Smart Emission

Building a Spatial Data Infrastructure for an Environmental Citizen Sensor Network

Michel Grothe
Geonovum
Amersfoort, The Netherlands
m.grothe@ geonovum.nl

Linda Carton
Radboud University
Nijmegen, The Netherlands
l.carton@fm.ru.nl

Just van den Broecke
JustObjects
Amstelveen, The Netherlands
just@justobjects.nl

Hester Volten
National Institute of Environment & Health
Bilthoven, The Netherlands
hester.volten@rivm.nl

Robert Kieboom
CityGIS
Den Haag, The Netherlands
Robert@citygis.nl

***Abstract*—Smart Emission is a citizen sensor network using low-cost sensors that enables citizens to gather data about environmental quality, like air quality, noise load, vibrations, light intensities and heat stress. This paper introduces the design and development of the data infrastructure for the Smart Emission initiative and discusses challenges for the future. The Spatial Data Infrastructure (SDI) is open and accessible on the Internet using open geospatial standards and (Web-) client applications. Smart Emission as a citizen sensor network offers several possibilities for heterogonous applications, from health determination to spatial planning purposes, environmental monitoring for sustainable traffic management, climate adaptation in cities and city planning.**

Keywords—Smart Emission, Citizens, Low-cost sensors, Spatial Data Infrastructure; Sensor Data; Geospatial Standards

I. INTRODUCTION

Today's technical advancements enable innovations of citizen sensor networks in cities, because more and more low-cost sensors are being invented, wireless communication infrastructures provide the means for information loops to be established over longer distances against relatively low costs, and big data tools are becoming available that make the handling of massive amounts of data flows affordable and doable [1].

At the basis of citizen sensor networks lies the observation that in multiple places, citizens organize themselves in networks (sometimes self-organized, sometimes participating in government-initiated participatory projects) with the aim to monitor and/or improve environmental qualities in their daily living environment [2]. This is strongly supported by increasingly available ICT technologies and low-cost sensing equipment [3], [4]. Technology is an enabling factor in this development. Moreover, the social trend of self-organization by citizens, taking responsibility of their own neighbourhood and region, is a driving force behind this emerging trend. In [5], the user perspective and the societal and policy dimensions of this emerging concept of citizen-sensor-networks has been analysed in more depth.

The Smart Emission initiative aims to establish an innovative citizen sensor network in a real-life ‘urban lab’ setting [2]. The initiative builds on the knowledge of technical and social innovation in implementing such a new citizen-sensor-network. It also reflects on the feedback provided by the information, and it’s potential consequence for citizens and government to explore new venues, options and (low-cost) strategies to further improve local environmental quality in dedicated places.

This paper is about the Smart Emission data infrastructure offering an open SDI, including the use of international, open standards to achieve interoperability and provide open access to the data on the Internet. The paper is structured in the following order. In the first section, the Smart Emission initiative is described in summary. Next, the data collection and data distribution infrastructures are outlined. Then, the cloud deployment architecture is introduced. In the next section, a short outline of the adopted open access principles and available user applications follows. The paper ends with a final section sketching challenges and outlook.

II. Smart Emission citizen sensor network

The Smart Emission initiative applies an innovative citizen sensor network, testing a network of low-cost sensors that is being put in place in the city together with experts and participating citizens, as a proof of concept.

The initiative aims to monitor the environmental qualities in a low-cost and efficient manner while simultaneously acquiring a detailed image in space-time, as the sensors are spread over many locations in the city. Ultimately, the sensing initiative serves to further improve environmental qualities in the built-up environment by monitoring and developing measures for environmental improvement in a bottom-up planning style, based on a collaborative and communicative planning philosophy [6], [7], [8].

To this end, it involves citizens living and working in the city who have been be invited to place a sensor on their house, garden, window or building property. As volunteers, citizens are involved in receiving the data and discussing the ‘bigger picture’ of analyzed data and visualizations in sense-making sessions for citizen feedback and participatory evaluation (building on existing knowledge from fields like participatory planning and citizen science). Citizens participate in this low-cost sensing initiative. Their use cases are the starting point for this environmental citizen-sensing SDI. The initiative organizes citizens sessions in which citizens perform collaborative sense-making: in dialogue with other citizens, the city government and scientists/researchers, the Smart Emission data is analyzed, visualized and interpreted in a collaborative way using associated tools, like Web apps and a Maptable.

