

Noname manuscript No.
(will be inserted by the editor)

Toward Self-monitoring Smart Cities: the OpenSense2 Approach

Jean-Paul Calbimonte · Julien Eberle · Karl Aberer

Received: date / Accepted: date

Abstract The sustained growth of urban settlements in the last years has had an inherent impact on the environment and the quality of life of their inhabitants. In order to support sustainability and improve quality of life in this context, we advocate the fostering of ICT-empowered initiatives that allow citizens to self-monitor their environment and assess the quality of the resources in their surroundings. More concretely, we present the case of such a self-monitoring Smart City platform for estimating the air quality in urban environments at high resolution and large scale. Our approach is a combination of mobile and human sensing that exploits both dedicated and participatory monitoring. We identify the main challenges in such a crowdsensing scenario for Smart Cities, and in particular we analyze issues related to scalability, accuracy, accessibility, privacy, and discoverability, among others. Moreover, we show that our approach has the potential to empower citizens to diagnose their environment using mobile and portable sensing devices, combining their personal data with a public higher accuracy air quality network.

Keywords Sensors · Internet of Things · Semantic sensors · Smart cities · Air Quality

1 Introduction

Cities are complex systems in which different actors and components interact, responding to internal and external stimuli. Every city has unique characteristics, such as the geography, demography, environmental conditions, etc., which gives them a certain identity [1]. The different components of a city include its natural resources, fauna, public infrastructure, residential buildings, industry, or traffic, to name a few. However, the main driving actor of a city are the citizens, who are the only who can knowingly take decisions on how to change and adapt to different situations during the lifetime of an urban environment.

Smart cities have been characterized by the investment in human and social capital and infrastructure, with the goal of attaining a sustainable economic growth and high quality of life, through participatory governance, as defined in [2]. This conceptualization puts emphasis on the participatory nature of decisions and actions in a smart city, changing the top-down paradigm that is currently in place in most urban environments.

Some of the key concerns of citizens in urban areas is related to their interactions with the local environment. This includes the sustainable use of natural resources, the efficient management of energy, the disposal of waste and residues, or the emission and exposure to pollutants. Pollution is a particularly sensible issue, as it affects the life and health of the citizens, and of the society in general. In particular, *air pollution* is one of the most important and most studied cases, given its direct impact, not only on the health and lifestyle of citizens, but also on the rest of the components of an urban settlement [3]. In fact, it is estimated that around 7 million premature deaths are attributable to air pollution [4] worldwide. Cardiopulmonary, cardio-

Jean-Paul Calbimonte
Institute of Information Systems, HES-SO Valais-Wallis
Techno Pole 3, CH-3960 Sierre, Switzerland
E-mail: jean-paul.calbimonte@hevs.ch

Jean-Paul Calbimonte · Julien Eberle · Karl Aberer
Distributed Information Systems Laboratory, EPFL
Station 14, CH-1015 Lausanne, Switzerland
E-mail: firstname.lastname@epfl.ch

vascular and ischemic heart diseases have been linked to sustained exposure to certain air pollutants in urban environments, making it a global health concern.

In order to explain air quality conditions in a city, it is important to understand how the air pollutants are produced and how they disperse in a certain geographical area. This challenge has been addressed from different points of view. One possibility is to elaborate a model that simulates the air pollutants behavior, e.g. based on dispersion models, pollution map interpolation models, regression models, etc. [5]. These typically require a set of input data in order to be generated, including accurate reference air quality measurements, pollutant sources, city street models, wind direction, temperature and other weather data, among others. However, it has been evidenced that air pollution conditions present highly localized patterns [6], which can vary greatly from one street to the next one. Moreover, coverage of city-wide air quality observations can be impractical and unfeasible using standard stations, as it would be prohibitive in terms of costs, and also because of mobility restrictions [7].

The OpenSense2¹ project aims at integrating air quality measurements captured by heterogeneous mobile and crowdsensing data sources, in order to understand the impact of urban air pollution exposure in the citizens' health. In this paper we use this project as an example of how we can empower citizens through the use of information and communication technologies, to enable city-wide monitoring solutions where we mix both participatory and infrastructure sensing in a coherent distributed system. Our experience in this complex deployment reveals that we are able to cope with many of the main outstanding challenges in smart city monitoring. Furthermore, through a real use case in the city of Lausanne, we show the feasibility of the approach, and the various advantages that it brings over alternative solutions. Finally, we provide a set of ICT open source tools, which can be reused in similar projects in different domains.

The remainder of the paper is structured as follows. First, we present the state of the art in Section 2. In Section 3 we introduce the main characteristics of a citizen-driven approach in this context. Then, we introduce the air quality monitoring use case as a motivating example. In Section 5 we identify the main features of ICT solutions for participatory monitoring in a Smart City. Afterwards, we describe the overall approach of OpenSense2 in Section 6. In Section 7 we provide more concrete details about the Lausanne deployment, before concluding in Section 8.

2 Related Work

Air Quality monitoring has been widely studied, and different approaches have been used, ranging from sensors on bicycles [8,9], vehicles [10,11] and hand-held devices [12]. Interactions between different types of sensors in this case, remains an important question, and it is still needed to understand how different types of sensor networks can be combined effectively. This becomes even more complex in the context of crowdsensing, where incentives will need to play a role [13], and malicious behavior needs to be countered.

Existing air quality sensing systems have focused on the device and sensor layers, while disregarding the data management aspect [14]. In most of these systems the semantics of the data is hidden or implicit in ad-hoc data schemas and data source descriptions.

In the more general spectrum of smart cities, different previous works have outlined the principles that govern these complex environments [15]. In particular, architectures such as the one described in [16] have shown how heterogeneous devices deployed in an urban settlement can be combined for monitoring purposes. Diving into a more personal level, other works have studied the use of ICT for self-tracking [17], and the social consequences of these trending practices. These technologies have also been widely used for different types of health related activities [18,19], going beyond existing diagnosis and prevention practices.

