CITYZER observation network and data delivery system

Walter Schmidt¹, Ari-Matti Harri¹, Timo Nousiainen¹, Harri Hohti¹, Lasse Johansson¹, Olli Ojanperä², Erkki Viitala³, Jarkko Niemi⁴, Jani Turpeinen⁵, Erkka Saukko⁶, Topi Rönkkö⁷, and Pekka Lahti⁸

¹Finnish Meteorological Institute, Erik Palménin aukio 1, 00560 Helsinki, Finland
²Vaisala Ltd, Vanha Nurmiäärventie 21, 01670 Vantaa, Finland
³Emtele Ltd, Tampere, Finland
⁴HSY Helsinki, Ilmalantori 1, 00240 Helsinki, Finland
⁵Sasken Finland Oy, Vissavedentie 1, 69600 Kaustinen, Finland
⁶Pegasor Ltd, Hatanpään valtatie 34 C, 33100 Tampere, Finland
⁷Tampere University of Technology, Kalevantie 4, 33100 Tampere, Finland
⁸Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland

Correspondence: Timo Nousiainen (timouusiainen@fmi.fi)

Received: 29 March 2020 – Discussion started: 16 April 2020
Revised: 30 June 2020 – Accepted: 21 July 2020 – Published: 13 October 2020

Abstract. CITYZER develops new digital services and products to support decision-making processes related to weather and air quality in cities. This includes, for example, early warnings and forecasts (0–24 h), which allow for avoiding weather-related accidents, mitigate human distress and costs from weather-related damage and bad air quality, and generally improve the resilience and safety of the society. The project takes advantage of the latest scientific know-how and directly exploits the expertise obtained from earlier projects. Central to the project is the Observation Network Manager NM10 developed by Vaisala, on which CITYZER defines and builds new commercial services and connects new sensor networks, for example, for air quality measurements, as well as the ENFUSER local-scale air quality modelling system developed by the Finnish Meteorological Institute, for real-time air quality forecasts and nowcasts.

1 The CITYZER ecosystem concept

1.1 Background

The atmosphere plays a central role in the global circulation of heat, water and volatiles, making our current life possible and convenient. It is the reservoir of the breathable gases and takes away the exhaled gases that would otherwise poison our immediate surroundings. Our civilization is closely linked with the atmosphere in a multitude of ways, whether we are considering simple agricultural communities or modern urban areas.

The CITYZER project concentrates on two closely linked aspects of the atmosphere with the focus on urban climate. On the one hand, we consider air quality, which is highly important in particular due to its possible adverse health effects. On the other hand, we consider weather phenomena, in particular those linked with precipitating weather systems. Weather events can cause episodes of high pollutant concentrations in urban areas, or they can help clean the urban air through washing and wet deposition or by bringing in cleaner air from adjacent areas. In their own right, the weather effects can be hazardous, causing infrastructural damages and loss of lives, or they may be bothersome and cause harmful friction for logistics planning and the everyday life of human society. Collecting sensor data in near real time from the area of interest, sophisticated interpolation and forecast software provides reliable information for the next hours concerning air quality and weather phenomena, which can be made available to the public or service providers via mobile application or warning systems (Harri et al., 2018; Smart & Clean, 2020).

1.2 CITYZER ecosystem

The main objective of the project was to develop new digital services and products to support decision-making processes...
related to weather and air quality. The overall idea was to create a new ecosystem of services open to third-party developers, based on open data. To that end, the project set out to design and implement an IoT-based platform for collecting, refining and delivering environmental data. Accordingly, the project was closely linked to several megatrends, namely open data, big data, platform economy and digitalization.

The CITYZER ecosystem consists of three levels:

1. Measurement data and observation level contain the data collection and management systems like sensor networks and public or restricted data repositories.

2. Diagnostics and modelling level is where data are collected, analysed and prepared for user-friendly utilization.

3. Services and product levels utilize subsets of the raw or refined data, possibly in combination with external data sources like map service data, to present information in a form and updated frequency suitable for the user, like authorities or the public.

