official monitoring is performed by the Regional Environmental Protection Agencies (in Italian, abbreviated as ARPA), through their fixed and mobile stations, scattered over the Regions and cities. ARPA monitoring stations can be classified as traffic, urban industrial or rural industrial. Official monitoring follows very stringent standards, so that the measured values can be considered correct and reliable. For this reason, the precision and bias error must not exceed +/-15%.

A specific line of scientific research is related to the development of low-cost sensors for air quality monitoring ([43, 6, 58]), as well as several contributions related to the benchmark of smartphone microphones that are usually involved in outdoor noise monitoring ([27, 26, 53]). Currently, a number of projects use low-cost sensors for monitoring noise and air quality. For In noise monitoring several apps are available for measuring sound levels for personal use, including the Advanced Decibel Meter, Sound Meter Pro or Decibel Meter Pro. Only a few research works address noise mapping, including the "Ear-Phone" project [44], where Nokia phones were used to predict sound levels in a given environment, "NoiseSPY" [25], which exploited smartphones carried by bicycle couriers to collect data in Cambridge, the "2Loud?" project [29] that uses iPhones to assess nocturnal noise within buildings near highways in Australia and Sonyc [1], which measures sound levels and visualizes sound maps in New York. One of the main limitations of such projects is that users are only involved as data collectors but no specific platform functionalities are tailored for them or for city managers for improving citizens' life quality. Therefore, if specific software solutions for noise mapping within urban contexts are needed, city managers have to consider professional systems and platforms, such as the software application suite developed by SoftNoise (SoftNoise GmbH), which provides a complete tool-set of products for environmental noise calculations ("Predictor - LimA") and mapping ("MapAtWork"), as well as for occupational noise mapping ("NoiseAtWork"). SoundPlan Acoustics (SoundPlan GmbH) represents a similar solution: it is a noise modelling software for technicians and professionals, which offers advanced noise mapping and animation functionalities for 3D scenaeries. The obvious drawback of such products is represented by their high cost and the necessity of skilled personnel capable of managing them properly. Consequently, city administrations typically cannot adopt them on a large scale. In cited project there is no evidence about the approaches to improve the trustworthiness of data collected by users. City SoundScape project[31] has been one of the first examples of how citizens would be systematically involved in participatory activities in the collection of data for supporting the smart planning of urban mobility. The same project has applied techniques of context awareness for improving the data provenance. Currently, even if noise is the second pollutant in our cities, it is still largely neglected in public opinion and a long path is required for both improving the quality and volume of collected data and to create the consciousness of its impacts on health [40]. For the monitoring of air quality using low-cost sensors networks, more projects have been developed over time, often related to the impact on health ([28], [7], [38]), even pushing the use of citizen science approaches to systematically involve citizens, as reported in [15]. The following section is specifically focused on this topic and on how it is evolved with the advent of digital technologies.

## 2.2. Citizen Science: A Fast-Growing Trend

Citizen Science has started several decades ago [19] as a collaboration of scientists and non-professionals in ongoing research projects to provide raw measurements, especially in the field of weather monitoring [49]. A more structured approach then followed, known as *collaboration* [19], in which citizens participate to the different phases of a scientific study (i.e., project design, data collection, data analysis, result dissemination), under the guidance or in collaboration with scientists and/or research institutions [48]. A

**Table 1**
Recap of air pollution main applications. Modified from [58].

| Tiers | Application area | Pollutants | Uncertainty | Data completeness |
|---|---|---|---|---|
| I | Education and information | All | $\leq 50\%$ | $\geq 50\%$ |
| II | Hotspot identification and characterization | All | $\leq 30\%$ | $\geq 75\%$ |
| III | Supplemental monitoring | Criteria Pollutants, Air Toxics (incl. VOCs) | $\leq 20\%$ | $\geq 80\%$ |
| IV | Personal exposure | All | $\leq 30\%$ | $\geq 80\%$ |
| V | Regulatory monitoring | $O_3$<br>$CO, SO_2$<br>$NO_2$<br>$PM_{2.5}, PM_{10}$ | $< 7\%$<br>$< 10\%$<br>$< 15\%$<br>$< 10\%$ | $\geq 75\%$<br>$\geq 75\%$<br>$\geq 75\%$<br>$\geq 75\%$ |

further step forward is now in place: non-professionals not only volunteer in collaboration with scientists but also define and perform scientific analyses on their own, thus behaving as "civic scientists". This category of engagement is defined as *co-creation* [19] and it is gaining progressive significance worldwide. The trend is so significant that a dedicated Web portal for finding, joining and contributing to citizen science projects worldwide has been founded by the American National Science Foundation and by the Arizona State University's Center for Engagement and Training in Science and Society. The portal, under the name of SciStarter [47], currently showcases "*more than 1600 formal and informal research projects, events and tools*" [47].

Several factors have enabled these advancements in citizen science: ICT technologies (e.g., GPS positioning; improvement in Bluetooth Low Energy, BLE, and wireless connectivity, cloud computing, data crowdsourcing, big data capabilities, visual analytics, social networks, etc.) and the increasing portability and pervasiveness of electronic devices (e.g., smartphones, tablets, wearable devices, etc.); the growing interest of citizens in local environment and its impact on their quality of life; the do-it-by-yourself approach in people, which is often motivated by an increased technical familiarity with electronic devices, especially in youngsters, due to their constant use of smartphones (e.g., video chats, multimedia streaming, online gaming, mobile Web browsing, etc.). The majority of these features push towards the incremental use of participatory approaches in citizen science, even adopting the Mobile Crowd Sensing Paradigm (MCS), as explained in the next Subsection. Two challenging aspects related to citizen science are the engagement potential and the citizen retention rate over time. Different approaches are used, to motivate citizens, as surveyed in [33], and analyzed in [45] and [46]. Of interest is the use of marketing approaches to improve retention as reported in [14] Moreover, often, even the usability of collaborative platforms that support citizen science projects impacts retention, as reported in [59]. The previously reported experiences are integrated in our approach, as described later.

### 2.3. Mobile Crowd Sensing (MCS) in Citizen Science

From the technical perspective, the involvement of people for enabling large-scale, pervasive and cost-effective physical monitoring is addressed thanks to the Mobile Crowd Sensing paradigm [20]. It exploits a variety of sensors, that are available as purpose-built units or are embedded in consumer electronic devices (e.g., wearables, smartphones, dashcams, etc.) and connected via USB or BLE/Wi-Fi to other systems (e.g., vehicular devices, Internet etc.). The MCS leverages the combination of mobile sensing and crowdsourcing in order to complement, widen and enrich the scope of traditional monitoring solutions.

The way users are involved in MCS activities can be twofold [18]. In *opportunistic* MCS¸ users grant to a dedicated mobile app the permission to collect data and to contribute to the sensing process without being directly involved (e.g., the user does not actively decide when performing a sensing operation), so that the entire process is managed opportunistically by the internal logic of the mobile app itself. In *participatory* MCS, on the other hand, users are consciously engaged so that they can actively decide when and how to trigger the sensing operation by using a dedicated mobile app. In both the modalities, several features of mobile devices can be exploited, ranging from reliable communication capabilities (both long- and short-ranged) to significant computational power.

Application areas are classified according to monitoring typology: road and traffic management are among the very first investigated domains and still attract interest [39]. Emergency management (e.g., earthquakes, landslides, tornadoes, floods, etc.), disaster relief and first-response coordination are also areas of application both for indoor and outdoor scenarios [35]. Also for environmental monitoring, several pollution typologies have been investigated: air [16], water [46], noise [62, 44] and electromagnetic fields [30].

We believe that participatory MCS fits perfectly into the citizen science scenario, as it eases the way for citizens to partake in scientific research activities by collecting data, performing monitoring campaigns and being part of specific inquiry processes. Thanks to the pervasive distribution of mobile devices, observations coming from professional measurement stations on the ground can be complemented by purposely designed MCS activities. These assumptions are backed up by several MCS initiatives in the smart citizenship and citizen science domains [61].

### 2.4. Communities of Interest

Citizen Science initiatives can benefit from gathering participants into *Communities of Interest* (CoIs), with the aims of planning and joining specific interventions and meeting like-minded people[52]. These communities are groups of citizen and scientists that share common interests and engage in knowledge sharing, collaborative learning and scientific inquiry during their participation in citizen science activities [42]. Therefore, members of a CoI can be people living or working in the same place (e.g., inhabitants of the same city, students of the same school) that are interested in knowing the pollution levels around them and improving the quality of their living/working environment. Since the joining factor is represented by scientific inquiry in a given citizen science experience, CoIs are also closely associated with so-called *Communities of Practice* (CoPs). CoPs represent an even more widely adopted concept describing a "*group of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly*" [56, 55].

## 3. Apollon

### 3.1. The Project

The Apollon Project (which stands for "environmentAl POLLution aNalyzer") is a research effort granted by the Apulia Region (Italy) in 2018. It targets the development of a platform for urban environmental (i.e. air quality, noise and UV) monitoring and analysis. The project is heavily rooted in the integration of heterogeneous data from several sources (e.g, citizens-owned personal devices, city-managed monitoring stations, etc.), and has the following aims:

1. integrating low-cost sensors scattered across the territory to create large observation areas;
2. engaging citizens in environmental monitoring actions and/or campaigns;
3. involving city administrators in proper management and exploitation of collected data.