The research conducted is shaped by the ideas of action research. This entails constructing a pilot version of a citizen sensor network in practice in a city, with the aim to become and remain operational during a certain time period.

To this end, a low-cost sensor unit called “Jose” was developed [9] and implemented to measure the spatial pattern and spread of environmental information such as air quality, noise load, vibration, light coloring and intensities and meteo like rainfall, temperature, air pressure and humidity in a fine-grained network constellation. As such, other citizen sensing initiatives in the environmental domain exist. However, the initiatives often have one aspect of environmental monitoring in scope, like air quality [10, 11, 12], noise load [13,14,15], meteo [16] or vibrations [17,18]. The Smart Emission initiative has a broad(er) environmental perspective. The heterogeneous sensor unit used can be applied for multiple applications.

Smart Emission is also about exploring how low-cost sensors can add value to high end sensing methods by collecting fine-grained urban measurement data, and which methods and scenarios can be used for processing and visualizing this data for involving citizens and connecting to broader city purposes.

At this moment, the Smart Emission initiative has started in the city of Nijmegen. In this midsize city in the Netherlands with approx. 170.000 inhabitants, about 35 Jose environmental sensor units are located at citizens’ homes. Negotiations to expand the Smart Emission concept to other cities in the Netherlands and outside the Netherlands (Belgium, Germany) have started.

III. Data collection

The Smart Emission data collection starts with the collection of data from environmental sensors. The Jose sensor unit (see figure 1) developed by the Dutch sensor company Intemo [8] collects different types of environmental indicators: air quality, noise load, vibration, light intensity and several meteo indicators.

Fig. 1. Smart Emission Jose sensor (top) and Jose sensor installation at site of the national air quality network (bottom)

This layered and extendable sensor unit offers the following environmental indicators: 1. Light intensity, 2. Light reflection, 3. Light (air) colour, 4. Earth vibration, 5. Carbon monoxide, 6. Nitric oxide, 7. Ozone, 8. Hydrogen, 9. Carbon dioxide, 10. Pressure, 11. Temperature (unit and environment), 12. Humidity, 13. Noise load. The Jose sensor unit is connected to a power supply by a USB phone adapter and to the Internet. Internet connection is made via WIFI or telecommunication network (using a GSM chip). Furthermore, Jose collects time and date and location by GPS (latitude, longitude). Jose has memory and a multi-colour display ring.

The data are encrypted as data streams and sent every 15 seconds from each individual Jose unit to the data production platform hosted by CityGIS (see figure 2).

The encrypted data is decrypted by a dedicated 'Jose Input Service' that also inserts the data streams into a MongoDB (www.mongodb.com) database using JavaScript Object Notation (JSON). This MongoDB database is the source production database, in which all raw sensor data streams of the Jose sensor units are permanently stored.

IV. DATA DISTRIBUTION

A dedicated Application Programming Interface (API), the 'Raw Sensor API', is developed for further distribution of the Smart Emission data to other platforms, like the Smart Emission open data distribution platform hosted at the FIWARE Lab NL[1]. Other applications of FIWARE in the environmental domain can be found in [20].

The data distribution infrastructure at the FIWARE LAB NL consists of several components (see figure 2):

1. Pre-processsing and post-processing algorithms based on Extract-Transform-Load (ETL) principles;
2. Data storage in Postgres/PostGIS database (DB);
3. Several Open Geospatial Consortium (OGC) based APIs;
4. Several apps / web viewers, like the "SmartApp" and "Heron".