This article concentrates on the first and the third level, leaving the details of the analysis forecast software for weather and air quality to other publications.

2 CITYZER system architecture

2.1 General structure

The system is very modular with well-defined interfaces allowing the replacement, removal or addition of different elements and their implementation in different environments from commercial hardware configurations to virtual cloud-based networks.

The six main modules of the CITYZER system architecture are presented in Fig. 2.

1. Observation networks consist of various sensors for all observation parameters to be included into the specific deployment together with the field data networks for forwarding the measurement data and controlling the sensors from remote supervising stations where applicable.

2. Observations and device management and storage include automatic or manual sensor supervision, data collection and possibly data re-formatting services.

3. CITYZER control system contains data storage covering about 1 week of sensor and modelling data and the control module which fetches new data once available and controls the forecasting software.

4. External data sources from public data providers can be accessed to utilize additional weather-related data.

5. One or several application data servers provide a standardized interface between the stored data and external application providers.

6. One or several application service providers allow users to access the information in a user-friendly way from stationary or mobile devices. The service providers might additionally access other data sources like map services to render the data in a user-optimized way.
2.2 Forecast software

Depending on the needs or the deployment one or several forecast model software packages can be implemented. These might reside inside the same physical computer system as the control system or could be implemented on different platforms and linked to the control system via file-sharing technology. In the course of the project the precipitation nowcasting software for rainfall was installed and operated on the same Linux-based server as the control module. It was adapted from the modelling software developed at the Finnish Meteorological Institute as part of the RAVAKE project (Heinonen et al., 2013). The source code size was 205 kB with a size of the executable file of 180 kB. A typical run time for one complete weather forecast covering the larger Helsinki area is about 8 to 13 s. A motion vector analysis for the complete area of Finland takes about 1 min.

The air quality forecast modelling software ENFUSER (ENFUSER, 2020; Johansson et al., 2015) was developed into an operational modelling system with test implementations for the CITYZER demonstrator in a virtual Linux environment, as cloud service and on a local Windows computer. It includes detailed treatment of traffic emissions for individual roads, shipping emissions and elevated point sources such as power plants taking into account urban morphology, atmospheric stability and rain forecast information.

2.3 Data interfaces

Sensor data can be provided in two different ways:

1. Public data are accessible via standard web interfaces. The CITYZER environment implements the Open Geospatial Consortium’s Web Feature Service (OGC/WFS) as described in the organization’s document OGC 04-094, (OGC, 2020) reference list. Most European and many worldwide weather service data are available using this standard.

2. Dedicated sensor networks cover the area of interest. These might be owned by different authorities and private providers and will usually need specialized hardware interfaces and protocol translation software before they can be integrated into the CITYZER data storage module. Various solutions as deployed in the Helsinki area are described below. They are based on a network manager either integrated with a group of sensors or provided separately as a network controller, which then uses the same OGC protocol as the public data for feeding the real-time data into the data storage module.

The sensor data are collected in the data storage in a unified format and sorted according to data type as grid data or coordinate-based point data, day of observation and sensor providing the data.

The analysis and forecast software modules are ingesting the data needed for the forecast interval, adapting to the extended or reduced availability of sensors automatically. The forecast results are written back into the same data storage using a format compatible with the sensor data but organized into different files for easy access.

A database system keeps track of the available data and provides this information on request to user applications. These may be mobile applications allowing the public to assess the weather and air quality situation in the near future in an area of interest. The data related to the inquiry are then provided as links to the respective file in the data storage for download by the application server. This usually will also access geographic data from external providers as a platform.
System security considerations

One critical aspect of any multi-interface system is the system and data security. Most of the sensor systems deployed in the field can usually be compromised without excessive efforts. The same is true of interfaces between service providers and end-user application systems like mobile phones. The CITYZER system architecture therefore provides hardware solutions to protecting the central data storage and forecast software resources. Any connection for incoming sensor data or requests from application servers is separated by a strict firewall without any external write access possibility. The control system sends information requests to each connected data provider at regular intervals. When new data are available it fetches these data, filters them through incoming filter software and writes the possibly reformatted data into its internal data storage, thereby eliminating any illegal data or possible command instructions.