From a technological perspective, a core requirement of the Apollon project is to build a hybrid data layer capable of properly integrating data flows originated by IoT sensors, mobile devices, open data, historical data, and social media feeds. The overall architecture exploits semantic technologies and geolocalized data analysis utilities, which enable near-/real-time monitoring services for several types of city end-users. Collected data are processed, aggregated and validated in near-time before making them available to final users via proper visual dashboards hosted by the platform front end.

One of the featuring aspects is the coexistence of several types of sensors that are geographically scattered. These devices range from smartphones (which collect measurements thanks to their embedded sensors as well as to external pluggable sensors) to low-cost mobile sensing stations (e.g., sensing boards based on Arduino and hosting several sensors). Two different categories of sensors have been considered.

- **Fixed stations**: this group encompasses low-cost metering equipment deployed by city administrators (when / where needed) and under city maintenance and control. These stations can be deployed in a fixed location for long periods or, alternatively, can be deployed on vehicles provided by city authorities and city responders (e.g., metropolitan police cars, traffic police cars, etc.) so that their sensing devices can be moved around the city without additional costs. It is worth pointing out that this category does not include the institutional monitoring stations used by public administrations for environmental control, as it collects data from low-cost stations only.

- **Mobile Crowd Sensing**: this category refers to data sources whose behavior complies with MCS requirements. Therefore, any citizen owning a mobile device can participate in monitoring campaigns and contributing her/his measurement data to the platform, provided that she/he installs a dedicated mobile app. Such a category exploits the self-scalability and dynamic infrastructures of edge/cloud computing

During the project design phase, different quantities have been estimated for the items from the two categories enlisted above. Indeed, while it is plausible to assume that few fixed stations would be installed within a city area, providing continuous data flows, the number of mobile devices equipped with a dedicated mobile app for collecting sensor data can scale considerably and they can be used opportunistically according to specific contexts. This is the reason why we can state that Apollon is configured as a full MCS platform, which covers all the aspects of the taxonomy described in [3]. In fact from the sensing scalability point of view, Apollon targets the three categories of separate sensing (when people gather data individually for personal use only), cluster sensing (through the definition of the CoI concept), and community sensing (because sensed data, properly anonymized, are exploited for Apollon services). At the citizen involvement level Apollon allows opportunistic, participatory, and hybrid sensing (a mixture of user control and device processing) and data are sensed either continuously or are executed depending on a specific context (certain time periods or places). Regarding the network infrastructure, Apollon can exploit existing infrastructures (e.g., access points and GSM), ad-hoc infrastructures and a hybrid ones.

### 3.2. The Architecture

As represented in Figure 1, the architecture of the Apollon platform includes four main layers:

- **Edge and IoT Layer**: This layer includes heterogeneous devices adopted as data sources. The project manages stationary sensors (e.g., monitoring stations provided by authorities and/or environmental protection agencies for pollution control) as well as mobile sensors (e.g., monitoring stations placed on top of vehicles routed across urban areas for mobile pollution

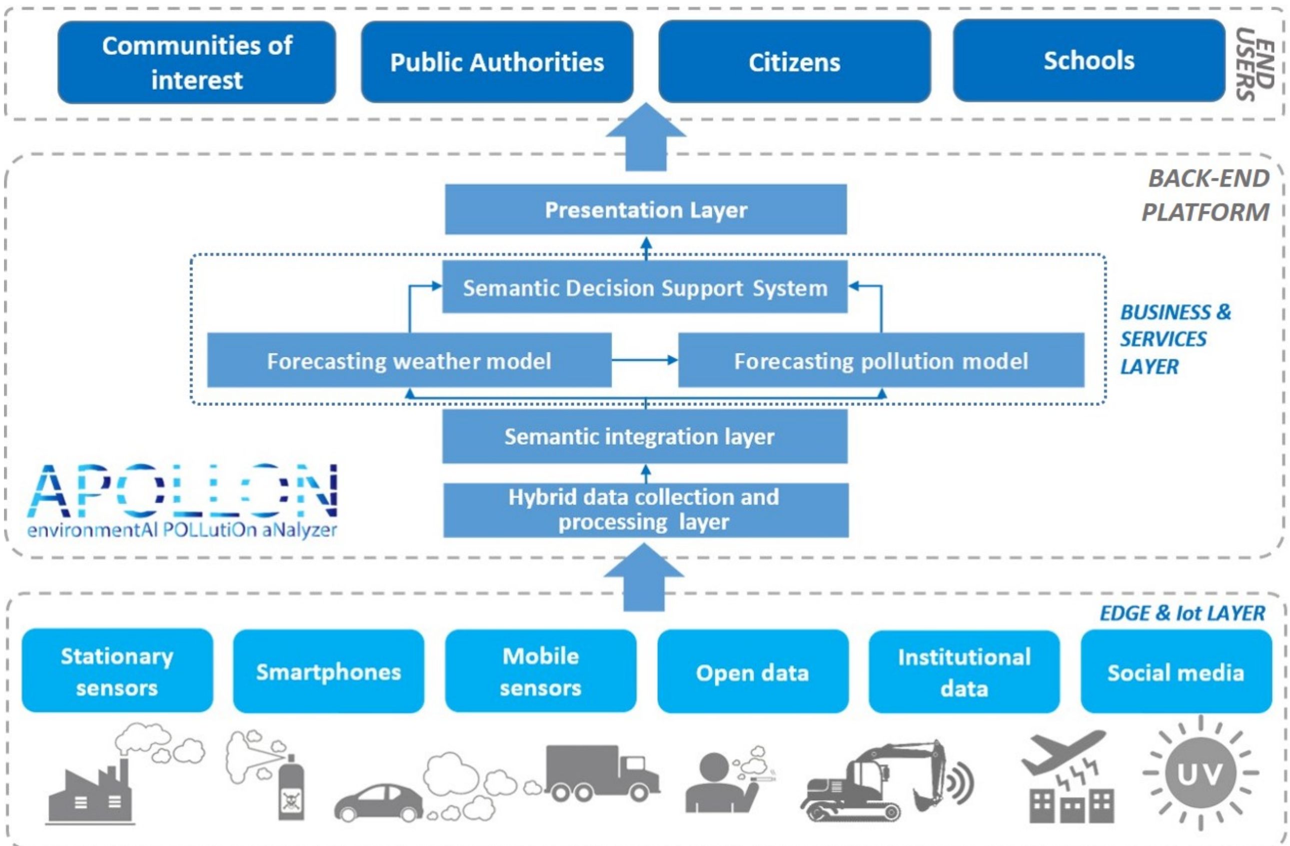


**Figure 1:** Apollon platform: architecture in the large.

control). A further type of mobile sensor is represented by personal electronic devices, such as smartphones and tablets, operating according to the MCS principles and collecting data via their embedded sensors or via external sensors. Additional data sources such as open data and institutional data repositories as well as social media streams belong to this layer, too. In the framework of the Apollon project, several physical parameters are monitored: noise levels, particulate matter (i.e., PM10, PM2.5, PM1), volatile organic compounds (VOCs), and UV-A/B rays.

- **Hybrid Data Collection and Processing Layer**: this layer is devoted to data storage, management, filtering and integration operations performed on the data received from the sources enlisted above. This layer also performs geo-referencing and time-stamping processes in order to reference measurement data in time and space properly. It communicates with the Business and Services Layer thanks to a dedicated Semantic Layer.

- **Business and services Layer**: This layer is in charge of performing complex operations on data coming from the data layer in order to feed services hosted in the Service sublayer. Business - to - Business (B2B) and Business-to-Consumer (B2C) services are exposed. Several specific modules are present in this layer, including advanced geo-referencing, sentiment analysis, user management, semantic analysis and integration and open data creation.

- **Presentation Layer**: This represents the multi-faceted interface used by project stakeholders to access the platform. The layer offers visual dashboards thanks to a dedicated Web portal, a mobile app and a Telegram BOT. The contents accessible via the presentation layer vary depending on the specific stakeholder type that request them.

As depicted at the top of Figure 1, several end-user types have been considered. They range from communities of interest (e.g., citizens groups and associations interested in performing pollution measurement, and factories and industries interested in their level of environmental pollution) to public authorities, and from citizens (both single individuals and associations) to other enduser categories such as schools or healthcare providers. This paper analyses the MCS component, for the aspects related to its effectiveness in citizen science initiatives. Other details about the MCS architecture are discussed in [32], where the osmotic computing paradigm is applied to move resources and processing towards the edge of the Internet.

### 3.3. Mobile App and Sensors

In the framework of the Apollon project, a mobile app has been designed and developed allowing users to participate in monitoring campaigns by measuring noise levels, particulate matter levels and other pollutants. Noise levels are sensed via internal or external, pluggable microphones. Particulate matter levels are measured by means of commercial devices for indoor/outdoor air quality control.