About semantic sensor data management, a good number of works has focused on Linked Data publishing, considering streams as the fundamental building block. More specifically, there have been projects focusing on different aspects related to sensing, including coastal flood monitoring [20], earth observation remote sensing [21], or natural hazards [22].

3 Citizen Participation in Smart Cities

Smart city projects and initiatives have initially put the focus on public administrations and government organizations as main drivers. The reason is not artificial: in fact, local governments have the financial and logistical means to put in place IT infrastructures and enforce policies at a city-wide level. Whether it is for deploying sensors on roads, or public transportation, or installing smart meter facilities, these governing organizations can coordinate with different public and private institutions to manage this infrastructures (See Figure 1). Even if at this level there may be well-established procedures to enforce city policies, this top-down approach may encounter issues when it reaches

¹ OpenSense2 project: <http://opensense.epfl.ch>

the most sensible and heterogeneous layer of the society: *the citizens*. It is often the case that there is a mismatch between expectations and the actual implementation of smart city policies, when these are put in place by public administrations. This may be due to a lack of a detailed requirement elicitation, planning discrepancies, or unforeseen changes in the local environment.

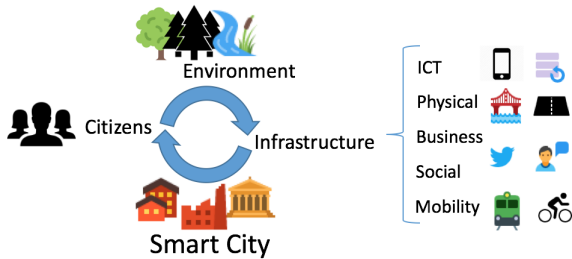


Fig. 1 Smart City interactions among citizens, environment and infrastructure, which may be physical, ICT, social, mobility, etc. Different governance models can be implemented, resulting in different kinds of interactions between these actors.

Complementary to top-down approaches, emergent initiatives from the citizens themselves can provide several advantages for the deployment of smart city projects. These include:

- **Acceptance.** As these initiatives originate on the actual users of the smart city infrastructure, they are easily adopted and the rejection risks are lower.
- **Impact.** The requirements in this case emerge from real-life situations and open problems that have direct impact over the lives of the targeted population.
- **Customization.** By definition, citizen-driven initiatives consider specific requirements and needs of the end users, and corresponding to the local context.
- **Relative Costs.** Although it is very hard to estimate the costs in this case, these initiatives in principle start with no budget from public administrations and can be managed autonomously.
- **Openness.** The inherent nature of these projects lends itself to be implemented with open-source software and hardware, which can be later consulted, revised and modified.
- **Reuse.** Once such an initiative is successful, it can be replicated in different locations under similar circumstances. Reuse can be done at multiple levels, e.g. data reuse, infrastructure reuse, adaptation of the ICT technologies, methodology, etc.

This approach, however, can entail several challenges. Of these, we can highlight the following:

- **Heterogeneity.** Spontaneous smart city projects may use different technologies, communication protocols, standards and data repositories, which may be managed independently from other urban infrastructures. When it comes the time to integrate these solutions with other facilities, there might be difficulties in orchestrating a more complex system that includes the citizen-driven parts.
- **Compatibility.** In bottom-up smart city projects, the technical decisions may have diverse motivations leading to solutions that are not necessarily compatible with other available infrastructures. It can even be the case that several similar initiatives are launched in the same metropolitan area, potentially overlapping in terms of goals but having compatibility issues.
- **Scalability.** The scope of participatory citizen initiatives can be limited to certain boundaries and contexts. Then, if the proposed solution is applied to a wider context, the system may not scale as expected.
- **Efficiency.** Similarly, the deployed technologies may not be efficient enough in terms of energy, processing querying, etc.
- **Reliability and Accuracy.** Specially for monitoring projects, the accuracy of the collected data may not always reach the desirable levels of quality needed for the actual needs.
- **Privacy.** Concerns about the privacy of data, especially for monitoring, are often disregarded in informal spontaneous citizen projects. Given that this is sometimes an orthogonal issue in the domain of application, it doesn't receive the attention it should have. However, very sensible information, including geo-location and contextual data, should be protected, while guaranteeing a desired utility.
- **Cyber-threats.** Data vulnerability in IoT deployments is an open issue that still requires to be studied, especially for crowd-based scenarios. Citizen sensing can also be the source of different threats, given the potential weaknesses of smartphones and wearables which are susceptible to different types of attacks, including tampering, denial of service, data falsification, identity theft, etc.

4 Air Quality Monitoring: a Use Case

The domain of application in this work is air quality monitoring. It is a particularly interesting case for

citizen-driven sensing, as it is a concern for urban population and it needs fine grained information about pollutant concentrations.

In the context of a city, ambient air quality can be assessed by analyzing the concentration of different pollutants present in the air. These pollutants are emitted by different sources and have distinct characteristics. In consequence, there are different ways and methods to detect them. Competent bodies such as the European Environment Agency, recommend monitoring a list of key pollutants for assessing air quality monitoring. These include: Carbon monoxide (CO), Nitrogen dioxide (NO_2), Nitrogen monoxide (NO), Sulfur dioxide (SO_2), Ground level Ozone (O_3), Particulate matter (PM) and Lead (Pb).

These pollutants appear under different conditions and require distinct technologies for sensing. CO is mostly produced by incomplete combustion processes, which include sources such as transportation, industry, and households. In the case of NO_2 and NO (commonly called NO_x) are mainly produced by the transportation sector. PM is also a product of transportation emissions, as well as heating and industrial processes. PM exists in different ranges of sizes, e.g. 10, 2.5 nm and ultrafine particles (UFP). O_3 is formed by reactions of other pollutants such as NO_x and volatile organic compounds.