The availability of new sensor or model output data is indicated to the application database outside the firewall, providing a complete link to these data. All data storage files are accessible by the application database software via a hardware read-only link without any write-back possibility. In the case of a compromised database, it can be re-built on the fly from the data inside the secure data storage.

Access to the application data server is controlled via an authentication mechanism. Each possible user has to register first with this server, indicating also the type of access intended: data polling and fetching by the application or automatic push service according to rules to be defined by the user service like update time, geographical boundary parameters to be provided etc.

Data flow control

In the example diagram, Fig. 3, one set of air quality sensors from Pegasor is connected via a public phone data network or via an Emtele-provided LoRa (LoRa Alliance, 2020) sub-network to Vaisala Observation Network Manager NM10. Another set of air quality sensors, AQT400 series from Vaisala, is connected via a Vaisala Beacon View data access and control system to the NM10. Weather radar
data are collected separately via the common Vaisala IRIS system for map generation. The connection between NM10 and the second layer control system is realized as Open Geospatial Consortiums’ Web Feature Service (OGC/WFS) to make it as universally adaptable as possible. Also other public data servers are connected to the control system via OGC/WFS compatible communication standards. The abbreviations used in the flow diagram are OD for open data, AO for air quality observations, WO for weather observations, MD for model data and CRC for command flow from the run control module.

The control unit synchronizes all activities in the CITYZER system. Based on regular timing or action events, different processes are started, data are collected or the database is updated, thereby indicating to the application processes that new data are available.

The initial configuration is defined as follows – see the timing information in the leftmost column in Fig. 4.

Once a minute requests are sent via OGC protocol to the attached and registered data providers to check the availability of new data. In the case of new data these are fetched, converted if necessary and stored in the respective file system of the server according to type, sensor and time of observation. All incoming data are expected to be calibrated in standard physical units. The data sets usually also contain quality information indicating whether the related sensors were correctly calibrated or offline. Such an update is also indicated to the database system to make observational data directly available to the users, possibly initiating an alert to the application servers if so requested.

Every 5 min new weather radar data are fetched and stored in the file system, and the rain nowcasting model is activated. Its results are available within a minute and stored in a different part of the file system. This information is also forwarded to the database system. At a special resolution of 250 m the size of one radar composite file varies between 50 and 1000 kB depending on the amount and type of observed precipitation. A forecast output motion vector field covering the whole area of Finland has a typical size of 60 kB.

Every 60 min the air quality data modelling system is activated providing new hourly and 1-day forecasts of the air quality development in the covered area. These forecasts both make use of fresh observational data as newly calculated rain nowcast results anticipating air quality modifications by imminent rainfall. One typical model run for the Helsinki area takes about 35 min and generates maps with a grid size of 13 m × 13 m. These approximately 3 million grid cells contain concentration estimates for NO₂ and O₃, dust concentrations PM₁₀ and PM₂.₅, and air quality index (AQI) estimates. For each forecast about 1000 local air quality measurements are ingested and used besides the hourly

Figure 4. CITYZER data and command flow control.
10 km resolution weather forecast maps and the 100 m resolution/5 min time resolution rain nowcasting data.

Alternatively any analysis software package might run autonomously at certain time intervals accessing those data available at the start of a new activation.

The output of each control script is logged. Figure 4 shows the various data flow paths, where the solid arrows indicate the flow of data between the different modules, and the dashed lines represent command flows for synchronization of the different modules and control of the attached sensors where possible and needed.

In the demonstration implementation the central control module is implemented as a set of scripts activated according to specific timing rules. A typical control and time diagram is shown in Fig 4.

2.6 User application access

User applications register usually with an application service provider which provides a downloadable application for displaying the requested data and a service which generates these data according to the requirements of the end user. The application server has to register with the CITYZER database to gain access to the data. At the time of registration the strategy of accessing new data can be defined either as notification to get alerted whenever data relevant for the application are updated or as polling, where the application requests data matching certain criteria, and the database provides these data if available. In both cases the returned information provides a complete data link address from where a file with the related data can be retrieved. This scheme is shown in Fig. 5.