Four screenshots from the Apollon mobile app (Android version), which is freely available on the Google Playstore, are depicted in Figure 2. Screenshot (A) is the opening app screen: from this panel, the user can start to monitor noise levels or particulate matter levels. Once the monitoring functionality has been selected, the user can decide whether to start an automatic measurement session (i.e., the mobile app works in the background and sends measurement values each 60 seconds to the Apollon backend) or a manual session (i.e., the user selects in advance the duration of the measurement session). These two working modes are representative of the two above-mentioned MCS categories: opportunistic sensing and participatory. Screenshot (B) shows how ongoing noise measurements are shown to the user. The app monitors the sound pressure level (SPL, an instantaneous quantity) and the equivalent continuous sound level (LEQ, which is the time-averaged SPL on a given time window). Both these quantities are shown in the numerical dashboard at the upper part of the screen and are also charted in the area graph below. Three interaction buttons (i.e., "start measurement", "stop measurement" and "show measurement suggestions", depicted with a microphone icon, a square icon and a question mark icon, respectively) complete the interface. Any measurement exceeding regulatory thresholds is communicated graphically to the user directly in the chart. Moreover, a detailed summary of the performed measurements, as depicted in screenshot (C), is provided as a series of gauges that indicate how the collected values are placed with respect to limit values in different acoustic classes, according to the Italian regulation. Finally, the user receives suggestions on how to use the app and the smartphone microphone properly, by means of the schematized measurement diagram shown in screenshot (D). Measurement suggestions, live charts and gauges activate the plan-execute-monitor cycle upon which a citizen science initiative should be grounded.

However, it is important to point out that, while the user, especially when involved in a participatory sensing activity, receives these suggestions from the app itself, thus directly intervening in the measurement quality, the same does not apply to the opportunistic sensing, which could produce large amount of data from untrusted or unreliable sources.

These aspects will be tackled by adopting a series of guidelines to improve the effectiveness of our approach (see Section 4.1) and by implementing proper data processing solutions capable of increasing data trustworthiness and reliability (see Section 5).

### 3.4. Applications of the Platform

The platform described so far is the third iteration of a five-year-long effort of the authors to promote pollution control experiences with an approach based on citizen science and MCS.

The very first implementation of the platform started as an academic laboratorial experience in MCS applied to electromagnetic monitoring [30]. Even if the research was mainly aimed at examining collaborative learning applicability and effectiveness, it allowed us to start addressing typical challenges of MCS-based environmental monitoring.

The second implementation, funded through the European entrepreneurial accelerator FrontierCities and named City Soundscape, was released in 2015 [61, 31]. It targeted noise pollution monitoring in outdoor urban environment.

The third implementation, named Apollon and described in Section 3, was released in July 2019: it superseded and extended City Soundscape by adding air pollution monitoring functionalities (i.e., fine particles PM1, PM2.5, PM10, NOx, SOx and other primary and secondary air pollutants) and a more flexible and scalable structure in order to cope with communities of interest and more end-user services. Moreover an alert service was included in the mobile app to notify users when they are in areas where PM10 and PM2.5 exceed acceptable values.

The City Soundscape web portal is still available online[2] and hosts the entire dataset collected so far, ranging from the initial City Soundscape experience to the latest measurement campaign in the framework of the Apollon project. The new Apollon Web portal has been published online[3] as well.

From a technical viewpoint, City Soundscape and Apollon have been developed in collaboration with industrial partners which have adopted their internal methodologies to test the developed platforms and to guarantee the performance level required for the purpose. In particular, Apollon has been designed and developed to scale up to hundreds of thousands of users spread all over the Italian regions.

## 4. Guidelines and Method

One of the contributions of this paper is related to a set of guidelines and tools designed to effectively apply citizen science to urban pollution monitoring using MCS. In this section we report the guidelines adopted to involve schools, students and teachers in our citizen science project according to the relevant literature previously mentioned. Moreover, according to the lessons learned during the first two implementations, as mentioned in section 3.4, we describe the method developed to effectively use the platform to collect and analyze data and to monitor the pollution levels.

### 4.1. Guidelines for an effetive citizen science approach in Apollon

The scientific community has shown two major streams of reactions to citizen science so far. On the one hand, scien-

[2] http://Apollon.x-scape.it:10080/
[3] http://web.apollon-project.it/

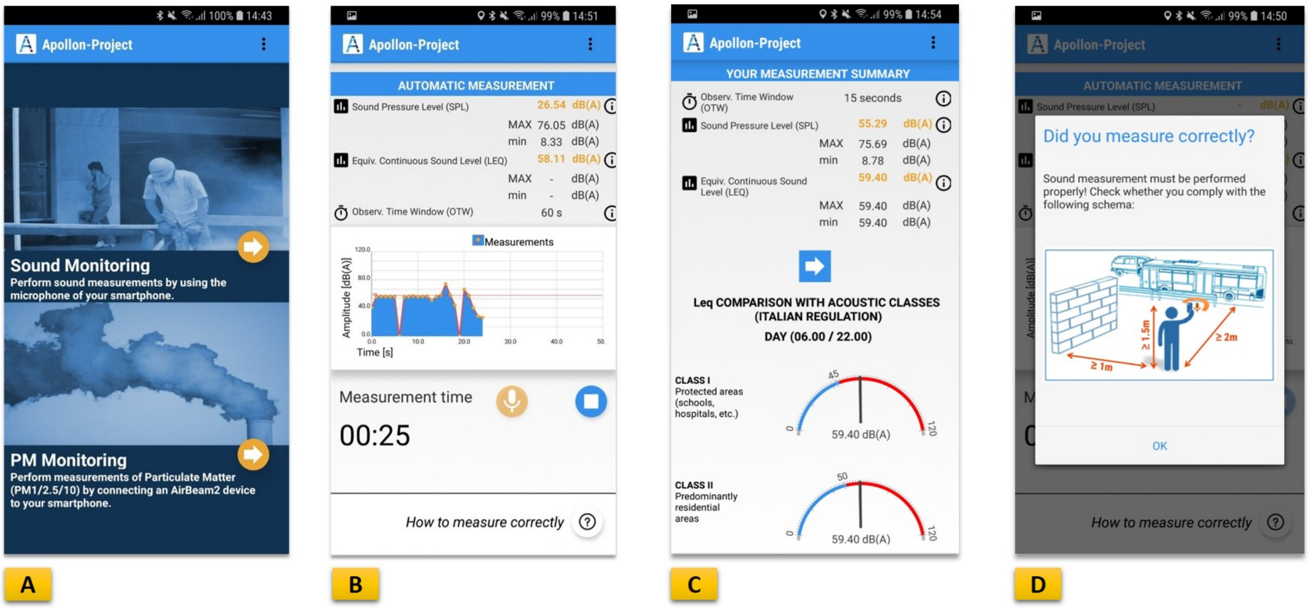


**Figure 2:** Screenshots from the project Apollon mobile app: opening screen (A); noise measurement dashboard (B); noise measurement summary (C); correct measurement best practices (D).

tists acknowledge several advantages, such as reduced budget limitations, cost reductions, expanded data collection campaigns and widened scientific investigation scopes [51]. On the other hand, the involvement of non-professionals in research activities is still a cause of skepticism and concerns about the challenges that citizen science poses regarding traditional research ethics, the vulnerability of specific classes of participants, the accuracy of collected data and the relevance of scientific findings [12]. In recent years, a consistent stream of scientific literature has addressed these problems, which we restate below as challenges (indicated as $C_n$), followed by the best mitigation strategies and/or guidelines (indicated as $G_n$) we have adapted:

$C_1$: the combination of citizen science and smartphone usage face several issues related to data quality, which is predominantly affected by measurement accuracy [41], is conditioned by measurement task difficulty, precision of the measurement tool and the level of experience of the involved non-professionals [50]. The result can be the presence of outliers (i.e. data points differing significantly from the dataset they belong to), which must be identified and whose cause must be assessed whether deriving from anomalies in measurement procedures or tools rather than by actual considerable variations in the monitored physical quantity.

$G_1$: to properly deal with data accuracy and data trustworthiness [2], a training strategy has been adopted (see Section 4.2) to prepare the involved citizens and ensure sufficient reliability of collected data. Moreover, a formal procedure has been developed (see Section 4.4) to validate the collected dataset and avoid excessive variability, misleading inferences and biased data (i.e. data affected by systematic errors [41]).

$C_2$: Citizen involvement is a critical challenge as MCS activities very often evolve from initial success to progressive reduction of participants, due to several reasons (e.g., repetitiveness of tasks, lack of incentives, scarce sharing and social opportunities, excessive technical and/or scientific contents, etc.). Involvement implies different aspects including how to engage and retain each participant and how to scale up the number of participants.

$G_2$: citizen science has proven to be more effective if supervised by professional researchers, who can keep citizens engaged by providing support and enrich the citizen scientists experience with updated news and information via focused groups on social networks and chats. Moreover retention can be improved using remuneration mechanisms according to the specific use case as reported in [33, 14]. To scale up the number of participants, pivots are identified who have acted with a multiplier effects on the number of participants. In our methods to face the challenge of citizens involvement we have adopted an educational-oriented approach by starting the citizen science projects from within education institutions, in order to involve students from middle and high schools first. We have directly trained a number of students to validate the training toolkit and the Apollon platform usability. As pivotal elements we have selected low and high secondary school teachers, who have been trained both on the inclusion of mobile apps in science lessons and on the use of the tools which become means for experimental collaborative activities. In our experience the inclusion of mobile apps has increased the interest of students in scientific education and provided teachers with efficient tools for implementing laboratory activities even in the case of scarce laboratory resources [63]. In order to trigger schools participation proper communication must be designed and implemented using official channels and the dedicated social networks.