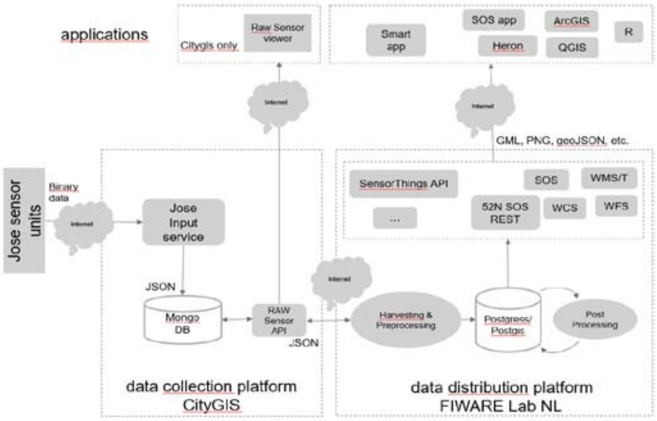

Fig. 2. Overall data architecture Smart Emission

In order to store the Smart Emission data in the distribution database, harvesting and pre-processing of the raw sensor data (from the CityGIS production platform) is performed. First, every minute a harvesting mechanism collects data from the production platform using the raw sensor API. The data encoded in JSON format is then processed by a multi-step ETL-based pre-processing mechanism. In several steps, the data streams are transformed to the Postgres DB.

Pre-processing is done specifically for the raw data of the air quality sensors. Based on a calibration activity, the raw data from the air quality sensors is transformed to 'better interpretable' values. For some of the environmental indicators, calibration procedures have been started. Especially the four air quality indicators (carbon monoxide, nitric oxide, ozone, carbon dioxide) need further calibration. Smart Emission air quality sensor data are compared to the measurements of two high-cost air quality sensor installations belonging to the national air quality network and operated by the National Institute for Public Health and the Environment (RIVM) and located in the city of Nijmegen (see figure 1). For calibration of the air quality indicators, the approach according to [20] was adopted and implemented.

Post-processing is the activity to transform the pre-processed values into new types of data using statistics (aggregations), spatial interpolations, etc..

The data distribution architecture of Smart Emission is further expanded below. Figure 3 sketches the architecture with an emphasis on the flow of data. This architecture sketches a multistep ETL approach which also is used within the 'INSPIRE SOS Pilot' approach for the implementation of European air quality e-reporting (see [21], [22] and sensors.geonovum.nl).

The multistep-ETL approach consists of three steps: harvesting, refinement (pre-processing and post-processing) and publishing (see figure 3).

---

[1] "The FIWARE platform provides a rather simple yet powerful set of APIs that ease the development of smart applications in multiple vertical sectors. The FIWARE Lab is a non-commercial sandbox environment where innovation and experimentation based on FIWARE technologies take place" (www.fiware.org).

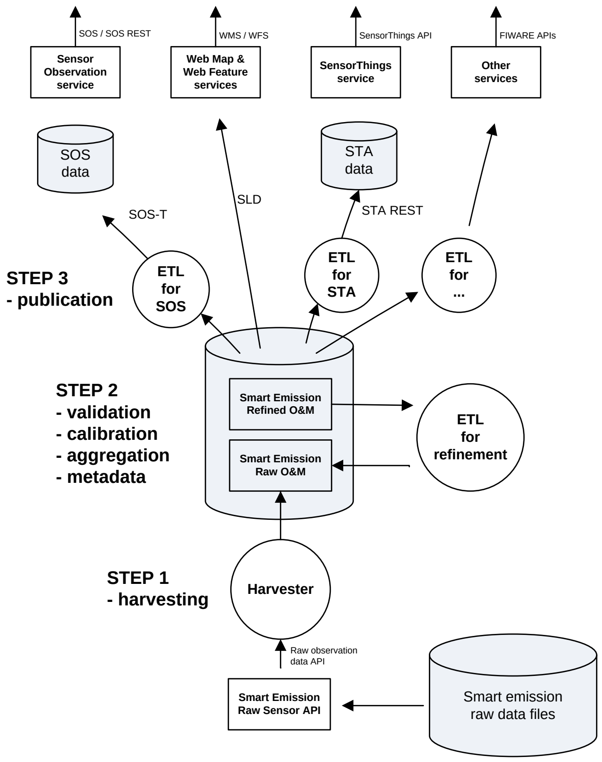

Fig. 3. Overall Architecture with ETL Steps[2]