The system is designed to support also the generation of automatic alerts. If one or several parameters exceed pre-defined limits, an alert message can be generated and sent to the connected mobile phone or user service centre. This feature is currently implemented in the Helsinki area to reschedule the street cleaning services in case microdust levels exceed pre-defined levels on major streets before they become a health hazard.

2.7 Data storage structure

The rolling storage of data is divided into two separate file systems:

- Local storage for incoming and cache-type data is on a local disc. The storage size of the demo version is 8 GB.

- Shared storage for output data is in an external file system, which is mounted writable for internal use and read-only for the CITYZER database server. The shared storage directory is physically mounted on an externally visible mount point. Its absolute path must be used by the application server software when sending notifications of new data to the CITYZER database server in JavaScript Object Notation format (JSON, 2020). The storage size of the demo version is 120 GB.

Radar data for rain nowcasting are stored in the ODIM HDF5 format (ODIM, 2020), developed by the European
Meteorological Service Network EUMETNET. Data originating from the various air quality sensors and air quality model data for single geographic points are stored in the XML-based GML format (GML, 2020), also endorsed by the INSPIRE directive. Coordinate grid data are stored in NetCDF format (NetCDF, 2020).

In the local storage there is a directory “Today”, which is linked every day at midnight to the actual directory of the new day (Mon, Tue, Wed etc.). Each daily directory is structured as follows:

- Nowcasting contains data for the rain nowcasting model RAVAKE.
- Sub-directories contain data in ODIM HDF5 format:
  - /Rain incoming radar reflectivity composite files
  - /Vectors incoming motion vector fields of radar composites
  - /Probability output rain accumulation exceedance probability fields
- /Ravake output rain intensity fields from rain nowcasting.
- Files *.dat contain ensemble member data and are used only by the exceedance probability analyser.
- Files *.h5 are in ODIM HDF5 format and contain deterministic rain rate nowcast data.
- ENFUSER contains model grid and point observation data for ENFUSER AQ model input:
  - /grid_data
  - /HIRLAM incoming NWP model HIRLAM data in NetCDF format
  - /SILAM incoming AQ model SILAM data in NetCDF format
  - /point_data
  - /FMI-AQ incoming AQ observation data from FMI stations in GML format

https://doi.org/10.5194/gi-9-397-2020
The Linux cronab-based data fetch process of the run control module copies and links new files from these directories for notification and fetching for CITYZER database server. The fetch process checks the status of dedicated log files in the shared output directories at defined time intervals (currently 10 min) to copy new data files to storage and link them for notification.

3 CITYZER ecosystem implementations

3.1 CITYZER demonstration implementation

As verification of the CITYZER ecosystem concept, all components were implemented in the Helsinki area (see also Fig. 3). Existing air quality sensor networks were augmented by additional sensors and their various data connections. Real-time data from the Finnish weather radar network were used as input for the rain forecast software. The data server, together with the modelling software and the application database system, was built using resources of the Finnish Meteorological institute. A dedicated application for web access and mobile phones was developed showing the forecast information for rain and air quality parameters projected onto a map of the larger Helsinki area and allowing also the specification of automatic alerts in case a specified parameter exceeds a pre-set limit. In preparation for the deployment of the demonstration system, several air quality measurement campaigns were executed (e.g. Hietikko et al., 2018; Järvinen et al., 2019; Kuula et al., 2019; Teinilä et al., 2019). An example of an air quality forecast map for the larger Helsinki area is shown in Fig. 6. The corresponding rain forecast map for the same area can be seen in Fig. 7, while the application was configured to show the probability of light rain. Alternatives

Figure 7. Rain forecast for the Helsinki area © OpenStreetMap contributors 2019, distributed under a Creative Commons BY-SA License.
for medium and heavy rain are also offered. An additional in-built service allows the end user to receive an alert via mobile phone when the probability of rain at a user-specified point exceeds a given level.