$C_3$: Privacy and citizens' data protection is another relevant concern: proper encryption, anonymization and data perturbation solutions have to be implemented to preserve data privacy and protect against misuse [9].

$G_3$: The use of a supervised approach via professional researchers enables definition of the policies concerning informed consent, data confidentiality, security, data publication and sharing and users privacy in advance, and provide specific answers to citizens' questions on the most significant ethical, legal and social implications [17]. Moreover the Apollon platform has developed a novel data protection service via the osmotic membrane which filters data according to specific policies defined by citizens and/or CoIs [32].

In its first international conference (Berlin, 2016) the European Citizen Science Association (ECSA) defined a set of recommendations addressing science, policy and society, in order to foster the further development of citizen science [23]. These guidelines include: demonstrating scientific benefits of citizen science to the public, branching out across multiple disciplines, embracing opportunities for policy monitoring and development, etc. From the ECSA recommendations and other studies ([15, 3] we extracted a subset of guidelines which have been adopted in the Apollon project and which are instantiated in the following subsections, namely:

1. provisioning of a kit including specific items needed to ensure a high-quality citizen science experience,
2. iterative development of task and tool design through a co-design approach,
3. use of standardized and calibrated equipment,
4. expert validation of performed measurements,
5. experiment replication and equipment calibration, (to be performed amongst volunteers),
6. skill-based statistical weighting of volunteer activity outcomes,
7. accounting for random errors and systematic bias.

Our citizen science solution, therefore, has been devised keeping in mind the key elements discussed so far. We started with an education oriented approach, which later also involved associations and groups of interests sensitive to urban pollution

### 4.2. Volunteer Training, Assessment and Enrollment

To achieve satisfactory data quality (i.e., data biases must be known in advance or at least estimable and data accuracy must be acceptable), as specified in the previous section 4.1, volunteers have been trained properly before participating in measurement campaigns. In Apollon this strategy has been preferred to the enrollment of untrained volunteers as it improves significantly the overall quality of the citizen science experience as well as that of the measured data. In addition, by referring to the concept of CoIs as described in Section 2.4 and by considering the outcomes of the first two implementations described above, we have implemented CoIs in Apollon, in order to raise both the engagement level of the participants and the systematic pervasiveness of monitoring the spatial-temporal domain.

In fact in Apollon participants are clustered into four classes of users depending on their numerical amount and on their anonymity in the platform (Figure 3):

- **Anonymous users**: untrained volunteers who do not belong to any specific CoI and do not login to the platform. They simply provide monitoring of data via the mobile app. However, they are allowed to check overall pollution levels thanks to the platform Web portal.
- **Profiled users**: trained volunteers (who participated in specific training sessions supervised by domain experts and by Apollon platform managers) who can perform individual measurements via the Apollon mobile app and enrich those measurements with comments.
- **Anonymous CoIs**: trained volunteers enrolled in specific CoIs who can perform measurement campaigns but who cannot enrich measurements with any additional comments.
- **Profiled CoIs**: trained volunteers enrolled in specific CoIs. These volunteers can perform measurement campaigns and enrich them with comments. In addition, they are engaged in co-designing monitoring campaigns, as in the case of schools, which can define specific monitoring campaigns by accessing a dedicated area.

During the project volunteers who attended the training sessions were briefed on the theoretical aspects and fundamentals of Mobile Crowd Sensing and mobile devices in general. Then, they were trained through practical sessions on pollution monitoring, how to perform group-reporting in experiments and how to analyze and interpret the achieved results.

### 4.3. Equipment Provisioning

The provisioning of a toolkit for volunteers improves the level of engagement in a supervised citizen science activity. The toolkit must contain everything necessary to perform the experiment, collect the results, engage the participants and ensure the reliability of the process.

In Apollon the toolkit is made up of:

- all services offered by the Apollon platform to gather data coming from volunteers' devices and depiction of the pollution maps produced thanks to their effort. The platform shows aggregated, geo-referenced and

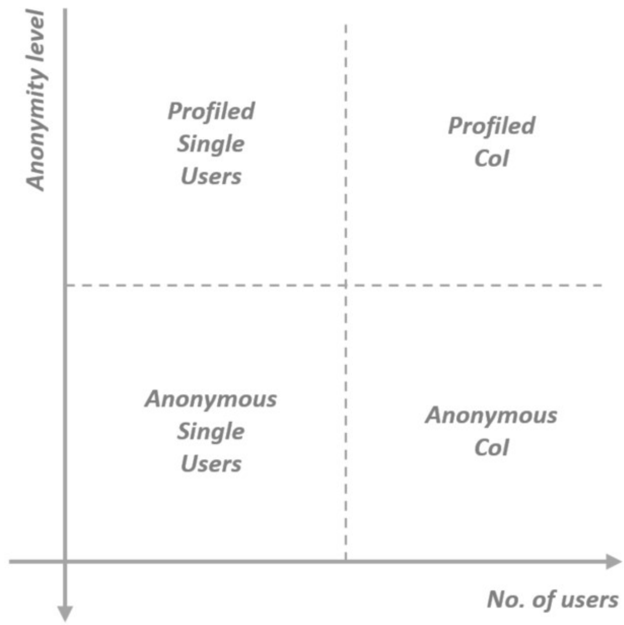


**Figure 3:** Clusters of users in Apollon.

time-located measurements and is capable of managing user profiles as well as CoIs, as described before. Similarly, it allows for the co-design of monitoring campaigns as CoI members can log in to the platform and customize their monitoring campaigns (e.g., by choosing location and duration).

- the mobile app, which is necessary to perform measurement sessions. As for the experimentation, the app is also complemented by ad-hoc Bluetooth/Wi-Fi connection capabilities, to connect the hosting mobile phone to other external sensors (i.e., portable sensors for detecting air pollutants) as described below.

- one or two external sensors capable of complementing the native sensors embedded in the mobile phone. While noise metering can be performed via mobiles, air pollution control requires participants to be provided with sensors they can carry with them during monitoring campaigns as portable devices paired with their smartphones. These devices enrich data streams coming from fixed stations (see Section 3).

Along with sensors, iterative calibration is required to tackle systematic biases in sensor readings. Since the Apollon project involves both acoustic and air pollution monitoring, two different calibration approaches have been adopted, which require two different types of calibration equipment:

- a class 2sound meter (e.g. Sauter SU 130) at least, and a procedure to calibrate the mobile phone of each participant, so that noise measurements can produce quantitative data and not only approximative indications,

- a procedure to maximize data quality for air monitoring. Due to the greater complexity of the involved equipment, calibration procedures for air pollution sensors have been performed before engaging students in monitoring activities. More specifically, portable sensors and stationary equipment have been calibrated by referring to institutional monitoring stations (whose data are available for the specific type of pollutants in the pilot sites). Metering biases have been identified so that correction factors have been applied to those sensors. In addition, portable sensors have been controlled and periodically re-calibrated according to the procedure detailed above for noise monitoring activities. Participants who were provided with sensors for air monitoring have been instructed on how to flag any issue in their devices and constant support has been provided to ensure service continuity and data accuracy checking.

### 4.4. Methods to Ensure Data Quality in Apollon

As anticipated in section 4.1, the Apollon platform exploits a data pre-processing algorithm to validate all datasets received from users. The algorithm estimates potentially wrong device positions and setups, so that corresponding measurements can be discarded or managed properly. We propose here a short summary of this process (i.e., Algorithm 1) and refer the interested reader to [31] for more details.

According to Algorithm 1, noise measurements ($m$, [dB]) are enriched by a set of contextual data: luminosity ($l$, [lux]), proximity ($p$, [cm]), and detected activity ($da$, as evaluated by built-in Android APIs on the basis of readings from the smartphone accelerometer and gyroscope). The reliability of noise measurements (namely, $R(m)$) is inferred from contextual parameters, when needed. Let us consider some examples. If noise readings exceed sensor metering boundaries (i.e., less than 20dB or more than 120dB), they can be considered as unreliable. If, conversely, a plausible noise metering is collected (e.g., between 50dB and 90 dB), the presence of any obstacle very close to the smartphone (indicated by low proximity values, i.e., less than 5cm) suggests low reliability, as the smartphone microphone could be obstructed. When the proximity sensor does not detect any obstacle (i.e., $p \geq$ 5cm), low reliable noise measurements can be spotted as well. If luminosity is low (i.e., $l \leq$ 50lux) but the timestamp indicates daytime and the device location (in terms of latitude and longitude) is not compatible with the timestamp (i.e., $s \perp t$), the algorithm marks the corresponding measurement as low reliable.

In addition, starting from Algorithm 1, a transductive machine learning approach has been adopted to leverage contextual awareness: the algorithm assesses data trustworthiness by training a classifier that infers unknown sensor readings on the basis of reliable measurements [60]. A similar approach has been implemented to maximize data quality for the air monitoring component of the project.