The ETL design comprises these main processing steps (see figure 4):

- Step 1 – Harvesting: fetch raw observation data via the Raw Sensor API;
- Step 2 – ETL for refinement: for data validation, calibration transformations and aggregations of the raw observation data; rendering 'refined' data with metadata;
- Step 3 – Publication: ETL for publishing to various services, some having internal data stores:
    - SOS ETL: transform and publish to the SOS DB via SOS-Transactional (SOS-T);
    - SensorThings ETL: transform and publish to the SensorThings API (STA) DB (via REST);
    - Direct publication in WMS (with SLDs) and WFS;
    - other ETL for custom services or FIWARE APIs (to be implemented).

Some additional notes for the data flows above and software used:

- The central datastore DB is PostgreSQL (www.postgresql.org) with PostGIS enabled;
- All ETL transformations are executed with Stetl, streaming ETL, a lightweight ETL framework for geospatial data conversion (github.com/geopython/stetl);
- The three ETL steps run continuously via Linux cronjobs;
- Each ETL process applies 'progress tracking' by maintaining persistent checkpoint data. Consequently, a process always knows where to resume, even after its (cron)job has been stopped or cancelled. All processes can even be replayed from 'time zero'.
- Refined O&M data can be directly used for WMS and WFS services via GeoServer using SLDs and the PostGIS datastore with selection VIEWs, e.g. last values of components by WMS time dimension (WMS-T);
- The SOS ETL process transforms refined data to SOS Observations and publishes these via the SOS-T InsertObservation operation. Stations are published once via the InsertSensor operation of the 52°North SOS server (www.52north.org);
- Publication to the SensorThings Server will go via a REST service. The SensorThings Server used is offered by SensorUp (www.sensorup.com);

The Smart Emission data infrastructure can be considered a spatial data infrastructure approach using geospatial standards to expose spatial sensor data to the Internet. Although the search for sensors and sensor data through sensor network metadata is not yet addressed, all data is exposed to the Internet for re-use by international OGC-based standards, like WMS, WFS, SOS and the latest SensorThings API (www.ogc.org). Other non-geospatial standards are considered as well. A multi-API approach offers a rich possibility of re-use of the environmental data to be explored in different environmental application fields and re-used by different developers communities, like GIS, Internet of Things and Web developers.

V. CLOUD DEPLOYMENT ARCHITECTURE

The cloud deployment architecture described above will be deployed on the FIWARE Platform provided by the FIWARE Lab NL (fiware-lab.nl). The FIWARE Lab NL offers a PAAS-based computing and storage cloud where instances for common images like Ubuntu can be created, provisioned (e.g. storage, networking, central processing unit), and deployed. Components from the Smart Emission data infrastructure as described in the architecture above will be deployed on the FIWARE cloud using Docker (www.docker.com). Docker is a common computing container technology also used extensively within FIWARE. By using Docker, we can create reusable high-level components, 'containers', which can be built and run within multiple contexts. Figure 4 sketches the Docker deployment strategy.

---

[2] The arrows represent the flow of data; circles depict harvesting/ETL processes; server instances are in rectangles and data stores are represented by the DB icons.

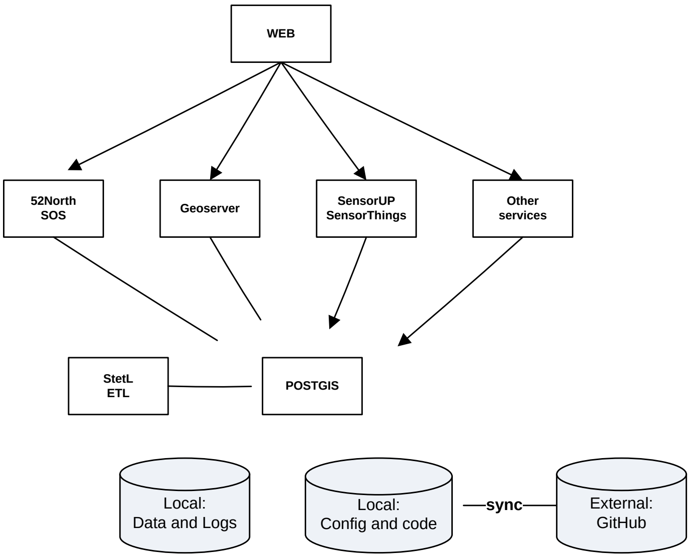

Fig. 4. Docker deployment strategy[3]