In order to provide accurate concentration measurements of these pollutants, it is necessary to deploy a sensing infrastructure that spans a vast geographical region, and at a considerably high resolution. In this paper we focus in our experience with an air quality sensing deployment in the city of Lausanne (Switzerland), in the context of the OpenSense2 project². Conventional air quality monitoring in urban settlements such as Lausanne has been traditionally done using high-accuracy reference stations. The main problem with this approach is that it fails to capture the fine grained dynamics of air pollutants, and instead provides only a local view of the location where the reference station is placed. A possible alternative would be to install several high-end stations spread over the city. However, the costs of the equipment and the maintenance they require would not be assumable. Furthermore, even with a dozen of such stations the measurements would still be highly localized.

These limitations motivate the investigation of alternative approaches that change the focus of the monitoring approach, from a centralized to a participatory paradigm.

5 Empowering Citizen Monitoring through ICT

Citizens have direct access to local events and places which are sometimes uncovered by official monitoring stations. In the case of air quality monitoring this can bring a competitive advantage, given that the air pollutant behavior can be very different even within a small spatial distance. Air street canyons and the diversity of emission sources can result in very different pollutant concentrations from one street to the next one. However, as we have seen in Section 3 there are several challenges and issues that may limit the applicability of citizen-driven monitoring solutions, many of these challenges related to technical constraints.

In order to address these challenges, we propose providing the citizens with a comprehensive –but easy-to-use– set of ICT services and equipment that can be adapted to different monitoring scenarios (Figure 2). This infrastructure should provide the means not only to connect to different sensing devices, but it should also be capable of interconnecting with other participants, so that a smart city monitoring network can be established.

More concretely, we identify the main features for the ICT infrastructure:

- **Mobile sensing.** The foremost requirement for such an infrastructure is the capability of sensing through portable mobile applications. We advocate the use of smartphones, which can themselves act as sensors (e.g. using their accelerometers, light sensor, GPS). In addition, external devices can be plugged to the smartphone, via short range communication (e.g. Bluetooth). Smartphones can also capture data that does not necessarily come from devices attached to it but also from human perception of the environment that is input as text, microposts, pictures, etc. Citizen sensing applications should allow providing input through these and other types of sources, in such a way that the smartphone acts as a gateway with minimal storage, processing and communication capabilities.
- **Data publishing and sharing.** The ability to communicate the sensed data is crucial in order to allow the crowd-sensing scenario to be possible. This includes not only the data flow among citizen sensors, but also external actors such as sensing stations, publishing servers, etc. For these interactions, it is key to provide a standards-based stack, and we propose to base it on REST services, using standard semantic annotations to enhance integration and discoverability.

² OpenSense2 project, funded by Nano-Tera.ch: <http://opensense.epfl.ch>

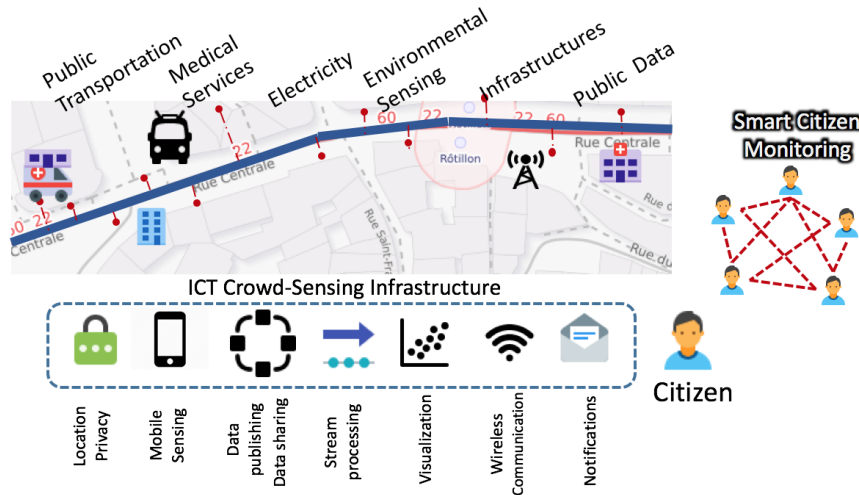


Fig. 2 ICT infrastructure for smart city monitoring. Citizens need to interact with different services and other infrastructure present in the city (e.g. transportation, medical services). ICT solutions are needed in order to provide communication with these external entities, as well as other features, including privacy protection, mobile sensing capabilities, visualization tools, stream processing, among others. It should also be possible to provide coordination mechanisms with other citizens in the network.

- **Stream processing.** The data collected by citizen devices arrives as continuous streams of information in potentially high rates. Online processing of real time data is a key feature for producing higher-level information from the incoming streams. Complex Event Processing (CEP) or Stream Reasoning approaches are examples of processing modules that should be pluggable to the system, either locally or applied to a network of citizen sensors.
- **Notifications.** Alerts in real-time are commonplace in monitoring applications. These are typically launched following continuous queries or rules applied over the sensor data streams. We advocate the use of push-based notifications and alerts, which could potentially be integrated into a publish-subscribe scheme.
- **Communication.** Different protocols exist for IoT devices and they target different situations and objectives. A sensing infrastructure should target standard protocols for communication and data exchange with special attention to energy efficiency and including conventional security mechanisms.
- **Data Privacy.** Contextual data from citizens can be directly revealed in a monitoring scenario. For example, the current location (obtained from the smartphone GPS) is continuously sent along other sensed properties and this could lead to disclose a citizen's daily patterns, habits and activities. We propose to include a built-in location privacy protection mechanism for all context-sensing data that is published or shared with external parties.
- **Visualization.** Plots and widgets are important in order to provide simple-but-effective means for monitoring and observing the behavior of sensed properties over time. Citizens can observe patterns or do quick visual analysis over the latest values collected by the sensing infrastructure.