When dealing with wireless sensor networks and mo-

**Algorithm 1** Reliability assessment of noise measurements

Given $m$ = noise level [dB]
at time $t$ [s], location $s$ [lat;long]
Given $l$ = luminosity [lux]
Given $p$ = proximity [cm]
Given $da$ = detected activity
Given $R(m)$ = measurement reliability
**Require:** $m > 0$
**for all** $m = 1, 2, \ldots, M$ **do**
  **if** $m \leq 20\text{dB} \vee m \geq 120\text{dB}$ **then**
    $R(m) \leftarrow$ unreliable
  **else if** $20\text{dB} < m \leq 50\text{dB}$ **then**
    **if** $p < 5\text{cm}$ **then**
      $R(m) \leftarrow$ poorly reliable
    **else**
      **if** $l < 100\text{lux} \wedge t = \text{daytime} \wedge s \perp t$ **then**
        $R(m) \leftarrow$ unreliable
      **else if** $100\text{lux} \leq l < 200\text{lux}$ **then**
        **if** $da \neq \text{STILL}$ **then**
          $R(m) \leftarrow$ poorly reliable
        **else**
          $R(m) \leftarrow$ reliable
        **end if**
      **else**
        $R(m) \leftarrow$ reliable
      **end if**
      $R(m) \leftarrow$ reliable
    **end if**
  **else if** $50\text{dB} < m \leq 90\text{dB}$ **then**
    **if** $p < 5\text{cm}$ **then**
      $R(m) \leftarrow$ poorly reliable
    **else**
      **if** $l < 50\text{lux} \wedge t = \text{daytime} \wedge s \perp t$ **then**
        $R(m) \leftarrow$ poorly reliable
      **else if** $50\text{lux} < l < 100\text{lux}$ **then**
        **if** $da \neq \text{STILL}$ **then**
          $R(m) \leftarrow$ poorly reliable
        **else**
          $R(m) \leftarrow$ reliable
        **end if**
      **else**
        $R(m) \leftarrow$ reliable
      **end if**
      $R(m) \leftarrow$ reliable
    **end if**
  **else**
    **if** $p > 5\text{cm} \wedge l > 50\text{lux}$ **then**
      $R(m) \leftarrow$ reliable
    **else**
      $R(m) \leftarrow$ poorly reliable
    **end if**
  **end ifreturn** $R(m)$
**end for**

bile networks in general, trust mechanisms must be considered, as sensors are typically deployed in open environments, where they can be attacked, disrupted or infected with malicious data [34]. Different types of trust can be envisaged. A *direct trust* is achievable by considering device-related features (e.g., sensor readings, time availability, energy consumption patterns, etc.), while an *indirect trust* is provided by trusted data coming from other (e.g., neighbouring) devices, so that the trusted node does not communicate directly with the data consumer but with intermediate nodes that have direct communication with it [22]. If direct and indirect trust are combined, a *hybrid trust* or *recommendation trust* [21] is obtained, which requires that trusted nodes provide recommendations about the nodes under observation.

In MCS, trust factors have been investigated mostly for the recruitment of trustable participants, especially when they provide subjective opinions in addition to raw sensor data [64] but trust factors are also inherently related to the sensing model adopted in a specific MCS scenario [54], since MCS can exploit direct sensing (i.e., data are exchanged directly amongst devices that are physically close) as well as indirect sensing (i.e., data are exchanged via a dedicated, centralized MCS platform) [8].

Due to this usage scenario, we used the direct trust, based on the knowledge of the trustee's context thanks to contextual information coming from the device (e.g., additional sensor readings, GPS location, timestamping, etc.). As for the indirect trust factor, we performed manual checks on specific data providers on the basis of readings coming from other data providers that were close in time and space. A full indirect trust factor management based on reputation and experience concepts (as explained in [54]) is under evaluation and it is expected to be implemented in the near future.

Moreover in Apollon we implemented a second algorithm (not reported here for sake of simplicity) to reduce the number of outliers and improve the overall data quality. The algorithm is based on a near real-time control procedure: during measurement campaigns, student groups are assigned in advance with specific time and location slots to be covered, thus enabling sample verification of data streams from their monitoring devices. A dedicated project tutor was in charge of checking whether incoming data had a correct geographical reference and whether adjacent (in place and in time) devices were providing homogeneous measurements.

With regard to the improvement of noise measurement data, it is worth to noting that, at the moment, no institutional noise monitoring stations are available across the Apulian regional territory (except for some dedicated sensors deployed along regional airport runways, whose datasets are not accessible), which prevents comparison of users' collected noise data against official data.

## 5. Results and Discussion

As of this writing, 12247 measurements have been collected worldwide since 2015, with an overall number of 500 students and teachers trained and involved in citizen science

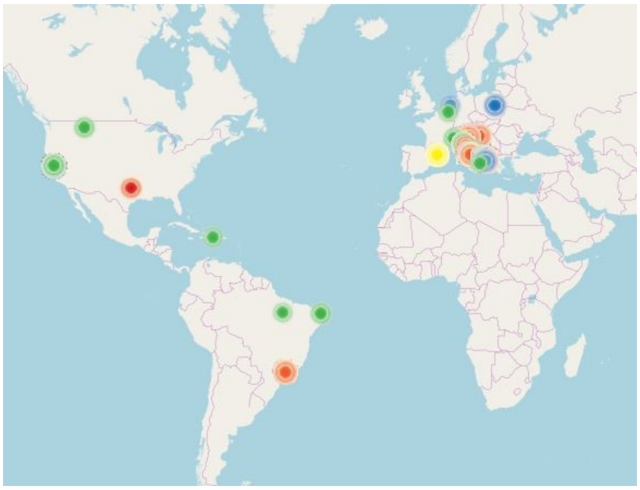

**Figure 4:** Spatial distribution of measurements collected worldwide.

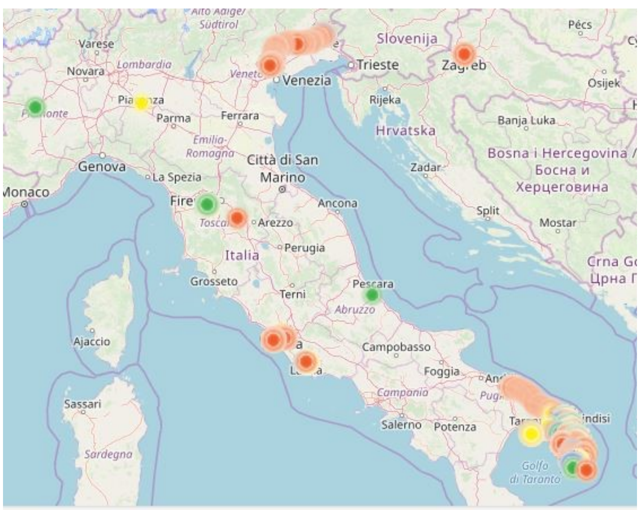


**Figure 5:** Spatial distribution of measurements collected on the national Italian territory.

activities. The overall spatial distribution of these measurements is presented in Figure 4, while Figure 5 details the spatial distribution across the national Italian territory only.

The first [30] and second run [61, 31] of our citizen science approach addressed the challenges of participant recruiting and training and of equipment calibration.

In this third implementation we have focused on scaling up the approach in terms of number of participants, and the amount and quality of collected measurements samples. The pilot projects have been run in the cities of Lecce (about 95.000 inhabitants), Brindisi (about 88.000 inhabitants) and Campi Salentina (nearly 10.000 inhabitants) which present very different urban pollution conditions: Lecce is mostly a tourism-based city, and pollution there is mainly due to road traffic and nightlife, Brindisi is an industrial city with a port, airport and the main provincial road that crosses its residential zones. Campi Salentina is a small artisan town, which was includedin the project to evaluate its scaling potential and the reactions of citizens to this kind of activities. The experimental phase started in January 2019 and lasted until March 2020. During the first six months, nearly 30 high-school students and about 25 high-school teachers have been involved. Students tested the supervised acoustic monitoring scenario only, and co-designed the training resources, which have been then used to train other students and pivotal users from associations and interest groups in the cities. Similarly, teachers were introduced to the use of mobile apps for acoustics and to the use of IoT devices for air pollution monitoring as tools for scientific experiments. Moreover, they were trained on citizen science and challenged to co-design laboratory experiments using the Apollon platform. These experiments have conducted with their students starting from September 2019. If we estimate that each class has 25 students on average, about 650 high-school students have been indirectly involved in the project, coached by their teachers. Noise measurements were collected via the Apollon project back-end and visualized on dashboards showcased by the project Web portal. Figure 6 depicts the dashboard for a city manager, once he/she has logged in. The city map is shown in the central area of the dashboard. Several informational layers are superimposed on the map including POIs (i.e., schools, hospitals and sport facilities, all represented as green squares), territorial zoning with respect to Italian acoustic classes (represented as polygons, according to the orange color scale in the top left corner of the map), and measurement points (represented as hexagons, according to the measurement color scale at the top left corner of the map). Filtering options are available at the top right corner of the dashboard and details about the selected city area (i.e., polygons) are enlisted at the right side of the dashboard. The overall measurement summary and regulatory thresholds are reported for the selected city area, along with two charts of measurement time series. The first one shows collected data in a three-month window, while the second one presents the current day's measurements with an hourly granularity.