In a first instance, Docker containers will be created for:

1. Web: front-end web serving (viewers/apps) and proxy to backend web APIs;
2. 5°2North SOS: container with Tomcat running 52°North SOS;
3. GeoServer: container with Tomcat running GeoServer for WMS en WFS;
4. SensorUp SensorThings: container running SensorUp SensorThings API server;
5. Stetl ETL: container for the Python-based ETL framework used;
6. PostGIS DB: container running PostgreSQL with PostGIS extension.

The networking and linking capabilities of Docker will be applied to link Docker containers, for example to link GeoServer and the other application servers to PostGIS. Docker networking may even be applied, independent of (VM) location. When required, containers may be distributed over VM instances. Another aspect in our Docker-approach is that all data, logging, configuration and custom code/(web)content is maintained 'locally', i.e. outside Docker containers. This will make the Docker containers more reusable and will provide better control, backup, and monitoring facilities. An administrative Docker component is also planned. Code, content and configuration is maintained and synced in and with GitHub. Custom Docker containers will be published to the Docker hub to facilitate immediate reuse.

As a result, FIWARE Lab NL will be used as a cloud-based computing platform. Standard FIWARE components for Internet of Things like Orion may be integrated at a later phase in the project. Also, several Smart Emission Docker containers will be generalized for potential addition to the FIWARE Platform as Generic Enablers (GEs) and included within the FIWARE Catalogue as components for FIWARE blueprints.

VI. OPEN ACCESS AND APPLICATIONS

The data infrastructure of Smart Emission is based on an open approach in many aspects, especially regarding open access to Smart Emission data for re-use and access offered through open, standardized APIs and some out-of-the-box client applications. The data infrastructure is as open as possible, as far as the privacy of the Smart Emission citizens is not violated. Also, the software infrastructure is mostly open source software, as well as documentation[4].

Open access means that all environmental data is open and accessible through web APIs for developers and some client applications for end-users, like students, professional researchers, but also citizens. In order to make access to the sensor data as easy as possible, several client applications are available for data re-use through these adopted APIs. Client applications that provide further processing and exploration are GIS applications (like ArcGIS and QGIS), statistical packages (like R) and out-of-the-box web applications (like the 52°North JavaScript SOS Client Helgoland). One specific client has already been developed during the Smart Emission project, the "SmartApp". SmartApp is a simple end-user Web application that uses the 52°North Sensor Web REST API to explore the last values of measurement in a simple, intuitive map application; all Jose sensor locations are shown on the map and the user can select a location and the last values of air quality and noise load are shown in a popup window (see figure 5).

Many choices are made and small improvements and innovations are developed while the system as a whole is being built and implemented. On the basis of requests from the citizens, a website is being set-up dedicated to the users. This portal serves as visible 'front end' to users and incorporates the data viewers, a forum, and documents exchanged at the citizen meetings: www.smartemission.ruhosting.nl.

Each month, the Smart Emission consortium holds meetings to figure out various aspects of the system's architecture along the way, and to learn from each other's progressing insights and learned lessons over the various parts of the project. On a lower frequency, meetings with citizens are organized. From the feedback by users that is generated in these interactions, the technical architecture and data processing mechanisms are further optimized in order to get a best achievable SDI under the constraints of the current initiative.

---

[3] The entities denote Docker containers, the arrows linking.

[4] All content authored within the project like (ETL) code, viewers, apps, Docker definitions, configurations are maintained in a dedicated project at GitHub: github.com/Geonovum/smartemission. Documentation is also maintained in this GH repo and published automatically on GH commits to smartplatform.readthedocs.io.