Fig. 3 Hybrid and heterogeneous monitoring environment. A combination of different types of sensing actors can help overcome the issues of coverage, accuracy, and dynamicity in the context of a Smart City.

6 OpenSense2: Participatory Air Quality Monitoring

In the OpenSense2 project, we advocate for a hybrid and heterogeneous sensing environment (see Figure 3) where reference stationary stations, mobile sensing devices on public transportation, and crowdsensing, contribute to an overall city-wide view of air quality conditions. This approach has well-identified advantages, which we can analyze from different perspectives:

Coverage In OpenSense2, we rely on sensors installed on public transportation vehicles, in order to reach the

main streets and avenues of the city, covering most of the urban geography. In this way, we reuse an existing mobile infrastructure with predictable schedules and regular patterns over time. However, this network does not have access to the complete geographical extent of the city, e.g. car-free alleyways, open markets, squares, parks, and streets out of the transportation network. For these fine-grained measurements and less accessible places, we need to rely on the mobile crowdsensing platform, even if their measurements are more unpredictable, and potentially less accurate.

Dynamicity The co-existence of different types of sensors in OpenSense2 introduces the need for coordination mechanisms between these devices. Depending on the location and accuracy requirements, one may reuse data from reference stations, public transportation, or crowdsensing. A reference station produces high fidelity observations but its value decreases when the targeted area is far away. The equipment mounted on public transportation constitutes the backbone of the OpenSense2 infrastructure, providing acceptable data accuracy for a wide area. These sensors are complementary to the crowdsensing network, which can be used for uncovered areas, if there is a sufficient number of reliable connected devices. While crowdsensing devices are cheaper but less accurate, the number of participants can compensate this factor, and still provide indicative measurements. This dynamic reuse of different sensing platforms is a key feature to enable a sustained measurement data flow in the system.

Extensibility Monitoring requirements differ from one use case to another, and even in the same deployment, the equipment can evolve over time. The OpenSense2 approach is extensible by-default: the devices can be replaced or upgraded, and even new modalities can be added progressively (e.g. measure new pollutants) without halting the data collection process. Such an ecosystem is also extensible in terms of scale. New mobile sensors and/or crowdsensing participants can integrate the network and extend the overall coverage.

Maintenance and Setup Reference stations require expensive setup and supervision, not only because of material costs, but also the need for specialized personnel. The mobile sensors on public transportation have the advantage of having their own power supply and virtually no transportation costs. Maintenance is still necessary, although components are less expensive and easier to replace if necessary.

Reliability Problems in one node or in one sector of the sensing platform can fall back on other part of the network, providing a resilient infrastructure that is flexible in case of unexpected events.

These features are vital for OpenSense2 (Figure 4), in order to attain the goal of establishing a city-wide monitoring platform. The project has two deployment sites (Zurich and Lausanne), which are used both as demonstrators of these technologies, and also as data science platforms for running health studies. Details on the Lausanne deployment can be found in Section 7.

Data management in OpenSense2 requires taking care of several aspects, given the scale and complexity of the system. Orchestrating a coherent and efficient flow of the data is necessary to ensure that raw measurements captured by the sensing devices, can be transformed into data products that can be used by scientists and the general public.

The data collection performed by the sensors is only the first step in this chain of operations. It produces time series that contain the measured values of the air pollutant measurements. This dataset consists of raw observations that cannot be directly interpreted and used by citizens and external applications. Before that, several tasks need to be performed, including data cleaning, interpolation, geo-location corrections, re-calibration, etc. Furthermore, it is required to exploit the spatio-temporal characteristics of the data, in order to understand its semantics. Given the highly localized nature of air pollutant concentrations, it is necessary to correct the location of the observations, and project them to segments that represent a street in the city. In this way, we can provide a view of the distribution of pollutants at street level, and even for sub-segments of a street. Similarly, in terms of time, we may be interested in different time granularity (e.g. hourly, weekly, monthly). It is easy to see that at different periods of the day, the pollutant concentrations differ (e.g. peak hours), and different daily patterns can be analyzed, depending on other contextual parameters. In the case of crowdsensing, the sampling can be very irregular and the distribution of participants over the city may be uneven. In these cases it is important to combine their measurements with those of the public transportation network, whenever it is possible.

After these spatio-temporal distributions of the pollutant concentrations are produced, they can be made available for further analysis and processing. Potential data products generated from these measurements include pollution maps, which can be complemented or validated with pollution models. Other outcomes are air quality models: e.g. log-linear regression models, lagrangian dispersion models, etc [5]. Finally, using these

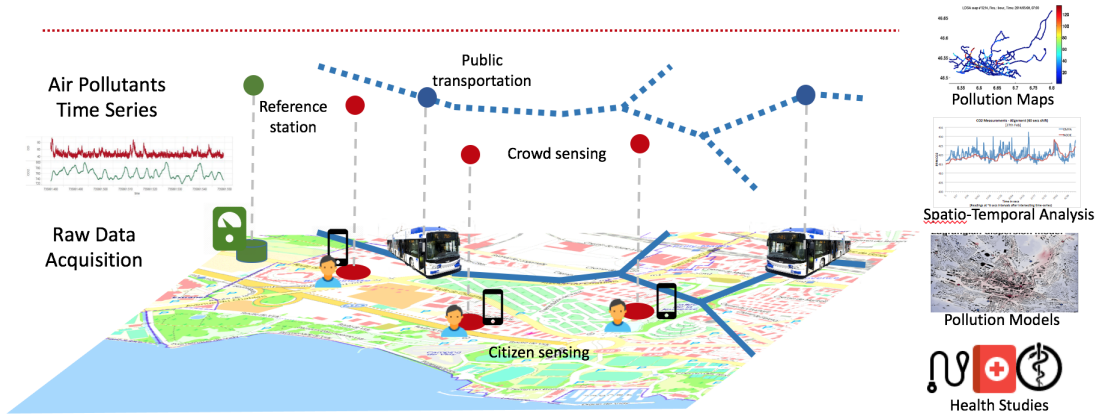


Fig. 4 The OpenSense2 approach for Air Quality monitoring in Smart Cities: different planes provide data at different granularities: raw observations, spatio-temporal aggregations, and air quality products and applications, including air pollution maps [23], models [24], mobile recommendations, etc.