### 5.1. Volunteer Training and Engagement

As discussed in section 4, in Apollon we introduced CoIs for raising both the engagement level of the participants and the pervasiveness of monitoring the spatial-temporal domain.

During the second experimental phase, which started in June 2019, 30 more high school students were trained on the Apollon framework for monitoring both air quality and noise in Lecce. Students took a 12 hours class on theoretical aspects of Mobile Crowd Sensing and mobile devices in general. They were then trained in a 20-hour practical session to develop the necessary skills for pollution measurement. In more detail, during dedicated hands-on sessions, students learned the internal structure of a smartphone and its sensing equipment, as well as its functionalities and how to use such devices in order to measure physical quantities profitably. Then, students attended short frontal and interactive lessons dealing with foundations of acoustics, foundations of metrology and measurement techniques. Some sample

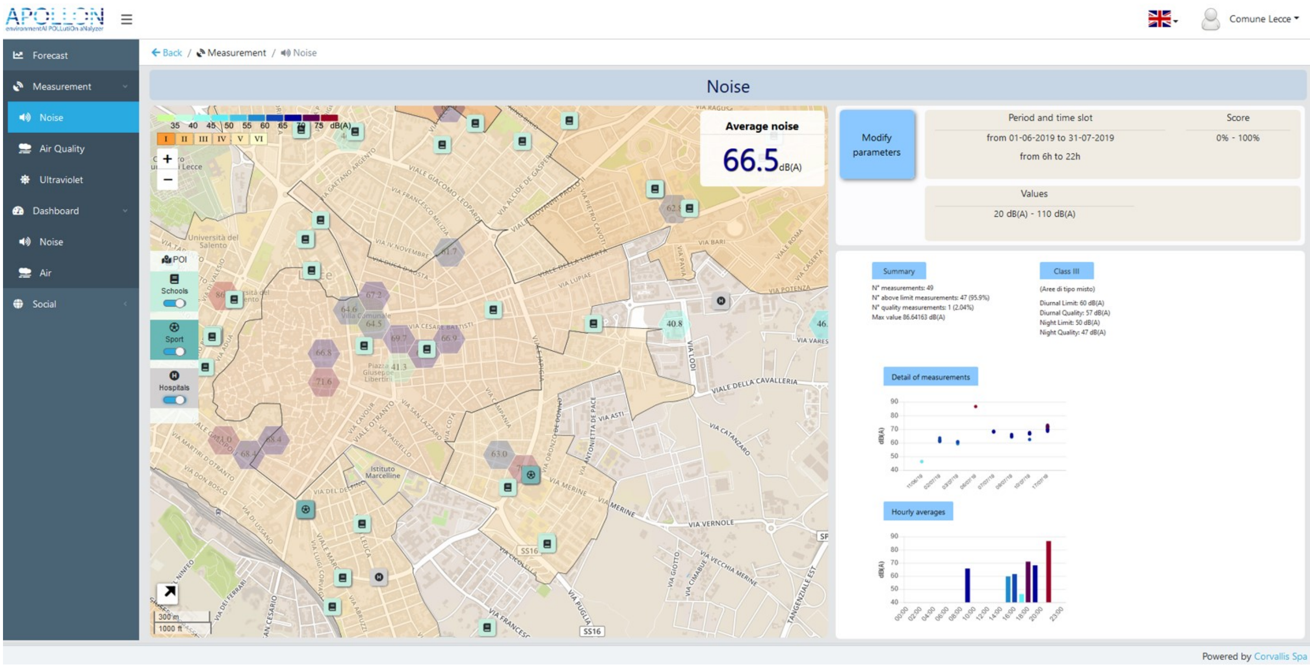


**Figure 6:** Apollon dashboard.

measurement scenarios were provided, in order to prompt them to co-design their activities. During the co-design process, they devised measurement settings in accordance with the Apollon platform managers and domain experts. Finally, students were trained on how to perform group-reporting on the participated experiments and how to analyze and interpret the achieved results. During the second session only two students dropped out, while teachers asked to be further involved in these kind of activities, using the Apollon platform for performing more experimental sessions with their students.

As for city administrations, their involvement started in June 2019, by means of their Environmental Departments. They provided part of the requirements and co-designed the dashboards. Moreover, they recruited their respective users and stakeholders to participate in the monitoring activities. For the pilot phase with the Municipal Administrations, starting in January 2020, two more associations in Lecce, the local police department and eight more schools in Brindisi, three coffee shops in Campi Salentina were involved in monitoring sessions. The monitoring campaigns in cities were managed in order to cover as much as possible of the whole cities' areas for noise pollution and to complement the location of official stations for air monitoring effectively.

The application of the guidelines for volunteers recruitment and training allowed us to reach satisfactory data quality levels. Indeed, once properly trained, students were able to measure physical quantities correctly, thus avoiding an excessive number of outliers.

## 5.2. Data Quality

As described in the method adopted during noise monitoring trials (which involved smartphone microphones only) we asked all participants to calibrate their mobile phones. The results showed a considerable variation of the measurement bias, mainly depending on: device type, presence/absence of smartphone cover; type of smartphone cover and the presence of dust on the microphone. In Table 2 a subset of measurements from sample smartphones is compared against the measurements provided by the *Sauter SU 130* class 2 sound meter. The calibration was performed indoor, in a quiet room, by using a dedicated educational mobile app as a sound generator (more specifically, a sinusoidal wave at 493.88Hz, B note, was used) over a time span of five minutes. From the results it emerged that some devices exhibited a significant bias between their measurements and the ones provided by the sound level meter. Therefore, a second set of calibration measurements has been performed outdoors. The results, shown in Table 3, confirm that a specific subset of devices underestimates the Leq values, thus, a proper correction factor can be used when considering measurements flowing from those devices.

More specifically, the lower bound (i.e., the smallest bias, nearly ±1.5dB) of the biasing window was achieved with new smartphone models, recently bought by their owners, without any cover and with clean microphone openings. The upper bound (i.e., the highest bias, between –10dB and -12dB) was experienced with older smartphone models, equipped with leather flipbook covers and showing visible traces of dust obstructing their microphones. It is worth mentioning that even a new smartphone model has showed a –16dB bias compared to the sound level meter. The device was a rugged

**Table 2**
Excerpts from indoor calibration sessions. Setting: indoors, quiet room, a-priori known sound source (493.88Hz, sinusoidal waveform). The sound level meter sensed 66.7[dB] in this scenario.

| Smartphone brand/model | $L_{eq(DUC)}$ * [dB] | RCF† [dB] |
|---|---|---|
| Samsung Galaxy A50 | 68.1 | -1.4 |
| Huawei Mate 20 Lite | 55.9 | +10.8 |
| Samsung Galaxy J3 | 63.8 | +2.9 |
| Huawei P20 Lite | 55.1 | +11.6 |
| Samsung Galaxy J5 | 64.0 | +2.7 |
| Caterpillar CAT S31 | 50.4 | +16.3 |

* $L_{eq(DUC)}$: Continuous equivalent sound level of the Device Under Calibration (DUC).
† RCF: Required Correction Factor. This represents the additive term to be applied to the measurement provided by the DUC to match the difference with the measurement provided by the sound level meter used as reference device (Sauter SU 130).

**Table 3**
Excerpts from outdoor calibration sessions. Setting: outdoors, various urban roads with heavy traffic.

| Scenario | RCF† device 1* [dB] | RCF† device 2* [dB] |
|---|---|---|
| Road 1 | +12.9 | +2.8 |
| Road 2 | +10.6 | +2.9 |
| Road 3 | +11.4 | +2.3 |

* Device 1: Huawei P20 Lite.
** Device 2: Samsung Galaxy J3.
† For the RCF calculation, see footnote † for Table 2.

smartphone for construction workers, shielded against dust and water, thus limiting the microphone quality significantly.

Thanks to the integrated actions of training and tools, the methods to check the quality of student-contributed data have rated all the samples of good quality, none discarded.

Qualitative benefits achieved by the Apollon project include:

- The increased awareness of students involved in the project, as well as their families, about the direct effects of daily life activities on air quality and noise pollution. Almost all involved students reported positive feedback (e.g. with specific measurements performed while doing sport, cleaning home, sleeping, preparing foods etc.)

- The clear perception of the limited value of the current prescription and thresholds enforced by law concerning pollution, which are based on a very small number of sensors (low spatial resolution) and measures (low temporal resolution), and are unable to account for strong pollution differences as measured by participants.

- The possibility to integrate the standard monitoring approach, based on a small number of high-cost, high-quality facilities managed by Public Administrations, using the mobile crowd sensing approach, and based on a large number of inexpensive connected sensors managed by trained volunteers, so as to achieve a new, low-cost and highly flexible approach more suitable for urban environments.

In our experience, sensor calibration was so effective in performing sensor readings correctly that we strongly suggest any citizen science activity to plan for the availability of proper calibration equipment as a mandatory asset in its toolkit. In our opinion, for activities related to the noise monitoring scenario, a portable class-2 sound level meter is sufficient. Indeed, it offers reliable measurements at affordable costs (no more than 200€). A suitable adoption plan could suggest that each CoI acquires such a device. In order to improve the quality of measurements external microphones can be added to smartphones [26].