3.2 HAQT (HELSINKI metropolitan Air Quality Testbed)

A subset of the CITYZER ecosystem concentrating on the forecasting of air quality parameters was developed into an operative service, implemented and maintained by the Finnish Meteorological Institute. It started its routine operation in the beginning of 2019 and is among other applications used by the city of Helsinki to optimize the deployment of street maintenance and cleaning resources to keep the dust level along traffic routes at a low level (HAQT, 2020). Updated air quality forecasts for the Helsinki region are also shown routinely on the electronic advertisement panels in the Helsinki metro trains (HSY Air quality map, 2020).

3.3 NAQT (Nanjing AQ Testbed)

A modified version of the HAQT environment was developed and deployed in the Nanjing region of China; see Harri et al. (2018).

3.4 HOPE (Healthy Outdoor Premises for Everyone)

The main purpose of the EU-funded project HOPE is to empower citizens to develop their own districts. The project focuses on three different districts with varying air quality challenges: one district with heavy port-related traffic, another with street canyons and with over 40,000 daily vehicles, and the third district which is affected by main roads and wood burning at homes. The change that the project wants to achieve is that the citizens will find air quality issues easily relatable and understandable by creating a feedback loop between high-resolution hyperlocal air quality data and actions of individuals and communities (HOPE, 2020).

3.5 Haaga-Helia student-generated applications

During the CITYZER project, students of the Haaga-Helia University of Applied Sciences in Helsinki were asked to develop mobile applications for air quality and rain forecast monitoring based on the CITYZER demonstrator implementation. Using an in-house developed learning ecosystem for business information technology, the students successfully demonstrated that new service concepts and mobile applications can be easily developed and linked to the data services provided by the CITYZER environment.

Data availability. The data used to generate Figs. 6 and 7 of this article are available on request from the data archive of the Finnish Meteorological Institute via link http://data.fmi.fi/, last access: 12 October 2020. Access instructions are given on the web page http://en.ilmatieteenlaitos.fi/open-data-manual, last access: 12 October 2020.

Author contributions. WS was part of the project administration team responsible for the methodology used. He wrote the original draft of the article and generated the figures for the text. AMH was the leader of the project responsible for project conceptualization, funding acquisition and supervision. TN was the project manager, wrote parts of the article, and reviewed and edited it in its various stages. HH designed and implemented the CITYZER control system and related software and wrote the draft version of the related article sections. LJ designed and implemented the air quality software ENFUSER and wrote the draft version of the related article section. OO was part of the project administration team, participated in developing the methodology of the CITYZER system and provided the network management resources. EV and ES provided sensor resources and input to the related article section. JT designed and implemented the user interface and database part of the system and wrote the draft version of the related section. His software was used to generate the air quality and rain data (Figs 6 and 7 of this article). TR was responsible for supporting air quality field studies. PL coordinated the worldwide applications of the CITYZER ecosystem and wrote the draft for the related sections.

Competing interests. Authors Ari-Matti Harri and Walter Schmidt are members of the editorial board of GI as associate editors.

Acknowledgements. The CITYZER research project was funded by Business Finland (Business Finland, 2020), grant no. Dnro 3021/31/2015, and contributions of the project consortium partners Finnish Meteorological Institute (FMI, 2020), Tampere University (Tampere University, 2020), Haaga-Helia University of Applied Sciences (Haaga-Helia, 2020), Vaisala Oyj (Vaisala, 2020), Pegasor Oy (Pegasor, 2020), Sasken Finland Oy (Sasken, 2020), Emtele Oy (Emtele, 2020) and Helsinki Region Environmental Services Authority (HSY, 2020).

Financial support. This research has been supported by Business Finland (grant no. Dnro 3021/31/2015).

Review statement. This paper was edited by Jean Dumoulin and reviewed by two anonymous referees.

References


https://doi.org/10.5194/gi-9-397-2020


Sasken Finland Oy: available at: https://www.sasken.fi/, last access 29 January 2020


Vaisala Oyj: available at: https://www.vaisala.com/en, last access: 29 January 2020