models, end-user applications can be built, leveraging on the available processed datasets. Applications include personalized recommendations for reducing the exposure to pollutants, or support for health studies related to diseases associated to air pollution.

7 OpenSense2 Lausanne Deployment

In this section we describe the deployment of the OpenSense2 platform in the city of Lausanne. The public transportation sensors are installed and maintained in collaboration with the local transport authority³, which operates the buses and trolleybuses. Ten buses have been equipped with electrochemical sensors for CO and NO₂, a resistive O₃ sensor, and a Naneos Partector for particulate matter. Even if the number of buses is limited, it already provides a good coverage of the city, given that a particular bus may be used for different lines, thus changing its path over successive days. These changes are defined depending on availability, days of the week, peak hours and vacation periods. The buses cover different areas of the city, ranging from the lake-side to the city center and the northern suburbs. The sensor boxes installed on top of these buses collect the measurements and send them via GPRS communication to a base station every five minutes. The base station runs an instance of the GSN middleware [25] (Global Sensor Networks), which is devoted to manage the data life-cycle of the sensor observations.

The sensors deployed on the buses constitute the backbone of the smart city air quality monitoring platform. The reliability of the sensors on these buses, as well as their number and sampling frequency offer minimal conditions for providing relatively good data qual-

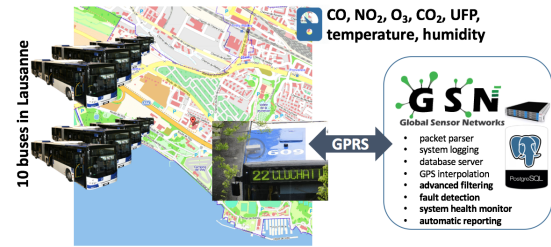


Fig. 5 OpenSense Lausanne deployment: 10 sensor boxes installed on top of buses of the local transportation network, are linked to GSN via GPRS communication. Sensor boxes report CO, NO_x, O₃, CO₂ and UFP measurements.

ity, at least in the main streets of the city. However, for the locations where buses are not accessible, OpenSense2 needs to rely on other sources of data. If this is the case, then the OpenSense2 Lausanne deployment needs to use the data from participatory sensing mobile devices. As part of this platform, we make use of tinyGSN [26], an android application that allows mobile phones to collect sensor data from built-in or external sensors, and –if necessary– report to a central GSN instance. With tinyGSN we are able to provide a common data acquisition, storage and processing environment for mobile and participatory sensing.

In the following, we describe the main characteristics of the GSN and tinyGSN, and how we use them in the project.

7.1 GSN: Distributed Data Processing

The GSN (Global Sensor Networks) middleware is deployed as the core back-end for stream processing and data publishing. Its mission is to handle the different stages of the data life-cycle of the incoming data streams.

³ Lausanne public transportation: <http://t-1.ch>

GSN is designed to work as a distributed deployment, which is one of the common needs in monitoring systems. The inherently decentralized architecture of GSN allows different instances to inter-operate, and each instance can expose a number of different virtual sensors (Figure 6).

Data input in GSN is managed by wrappers that provide a dedicated connection with the sensors, and are extensible enough to support different protocols and access methods. A set of ready-to-use wrappers (e.g. UDP, serial port, HTTP, etc.) is provided by default, and it is easy to extend them and add new data sources if needed.

The main data abstraction in GSN is the virtual sensor, which encapsulates the data captured by a certain wrapper. These virtual sensors can aggregate and reuse data from other virtual sensors, which allows defining different layers of data, according to specific use case needs. The streams produced by each virtual sensor have an output structure composed of one or more fields, which can be defined in terms of a continuous query running over one or more sources. After the data acquisition process, different custom data processing tasks can be applied on top of the virtual sensor data (e.g. data cleaning, classification, outlier detection, etc.). Finally, the data can be stored in a persistent repository, which can be customized (e.g. relational database, time series, triple store, column store, etc.).

GSN instances communicate remotely with each other through a native interface (inter-GSN communication) implemented on top of a MQ (ZeroMQ)⁴ message queue mechanism. This native interface helps providing a wide range of possible deployment set-ups, such that GSN instances can be located in different physical machines or data centers. Finally, an access layer on top of the services allows defining permissions over the virtual sensors and the observations they produce. The system has been implemented in Java, while some out-of-the box wrappers are implemented in other languages. The entire project is open-source, and is available in Github⁵, as a standalone project, with an existing and growing community of users and developers.

In the case of the Lausanne deployment, we had to surpass a series of challenges to successfully deliver the measurements as a coherent and searchable dataset. These challenges include: the particularities of the communication protocol with the bus sensor boxes, the data acquisition flow, and the need for health monitoring of the system. The modularity and flexibility of GSN contributed to address these issues satisfactorily. Communication among layers is operated through the ZeroMQ

wrapper, providing asynchronous communication, and data messaging with other remote instances of GSN. Experiments on this architecture have shown [27] that GSN can scale to sustain the load of low to mid-level rates of stream elements per virtual sensor when using the ZeroMQ-based internal communication system. Details of this configuration can be found on [28].