## 6. Conclusions

The advent of low cost sensors, the pervasiveness of IoT and mobile devices and the novel architectures for big data management have fostered the development of novel frontiers for urban pollution monitoring, involving people in MCS initiatives for education and creating informed opinions about urban pollution and sustainability. In this paper, the Apollon project and the applied citizen science approach for urban environmental monitoring are presented. According to the application scenarios in [58] the aim of Apollon can be classified at Tier I. This is the reason why we initially targeted schools. In Apollon data trust processes for improving data certainty and organized monitoring campaigns for widening the area coverage have promoted the application area to Tiers II, III and IV in [58] according to the users devices and objectives. Overall, the mobile crowdsensing approach is gaining momentum and several schools, as well as several municipalities, in addition to those involved in our projects, have asked to receive our training and to participate in environmental monitoring campaigns. The same positive feedback and requests have been received by CoIs, involved in environmental sustainability and social support initiatives. Moreover in our University program addressing sustainability and links with the territory, the whole student population (about 17k students) will be offered the chance to participate in our campaigns, which promises to act as a real boost for data gathering activities. The experiment can easily scale up because of the presence of a cloud system designed accordingly. This favorable scenario allows us to further tailor the system and the process at a larger scale, thanks to the adoption of guided methodology based on citizen science principles. The results achieved so far demonstrate that a supervised citizen science experience involving trained volunteers, schools and public administrations, organized into Communities of Interest, is able to increase the quality of data coming from citizens, increase public awareness on environmental pollution and induce positive changes in participants' mindsets and behaviors, as well as facilitate deeper analysis performed by the official environmental agencies in case of uncertain situations.

## 7. Acknowledgement

This work is partially supported by the research project "environmentAl POLLution aNalyzer (APOLLON)," within the "INNONETWORK 2017" call funded by Regione Puglia (Italy) in the framework of the "Fondo Europeo di Sviluppo Regionale (FESR)".

## References

[1] Bello, J.P., Silva, C., Nov, O., Dubois, R.L., Arora, A., Salamon, J., Mydlarz, C., Doraiswamy, H., 2019. SONYC. Communications of the ACM 62, 68–77. URL: https://dl.acm.org/doi/10.1145/3224204, doi:10.1145/3224204.

[2] Bijarbooneh, F., Flener, P., Ngai, E., Pearson, J., . Optimising quality of information in data collection for mobile sensor networks, in: Proc. 2013 IEEE/ACM 21st Int. Symp. Qual. Serv, p. 1–10.

[3] Boubiche, D.E., Imran, M., Maqsood, A., Shoaib, M., 2019. Mobile crowd sensing – Taxonomy, applications, challenges, and solutions. Computers in Human Behavior 101, 352–370. URL: https://doi.org/10.1016/j.chb.2018.10.028https://linkinghub.elsevier.com/retrieve/pii/S0747563218305181, doi:10.1016/j.chb.2018.10.028.

[4] Bruce-Keller, A.J., Salbaum, J.M., Luo, M., Blanchard, E., Taylor, C.M., Welsh, D.A., Berthoud, H.R., 2018. Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications. Technical Report 7. WMO - World Meteorological Organization. URL: http://www.wmo.int/pages/prog/arep/gaw/documents/Draft_low_cost_sensors.pdfhttps://linkinghub.elsevier.com/retrieve/pii/S0006322314005204.

[5] Budde, M., El Masri, R., Riedel, T., Beigl, M., 2013. Enabling Low-Cost Particulate Matter Measurement for Participatory Sensing Scenarios, in: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. pp. 1–10. URL: https://doi.org/10.1145/2541831.2541859, doi:10.1145/2541831.2541859.

[6] Caquilpán P., V., Aros G., G., Elgueta A., S., Díaz S., R., Sepúlveda K., G., Sierralta J., C., 2019. Advantages and challenges of the implementation of a low-cost particulate matter monitoring system as a decision-making tool. Environmental Monitoring and Assessment 191, 667. URL: http://link.springer.com/10.1007/s10661-019-7875-4, doi:10.1007/s10661-019-7875-4.

[7] Chen, M., Yang, J., Hu, L., Shamim Hossain, M., Muhammad, G., 2018. Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators. IEEE Communications Magazine 56, 14–20. doi:10.1109/MCOM.2018.1700571.

[8] Chen, P.Y., Cheng, S.M., Ting, P.S., Lien, C.W., Liu, F.J., 2015. When crowdsourcing meets mobile sensing: a social network perspective. IEEE Communications Magazine 53, 157–163.

[9] Christin, D., . Privacy in mobile participatory sensing: Current trends and future challenges. Journals Syst. Softw 116, 57–68. doi:10.1016/j.jss.2015.03.067.

[10] Conti, M., Passarella, A., . The internet of people: A human and data-centric paradigm for the next generation internet. Comput. Commun 131, 51–65. doi:10.1016/j.comcom.2018.07.034.

[11] Conti, M., Passarella, A., Das, S., . The internet of people (iop): A new wave in pervasive mobile computing. Pervasive Mob. Comput 41, 1–27. doi:10.1016/j.pmcj.2017.07.009.

[12] Cooper, C., Hochachka, W., Dhondt, A., . The opportunities and challenges of citizen science as a tool for ecological research. Citiz. Sci. Public Particip. Environ. Res , 99–113.

[13] Cordero, J.M., Borge, R., Narros, A., 2018. Using statistical methods to carry out in field calibrations of low cost air quality sensors. Sensors and Actuators B: Chemical 267, 245–254. URL: https://doi.org/10.1016/j.snb.2018.04.021https://linkinghub.elsevier.com/retrieve/pii/S0925400518307159, doi:10.1016/j.snb.2018.04.021.

[14] Crall, A., Kosmala, M., Cheng, R., Brier, J., Cavalier, D., Henderson, S., Richardson, A., 2017. Volunteer recruitment and retention in online citizen science projects using marketing strategies: lessons from Season Spotter. Journal of Science Communication 16, 1–29. URL: https://jcom.sissa.it/archive/16/01/JCOM_1601_2017_A01, doi:10.22323/2.16010201.

[15] EEA (European Environment Agency), 2019. Assessing air quality through citizen science. Technical Report 19. EEA. URL: https://www.eea.europa.eu/publications/assessing-air-quality-through-citizen-science/at_download/file, doi:10.2800/619.

[16] Feng, C., Wang, W., Tian, Y., Que, X., Gong, X., . Estimate air quality based on mobile crowd sensing and big data, in: 2017 IEEE 18th Int. Symp. A World Wireless, Mob. Multimed. Networks, p. 1–9. doi:10.1109/WoWMoM.2017.7974308.

[17] Freitag, A., Meyer, R., Whiteman, L., . Correction: Strategies employed by citizen science programs to increase the credibility of their data. Citiz. Sci. Theory Pract 1. doi:10.5334/cstp.91.

[18] Ganti, R., . Mobile crowdsensing: Current state and future challenges. IEEE Commun. Mag 49, 32–39.

[19] Grossberndt, S., Liu, H.Y., . Citizen participation approaches in environmental health. Appl. Environ. Educ. Commun 11, 178–188.

[20] Guo, B., Yu, Z., Zhou, X., Zhang, D., . From participatory sensing to mobile crowd sensing, in: 2014 IEEE Int. Conf. Pervasive Comput. Commun. Work. PERCOM Work. 2014, p. 593–598. doi:10.1109/PerComW.2014.6815273.

[21] Guo, J., Marshall, A., Zhou, B., 2014. A multi-parameter trust framework for mobile ad hoc networks, in: Rawat, D., Bista, B., Yan, G. (Eds.), Security, Privacy, Trust, and Resource Management in Mobile and Wireless Communications. IGI Global. chapter 11, p. 245–277. doi:10.4018/978-1-4666-4691-9.ch011.

[22] Habib, S.J., Marimuthu, P.N., 2017. Reputation analysis of sensors' trust within tabu search, in: Rocha, Á., Correia, A.M., Adeli, H., Reis, L.P., Costanzo, S. (Eds.), Recent Advances in Information Systems and Technologies, Springer International Publishing, Cham. pp. 343–352.

[23] Hecker, S., Bonney, R., Haklay, M., Hölker, F., Hofer, H., Goebel, C., Gold, M., Makuch, Z., Ponti, M., Richter, A., Robinson, L., Iglesias, J., Owen, R., Peltola, T., Sforzi, A., Shirk, J., Vogel, J., Vohland, K., Witt, T., Bonn, A., a. Innovation in citizen science – perspectives on science-policy advances. Citiz. Sci. Theory Pract 3, 1–4. doi:10.5334/cstp.114.

[24] Hecker, S., Garbe, L., Bonn, A., b. Mapping the citizen science landscape across europe.

[25] Kanjo, E., 2010. NoiseSPY: A real-time mobile phone platform for urban noise monitoring and mapping. Mobile Networks and Applications 15, 562–574. doi:10.1007/s11036-009-0217-y.