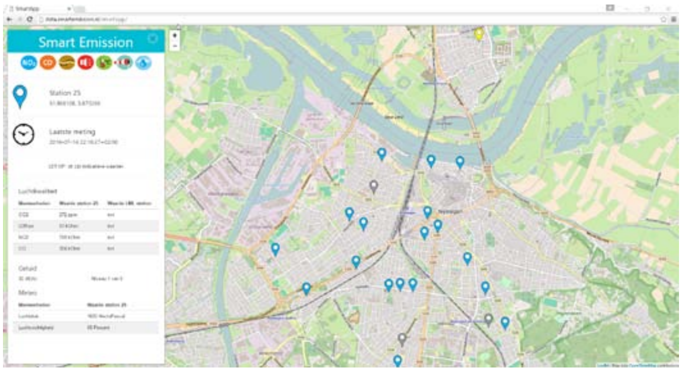

Fig. 5. SmartApp

VII. CHALLENGES AND OUTLOOK

As an innovation initiative, Smart Emission explores several research questions regarding environmental citizens-sensor-networks. However, several questions still remain to be answered:

1. Do low-cost sensors add to the fine-grained picture of air quality indicators? Can we trace an ‘air pollution cloud’ accumulating in certain places in the built environment?
2. Which methods (relatively simple spatial interpolation, spatial regression and visualization techniques) and scenarios can be used for processing and visualizing this data for involving citizens and connecting to broader municipal purposes? Can we combine these measurements with other (modelling) information for informed citizen and government?
3. What about data ownership in citizen-sensor-networks and privacy related issues of using low-cost environmental sensors in cities?
4. Does sense-making with citizens work? What is the citizen science contribution?
5. If the concept works, does this open up opportunities for bottom-up spatial/traffic/urban planning to further improve quality of living and health?
6. Reflective: (How) do roles of government and citizen change? Central elements in the research questions are the notion of ‘fine-grained’ constellation, tracing unevenly spread accumulations or ‘pollution clouds’, and ‘collective sense-making’.

These questions need further exploration and attention. The Smart Emission initiative is still in its infancy, working on the citizens network, the collection (and calibration) of sensor data, the distribution of data by APIs and the search for smart applications.

Smart Emission aims to connect the retrieved data flow to other data sources and embed this information in the dynamic process of city governance. The foremost important challenge in the next period is to stimulate the application of the Smart Emission data infrastructure. Gaining better insight in sensing, individual environmental issues, and in general citizen involvement, commitment and corporation is valuable in itself. Besides of course the experiences gained from the citizen’s network and their environmental situation, the Smart Emission data infrastructure also has the ambition to explore the role and potential of low-cost sensing to heterogeneous application fields. Several potential application areas exist, especially regarding investigating relationships between environmental factors and health problems, like air pollution mapping, noise mapping and heat stress mapping. In addition to health related applications, the Smart Emission low-cost sensing for spatial planning and climate adaptation purposes is also worth considering in more detail.

ACKNOWLEDGMENT

We would like to thank all members of the Smart Emission consortium: Municipality of Nijmegen, Radboud University, Geonovum, Intemo, CityGIS, National Institute for Public Health and the Environment. We would also like to thank SensorUp, in particular Steve Liang for providing the SensorThings Software.

REFERENCES

[1] M. Swan, “Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0,” Journal of Sensor and Actuator Networks, vol. 1, Issue 3, pp. 217-253, 2012.

[2] L.J.Carton, P. Ache and consortium partners, 2015, “Filling the feedback gap of place-related ‘externalities’ in smart cities: Empowering citizen-sensor-networks for participatory monitoring and planning for a responsible distribution of urban air quality,” Paper presented at AESOP 2015, Association of European Schools of Planning Annual Congress, Czech Republic, Prague, 13 - 16 July 2015.

[3] C. Gouveia, A. Fonseca, A. Câmara and F. Ferreira, “Promoting the use of environmental data collected by concerned citizens through information and communication technologies,” Journal of Environmental Management, 2004, vol. 71, pp. 135-154.