7.2 Crowdsensing with tinyGSN

The tinyGSN mobile platform provides a native application for Android that enables the configuration of virtual sensors –as in GSN– that directly gather the data captured by the available sensors installed in the phone (see Figure 7). These may include accelerometer, GPS, WiFi, etc., and these data can be collected internally, through a wrapper mechanism. Following the GSN principles, more virtual sensors can be added using custom wrappers, even if these are external to the smartphone, e.g. a NO sensor connected via Bluetooth. As it is the case in GSN too, it is possible to set up additional processing features on top of the sensed data. These may include local storage, cleaning, or privacy protection. The main limitation for these features is the limited battery life of the phone, which can be drained if these functionalities are overused. The data collected by tinyGSN is managed as a stream, and storage is optional. If data is not stored, it is temporarily kept during a certain window of time before being discarded.

While most of the data collection and processing is optimized to be executed locally, tinyGSN provides a data pull and push mechanism that allows sending the data to a centralized GSN server, which can concentrate measurements from different tinyGSN instances (e.g. for aggregation and data collection and integration purposes).

7.3 OpenSense2 Data Semantics

One key aspect in OpenSense2 is to provide the means to generate meaningful and semantically understandable data. This is possible thanks to the inclusion of human and machine-understandable metadata that describes the datasets provided by the platform. This allows offering transparent access to the data that is produced, so that it can be reused for different purposes, e.g. as feedback to citizens, as input for citizen scientists, or as validation data for pollution models, among others. Without proper meaningful information this goal could not be effectively achieved. As an example, consider the raw CSV data that GSN originally provides from the sensors in the buses:

⁴ ZeroMQ: <http://zeromq.org/>

⁵ GSN Github: <https://github.com/LSIR/gsn>

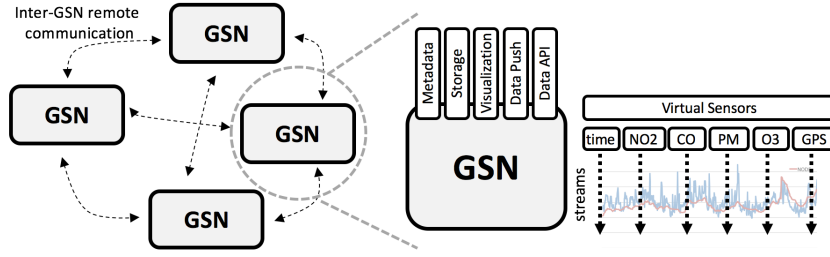


Fig. 6 GSN distributed architecture: different GSN instances may communicate remotely through inter-GSN communication. Each GSN instance may use different wrappers for data acquisition, and processing classes for additional data operations. Virtual sensors are the main data abstraction in GSN and can encapsulate real sensor data or aggregations and compositions of other virtual sensors.

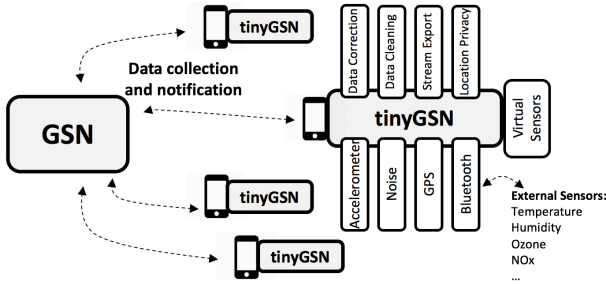


Fig. 7 Distributed architecture of tinyGSN: different tinyGSN-enabled smartphones may communicate remotely with a central GSN base station. Each tinyGSN device may use different wrappers for data acquisition, and processing classes for additional data operations. Each of them can also make use of native sensors included on the smart phone, or external ones plugged in through local low range communication (e.g. Bluetooth)

```
29-02-2016T16:41:24,47,369,46.52104,6.63579
29-02-2016T16:41:34,47,358,46.52344,6.63595
29-02-2016T16:41:44,47,354,46.52632,6.63634
29-02-2016T16:41:54,47,355,46.52684,6.63729
...
```

Listing 1 Example of mobile NO2 sensor readings.

Every line in the CSV file represents a measurement, and the same tabular structure is used for all buses, differentiated by the `station` column (2nd). Although some minimal metadata is provided (e.g. headers) we are missing important metadata such as the sensor specifications, measurements capabilities, units of measurement, etc. Furthermore, there is no information about the streets or avenues where data was collected, and in fact it is also mixed among concurrently reporting buses (i.e. data from different buses may be interleaved). Without any data pre-processing and contextualization, the data provided is of little use. To counter this problem, in OpenSense2 we propose a layered semantic data management model, where at each layer we provide additional value to the data, e.g. through semantic annotations that describe data clean-

ing and pre-processing, temporal and spatial aggregations, and finally event annotations [28].