[26] Kardous, C., Shaw, P., . Evaluation of smartphone sound measurement applications ( apps ) using external microphones—a follow-up study. J. Acoust. Soc. Am 140, 327– 333. doi:10.1121/1.4964639.

[27] Kardous, C.A., Shaw, P.B., 2014. Evaluation of smartphone sound measurement applications. The Journal of the Acoustical Society of America 135, EL186–EL192. URL: http://asa.scitation.org/doi/10.1121/1.4865269, doi:10.1121/1.4865269.

[28] Komninos, N., Pallot, M., Schaffers, H., 2013. Special Issue on Smart Cities and the Future Internet in Europe. Journal of the Knowledge Economy 4, 119–134. URL: http://link.springer.com/10.1007/s13132-012-0083-x, doi:10.1007/s13132-012-0083-x.

[29] Leao, S., Ong, K.L., Krezel, A., 2014. 2Loud?: Community mapping of exposure to traffic noise with mobile phones, in: Environmental Monitoring and Assessment, pp. 6193–6206. doi:10.1007/s10661-014-3848-9.

[30] Longo, A., Zappatore, M., Bochicchio, M., . Collaborative learning from mobile crowd sensing: a case study in electromagnetic monitoring, in: 2015 IEEE Glob. Eng. Educ. Conf, p. 742–750.

[31] Longo, A., Zappatore, M., Bochicchio, M., 2017. Crowd-Sourced Data Collection for Urban Monitoring via Mobile Sensors. ACM Trans. Internet Technol 18, 1–21. URL: https://doi.org/10.1145/3093895, doi:10.1145/3093895.

[32] Longo, A., Zappatore, M., De Matteis, A., 2019. An osmotic computing infrastructure for urban pollution monitoring. Software: Practice and Experience 50, 533–557. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.2721, doi:10.1002/spe.2721.

[33] McCrory, G., Veeckman, C., Claeys, L., 2017. Citizen Science Is in the Air – Engagement Mechanisms from Technology-Mediated Citizen Science Projects Addressing Air Pollution, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). volume 10673 LNCS, pp. 28–38. URL: http://link.springer.com/10.1007/978-3-319-70284-1_3, doi:10.1007/978-3-319-70284-1{_}3.

[34] Momani, M., Challa, S., 2010. Survey of trust models in different network domains. CoRR abs/1010.0168. URL: http://arxiv.org/abs/1010.0168, arXiv:1010.0168.

[35] Mori, H., Ito, M., Sezaki, K., . Early fire alert system during an evacuation with mobile sensing technology. JDR 12, 320–328.

[36] Mukherjee, A., Stanton, L.G., Graham, A.R., Roberts, P.T., 2017. Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama valley of California. Sensors (Switzerland) 17. doi:10.3390/s17081805.

[37] Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K., . The future of sitizen science: Emerging technologies and shifting paradigms. Front. Ecol. Environ 10, 298–304. doi:10.1890/110294.

[38] Ottaviano, M., Beltrán-Jaunsarás, M.E., Teriús-Padrón, J.G., García-Betances, R.I., González-Martínez, S., Cea, G., Vera, C., Cabrera-Umpiérrez, M.F., Waldmeyer, M.T.A., 2019. Empowering citizens through perceptual sensing of urban environmental and health data following a participative citizen science approach. Sensors (Switzerland) 19. doi:10.3390/s19132940.

[39] Panichpapiboon, S., Leakkaw, P., . Traffic density estimation: A mobile sensing approach. IEEE Commun. Mag 55, 126–131. doi:10.1109/MCOM.2017.1700693.

[40] Peris, E., 2020. Environmental noise in Europe - 2020. Technical Report 22/2019. European Environment Agency. doi:10.2800/686249.

[41] Pipino, L., Lee, Y., Wang, R., . Data quality assessment. Commun. ACM 45, 211–218. doi:10.1145/505248.506010.

[42] Preece, J., . Citizen science: new research challenges for human-computer interaction. Int. J. Hum. Comput. Interact 32, 585–612.

[43] Rai, A.C., Kumar, P., 2017. Summary of air quality sensors and recommendations for application. Technical Report 689954. Ref. Ares. URL: https://www.iscapeproject.eu/wp-content/uploads/2017/09/iSCAPE_D1.5_Summary-of-air-quality-sensors-and-recommendations-for-application.pdf.

[44] Rana, R., Tung, C., Bulusu, N., Kanhere, S., Hu, W., . Ear-phone : A context-aware noise mapping using smart phones. Pervasive Mob. Comput 17, 1–22. doi:10.1016/j.pmcj.2014.02.001.

[45] Reyna, K.S., Rollins, D., Rollins, D., 2017. Retention and Efficacy of Citizen Scientist Volunteers of the Texas Quail Index, in: National Quail Symposium Proceedings, p. 7.

[46] Rutten, M., Minkman, E., van der Sanden, M., 2017. How to get and keep citizens involved in mobile crowd sensing for water management? A review of key success factors and motivational aspects. Wiley Interdisciplinary Reviews: Water 4, 1–10. URL: http://doi.wiley.com/10.1002/wat2.1218, doi:10.1002/wat2.1218.

[47] SciStarter, . Scistarter. URL: https://scistarter.com/about. accessed January 22, 2018).

[48] Shirk, J., Ballard, H., Wilderman, C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, C., Krasny, M., Bonney, R., . Public participation in scientific research: a framework for intentional design. Ecol. Soc 17, 29. doi:10.5751/ES-04705-170229.

[49] Silvertown, J., . A new dawn for citizen science. Trends Ecol. Evol 24, 467–471. doi:10.1016/j.tree.2009.03.017.

[50] Swanson, A., Kosmala, M., Lintott, C., Packer, C., . A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol 30, 520–531. doi:10.1111/cobi.12695.

[51] Thornhill, I., Loiselle, S., Lind, K., Ophof, D., . The citizen science opportunity for researchers and agencies. Bioscience 66, 720–721.

[52] Townsend, A., Chisholm, A., . Citizen urban science - new partnerships for advancing knowledge.

[53] Travis, M., Shivangi, P., Alberto, B., Mohammad, A.E., 2019. Evaluation of smartphone sound level meter applications as a reliable tool for noise monitoring. Journal of Occupational and Environmental Hygiene 16, 620–627. doi:10.1080/15459624.2019.1639718.

[54] Truong, N.B., Lee, G.M., Um, T.W., Mackay, M., 2019. Trust evaluation mechanism for user recruitment in mobile crowd-sensing in the internet of things. IEEE Transactions on Information Forensics and Security 14, 2705–2719.

[55] Wenger, E., a. Communities of practice and social learning systems: the career of a concept, in: Blackmore, C. (Ed.), Soc. Learn. Syst. Communities Pract, Springer Verlag and the Open University. p. 179–198.

[56] Wenger, E., b. Communities of Practice: Learning, Meaning and Identity. Cambridge University Press.

[57] W.H.O., . Ambient air pollution: A global assessment of exposure and burden of disease.

[58] Williams, R., Kilaru, V.J., Snyder, E.G., Kaufman, A., Dye, T., Rutter, A., Russell, A., Hafner, H., 2014. Summary for Policymakers, in: Intergovernmental Panel on Climate Change (Ed.), Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge. 1, pp. 1–30. URL: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=277996&simpleSearch=1&searchAll=air+sensor+guidebookhttps://www.cambridge.org/core/product/identifier/CBO9781107415324A009/type/book_part, doi:10.1017/CBO9781107415324.004.

[59] Yadav, P., Darlington, J., 2016. Design Guidelines for the User-Centred Collaborative Citizen Science Platforms. Human Computation 3. URL: http://arxiv.org/abs/1605.00910http://hcjournal.org/ojs/index.php?journal=jhc&page=article&op=view&path[]=10.15346%2Fhc.v3i1.11, doi:10.15346/hc.v3i1.11.

[60] Zappatore, M., Loglisci, C., Longo, A., Bochicchio, M., Vaira, L., Malerba, D., 2019. Trustworthiness of context-aware urban pollution data in mobile crowd sensing. IEEE Access PP, 1–1. doi:10.1109/ACCESS.2019.2948757.

[61] Zappatore, M., Longo, A., Bochicchio, M., a. Empowering active citizenship via byod-mediated learning activities: an experience in acoustics, in: 2017 IEEE Glob. Eng. Glob. Educ. Conf, p. 1174–1182. doi:https://doi.org/978-1-5090-5467-1/17/.

[62] Zappatore, M., Longo, A., Bochicchio, M., Zappatore, D., Morrone, A., Mitri, G., b. A crowdsensing approach for mobile learning in acoustics and noise monitoring, in: Proc. 31st Annu. ACM Symp. Appl. Comput. - SAC '16, p. 219–224. doi:10.1145/2851613.2851699.

[63] Zappatore, M., Longo, A., Bochicchio, M.A., 2016. Enabling MOOL in acoustics by mobile crowd-sensing paradigm, in: 2016 IEEE Global Engineering Education Conference (EDUCON), IEEE. pp. 733–740. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7474633, doi:10.1109/EDUCON.2016.7474633.

[64] Zupančič, E., Zalik, B., 2019. Data trustworthiness evaluation in mobile crowdsensing systems with users' trust dispositions' consideration. Sensors 19, 1326. doi:10.3390/s19061326.