[4] M. Hacklay,. “Neogeography and the delusion of democratization,” Environment and Planning A, vol. 45, 2013, pp. 55-69.

[5] L.J Carton and P.M. Ache, “Citizen-sensor-networks serving as countervailing power through bottom-up planning: An analysis of how two grassroots alliances creatively use Geographic Information in a networked manner for monitoring environmental externalities,” 2016, unpublished.

[6] P. Healey, “Collaborative Planning: Shaping Places in Fragmented

[7] J.E. Innes, “Information in Communicative Planning”, Journal of the American Planning Association, vol. 64:1, pp. 52-63, 1998.

[8] P.M. Ache, and L.J. Carton, “Smart citizens 4 smart ruimte - het verkennen van vergezichten voor co-creatie van de stad van de toekomst,” in Toevoegen van ruimtelijke kwaliteit. Ruimtelijke kennis voor het Jaar van de Ruimte, W.Salet, R.Vermeulen and R. van der Wouden, R. (ed.), 2015, pp. 124-135.

[9] JOSENE - Joined Sensor Networks. www.intemo.nl (accessed on 1 July 2016).

[10] K. Austen, “Pollution Patrol,” Nature, vol. 517, , 2015, pp. 136-138.

[11] A. Bröring, Remke, A., and Lasnia, D., “SenseBox – A Generic Sensor Platform for the Web of Things,” In: Mobile and Ubiquitous Systems: Computing, Networking, and Services pp. 186-196, 2011, Springer Berlin Heidelberg.

[12] Q. Jiang, F. Kresin, A.K. Bregt, L. Kooistra, E.Pareschi, E. van Putten, Hester Volten, and J. Wesseling, “Citizen Sensing for Improved Urban Environmental Monitoring,” Journal of Sensors, vol. 2016, 2016, 9 pages.

[13] M.C. Bell and F. Galatioto, “Novel wireless pervasive sensor network to improve the understanding of noise in street canyons,” Journal of Applied Acoustics, vol. 74, Issue 1, pp. 169-180, 2013.

[14] Geluidsnet/Sensornet. http://www.sensornet.nl/english/ (accessed on 1 July 2016).

[15] N. Maisonneuve, Stevens, M. and Ochab, B., “Participatory noise pollution monitoring using mobile phones,” Information Polity, vol. 15, pp. 51–71, 2010.

[16] S. Bell, D. Cornford, and L. Bastin, “The state of automated amateur weather observations,” Weather, vol. 68, no. 2, pp. 36–41, 2013.

[17] Open Seismic Sensor Grid Groningen (OSSG). http://www.ossg.nl/ (accessed on 1 July 2016).

[18] E.S. Cochran, Lawrence, J.F., Christensen, C and Jakka, R.S., “The quake-catcher network: Citizen science expanding seismic horizons.” Seismological Research Letters, vol. 80, 2009, pp. 26–30.

[19] T. Usländer, Berre, A. J., Granell, C., Havlik, D., Lorenzo, J., Sabeur, Z. and Modafferi, S., “The future internet enablement of the environment information space,” In: Environmental Software Systems. Fostering Information Sharing, pp. 109-120, 2013. Springer Berlin Heidelberg.

[20] L. Spinellea, M. Gerbolesa, M. G. Villanib, M. Aleixandrec and F. Bonavitacolad, “Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide,” Sensors and Actuators B: Chemical, vol. 215, August 2015, pp. 249–257.

[21] A. Kotsev, O. Peeters, P. Smits and M. Grothe, “Building bridges: experiences and lessons learned from the implementation of INSPIRE and e-reporting of air quality data in Europe,” Earth Science Informatics, vol. 8, Issue 2, June 2015, pp. 353–365.

[22] [Schleidt, K., J, Hřebíček, G. Schimak, M. Kubásek, A.E. Rizzoli, “INSPIREd Air Quality Reporting,” In: Proceedings of the Environmental Software Systems. Fostering Information Sharing: 10th IFIP WG 5.11 International Symposium, ISESS 2013, pp 439- 450, 2013, Springer Berlin Heidelberg.