The semantic annotations are described following the RDF standard (Resource Description Framework), which can use commonly agreed vocabularies to represent metadata information. For the sensor observations, RDF is not the most convenient representation for data sharing and transmission, as it can be too verbose and complex. As an alternative, we advocate the use of CSV as the common format for observation measurements, but augmented with rich semantic descriptions that follow the specifications of the CSV on the Web Working Group⁶. As an example of embedded metadata that can be provided for the OpenSense2 data sets, consider the JSON snippet below (Listing 2). It represents a description of a CSV output of sensor data from the GSN middleware, using the metadata model defined by the CSV on the Web group [29].

```
{ "@context": [ "http://www.w3.org/ns/csvw",
  { "@language": "en",
    "base": "http://opensense.epfl.ch/data/",
    "time": "http://www.w3.org/2006/time#",
    "ssn": "http://purl.oclc.org/NET/ssnx/ssn#",
    "qu": "http://purl.oclc.org/NET/ssnx/qu/qu#",
    "opensense": "http://opensense.epfl.ch/onto/os#" } ],
  "tableSchema": {
    "columns": [ {
      "name": "time",
      "titles": "Timestamp",
      "aboutUrl": "Instant_{time}",
      "propertyUrl": "time:inXSDDateTime",
      "datatype": { "base": "dateTime",
        "format": "yyyy-MM-ddTHH:mm:ss" }
    }, {
      "name": "sensor",
      "titles": "Bus sensor",
      "aboutUrl": "Obs_NO2_{sensor}_{time}",
      "propertyUrl": "ssn:observedBy",
      "valueUrl": "Sensor_{sensor}"
    }, {
      "name": "obsProperty",
      "virtual": true,
      "aboutUrl": "Obs_NO2_{sensor}_{time}",
      "propertyUrl": "ssn:observedProperty",
      "valueUrl": "opensense:NO2"
    }, {
      "name": "no2",
      "titles": "NO2 concentration",
      "aboutUrl": "ObsResult_NO2_{sensor}_{time}",
      "propertyUrl": "qu:numericalValue"
    }
  ]
}
```

⁶ CSV on the Web <http://www.w3.org/TR/csv2rdf/>

```

}, {
  "name": "unit",
  "virtual": true,
  "aboutUrl": "ObsResult_NO2_{sensor}_{time}",
  "propertyUrl": "qu:unit",
  "valueUrl": "opensense:ppm"
}]
}

```

Listing 2 Example of a JSON metadata for a CSV of an OpenSense NO2 dataset.

This simplified JSON metadata description provides URI descriptions of properties such as the time format/type, the units of measurement (e.g. `opensense:ppm`) or the observed property (e.g. `opensense:NO2`). These URIs are identifiers that could also be browsed in order to obtain more information about them, following the Linked Data principles [30].

7.4 Air Quality and Health Studies

Studying the impact of air pollutants exposure on citizen's health is one of the medium-long term goals of OpenSense2. However, it is often not trivial to carry out experiments in this area, given the many technical difficulties of human sensing. In particular, we have tackled some of the technical issues for performing this type of health studies, focusing mainly on the mobile sensing platform, and the activity recognition mechanisms. The mobile sensing platform is based on tinyGSN, and as we have seen, provides customizable and reliable data acquisition functionalities, as well as communication and integration with GSN, and basic processing and cleaning.

For this type of health studies it is not only necessary to measure the pollution levels, but also the estimated exposure of a participant. This estimation depends not only on the concentration of a pollutant at a certain geographical point, but also on the breath air intake. The amount of air (and therefore pollutants) that enters the respiratory system is evidently different depending on the type of activity that the participant is carrying out. The impact of certain pollutants can have different consequences for people's health, even for short term exposures. For these reasons, it is also key to have accurate activity recognition mechanisms that provide a model, which can be combined with the pollutant concentration in order to estimate the intake and exposure levels.

The crowdsensing platform based in tinyGSN will allow us to achieve this goal. Using tinyGSN-powered Android phones we can capture location (Figure 8) and accelerometer measurements of every participant. As it can be seen in the example, users can be traced as they wander around the city, and we are able to character-

ize their activities based on the accelerometer data and their location. Activity recognition is also a key element to provide personalized recommendations based on air quality levels. We plan to explore different types of recommendations, guided by the results of the data collection that is currently carried out in Lausanne in collaboration with the CHUV university hospital.

8 Conclusions & Future Work

In this paper we have analyzed different challenges for a participatory and collaborative city-wide monitoring system. We believe that the combination of government-led actions with citizen bottom-up approaches is the only way to guarantee the success of this type of initiatives, as they require a high level of commitment and adoption by the society. This idea could potentially be applied to different use cases and scenarios in different domains, and we have provided an overview of the main technical challenges for achieving it. We have also described a technical solution that we have developed and deployed, for the particular case of Air quality Monitoring, in the context of the OpenSense2 project. Our experience is focused on the data management aspects, which are orthogonal to domain specific issues, and we think that they can be of use for similar efforts in other latitudes and other types of applications for Smart Cities.

The deployment of this ICT infrastructure serves as a concrete example that smart cities can be able to self-diagnose (in this case, with regards to air pollution), and react to events that it detects through a community-based monitoring system. Then the citizens themselves can take action and propose measures and policies to minimize adverse effects. In the concrete case of air quality, these actions can help reduce the risks of morbidity related to air pollutants exposure.

In the future, we expect that this type of infrastructure, based on the use of mobile sensing platform, or even wearable devices, can help shifting the focus of smart city monitoring towards the citizens. We strongly believe that initiative that emerge from local populations have better chances of success, given that they have the knowledge and experience that other authorities often lack. While one of the main impediments for community-driven monitoring is related to the availability and understanding of related ICT solutions, this reality is rapidly changing. Nowadays, available technology is ready for end-user setup and utilization, under favorable conditions for allowing the development of this type of solutions.

The proposed platform that we have presented, is only a small example of what can be done in many

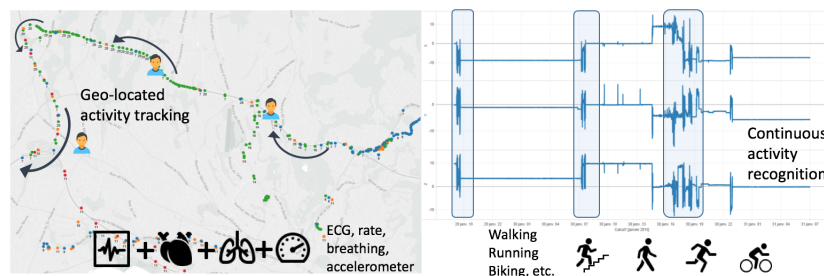


Fig. 8 OpenSense location and accelerometer traces captured by tinyGSN. In the left Figure (a), the location of a participant is plotted. Her trace allows to discern where she was and also how she moved (means of transportation). In the right hand side Figure (b), different accelerometer traces allow discovering basic activity patterns: stationary, walking, cycling, running, climbing stairs, etc.

other domains. We foresee that this type of deployment can be the basis for an ecosystem of applications on a city-wide infrastructure, which is devoted not only to monitoring, but also to take concrete community-based actions.

Acknowledgments

Partially supported Nano-Tera.ch OpenSense2 project, evaluated by the Swiss National Science Foundation.

References

1. A. Bagnasco, et al., *Cities in contemporary Europe* (Cambridge University Press, 2000)
2. A. Caragliu, C. Del Bo, P. Nijkamp, *Journal of urban technology* **18**(2), 65 (2011)
3. D.H. Tsai, I. Guessous, M. Riediker, F. Paccaud, J.M. Gaspoz, J.M. Theler, G. Waeber, P. Vollenweider, M. Bochud, *Journal of hypertension* **33**(6), 1144 (2015)
4. W.H. Organization. News release. <http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/>. (2014)
5. P. Zannetti, *Air pollution modeling: theories, computational methods and available software* (Springer Science & Business Media, 2013)
6. S. Vardoulakis, B.E. Fisher, K. Pericleous, N. Gonzalez-Flesca, *Atmospheric environment* **37**(2), 155 (2003)
7. D. Hasenfratz, O. Saukh, S. Sturzenegger, L. Thiele, *Mobile Sensing* pp. 1–5 (2012)
8. B. Elen, J. Peters, M.V. Poppel, N. Bleux, J. Theunis, M. Reggente, A. Standaert, *Sensors* **13**(1), 221 (2012)
9. J. Peters, J. Van den Bossche, M. Reggente, M. Van Poppel, B. De Baets, J. Theunis, *Atmospheric Environment* **92**, 31 (2014)
10. B. Buchli, M. Yuecel, R. Lim, T. Gsell, J. Beutel, in *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on* (IEEE, 2011), pp. 115–116
11. W. Hedgecock, P. Völgyesi, A. Ledecz, X. Koutsoukos, A. Aldroubi, A. Szalay, A. Terzis, in *Proceedings of the 2010 ACM Symposium on Applied Computing* (ACM, 2010), pp. 795–796
12. B. Elen, J. Theunis, S. Ingarra, A. Molino, J. Van den Bossche, M. Reggente, V. Loreto, *Sensing a Changing World* (2012)
13. G. Radanovic, B. Faltings, in *Twenty-Eighth AAAI Conference on Artificial Intelligence* (2014)
14. L. Sanchez, L. Muñoz, J.A. Galache, P. Sotres, J.R. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis, et al., *Computer Networks* **61**, 217 (2014)
15. T. Nam, T.A. Pardo, in *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (ACM, 2011), pp. 282–291
16. L. Filipponi, A. Vitaletti, G. Landi, V. Memeo, G. Laura, P. Pucci, in *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on* (IEEE, 2010), pp. 281–286
17. D. Lupton, Available at SSRN 2483549 (2014)
18. L.E. Burke, M.A. Styn, K. Glanz, L.J. Ewing, O.U. Elci, M.B. Conroy, S.M. Sereika, S.D. Acharya, E. Music, A.L. Keating, et al., *Contemporary Clinical Trials* **30**(6), 540 (2009)
19. A. Solanas, C. Patsakis, M. Conti, I.S. Vlachos, V. Ramos, F. Falcone, O. Postolache, P.A. Pérez-Martínez, R. Di Pietro, D.N. Perrea, et al., *IEEE Communications Magazine* **52**(8), 74 (2014)
20. A.J. Gray, J. Sadler, O. Kit, K. Kyzirakos, M. Karpathiotakis, J.P. Calbimonte, K. Page, R. García-Castro, A. Frazer, I. Galpin, et al., *Sensors* **11**(9), 8855 (2011)
21. M. Koubarakis, M. Sioutis, G. Garbis, M. Karpathiotakis, K. Kyzirakos, C. Nikolaou, K. Bereta, S. Vassos, C.O. Dumitru, D. Espinoza-Molina, et al., in *On the Move to Meaningful Internet Systems: OTM 2012* (Springer, 2012), pp. 932–949
22. A. Sheth, C. Henson, S.S. Sahoo, *Internet Computing, IEEE* **12**(4), 78 (2008)
23. D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, L. Thiele, in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on* (IEEE, 2014), pp. 69–77
24. E.C. Un, J. Eberle, Y. Kim, K. Aberer, in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (ACM, 2013), pp. 1143–1150
25. K. Aberer, M. Hauswirth, A. Salehi, in *Proc. 32nd International Conference on Very Large Data Bases VLDB (VLDB Endowment, 2006)*, pp. 1199–1202
26. J. Eberle, J.P. Calbimonte, K. Aberer. Efficiently gathering contextual information for health studies. <http://www.nano-tera.ch/pdf/posters2015/OpenSense252.pdf> (2015)
27. J.P. Calbimonte, S. Sarni, J. Eberle, K. Aberer, in *Proc. of the 7th International Workshop on Semantic Sensor Networks* (2014)

28. J.P. Calbimonte, J. Eberle, K. Aberer, in *In Proc. of the 6th Workshop on Semantics for Smarter Cities S4SC 2015, at ISWC* (2015)
29. J. Tennison, G. Kellogg, I. Herman. Model for tabular data and metadata on the web. <http://www.w3.org/TR/tabular-data-model/> (2015)
30. T. Berners-Lee, C. Bizer, T. Heath, IJSWIS **5**(3), 1 (2009)