IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Development of computational technologies and software for assessment of air pollution in Krasnoyarsk

To cite this article: O E Yakubailik *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **211** 012080

Development of computational technologies and software for assessment of air pollution in Krasnoyarsk

O E Yakubailik[1,2], A A Kadochnikov[1], and A V Tokarev[1]

[1] Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Science, Russia, 660036, Krasnoyarsk, Akademgorodok 50/44.
[2] Siberian Federal University, Russia, 660041, Krasnoyarsk, Svobodny pr., 79.

This study was performed with financial support of RFBR and the Government of the Krasnoyarsk Territory, research project no. 18-416-242002

Abstract. Experience of research and development of hardware and software for air pollution monitoring of industrial cities carried out at the Institute of Computational Modeling of the Siberian Branch of the Russian Academy of Sciences is presented. The functional possibilities, architectural features, and user interface of a distributed web geographic information system of ecological monitoring of the city of Krasnoyarsk, which is being developed in a service-oriented architecture, are discussed. Data from automated posts for observing the atmospheric air are collected. The services developed for data aggregation and information web presentation in the form of interactive graphs and thematic maps are discussed. An original device has been developed for measuring fine dust (PM2.5) pollution in the atmosphere to feed the data in real time to the system via a cellular network. A system for integral assessment of pollution in the atmosphere based on Air Quality Index (AQI) calculations has been created.

1. Introduction

Systems for monitoring of the atmospheric air pollution are developed and introduced in many cities in Russia and abroad. For example, in Kaluga region a system for analyzing and managing the air atmospheric air quality is operational, and its information support is based on the “Vozdukh-Gorod” original analytical software [1]. The Novosibirsk-based TION Company formed a network of 16 small stations with sensors that continuously measure the concentration of aerosol particles PM2.5 and PM10, as well as humidity, temperature, and atmospheric pressure. Data from this system is available online [2]. The automated system for monitoring of the atmospheric air developed in St. Petersburg regularly provides information on the air in the city to the authorities and other professional organizations [3]. A hardware-software system for estimating air pollution was developed in the city of Tula. This system collects information by various types of sensors, uses different engineering devices, carries out comparative analysis of the data obtained, and solves the most important problems, such as determination of the contribution of particular facilities to the atmospheric air pollution at a certain moment of time [4].

It should be noted that there is a German system for monitoring the atmosphere, which is based on web interfaces and services. This system was developed by the European Union in accordance with the specifications of the INSPIRE spatial data infrastructure [5]. Romanian researches developed a distributed system for atmospheric pollution monitoring by PM2.5 fine dust with a web interface on the basis of a geographic information system (GIS) [6]. Chinese specialists worked on special means for air pollution monitoring, such as the TaskML script language and a Sensor Web system for processing and analyzing data from atmospheric pollution sensors [7]. Italian developers created a system for monitoring of the air quality and traffic streams. Self-made sensors of this system were created on the basis of an Arduino microcontroller platform, and the results are published on a web portal [8]. The absolute trend today is the introduction of web-based solutions based on a distributed network of cheap sensors.

Monitoring of the state of atmospheric air pollution in Krasnoyarsk is carried out by a number of organizations of federal and regional levels. Each of them has its own methods, technologies, and systems for collecting, storing, and processing data. Recently some public environmental organizations, independent activists, and bloggers have started their participation in the collection of information on the pollution level. The diversity of the solutions being used, interdepartmental and organizational dissociation have lead to the fact that comprehensive analysis and rapid assessment of the entire array of recorded information is now technically difficult and is practically not carried out. The data collected are often published too late and are not easily understood by the general public.

Research and development in the field of air pollution monitoring conducted at the Institute of Computational Modelling SB RAS (ICM SB RAS) are aimed at solving the following problems [9]:

- collecting information on the air pollution level of the city coming from various sources into a centralized database;
- development of a geo-information analytical web-based system for air pollution monitoring with the possibility of presenting information on the air pollution level in a simple visual form based on the air quality index;
- creation of a prototype device for measuring the level of fine dust and automated transmission of collected data to the geoportal server.

In the present paper, a computing technology developed for information support of the air pollution evaluation in Krasnoyarsk is presented. To solve the problem under consideration, a subsystem "Operational Monitoring Data" was developed within the framework of the ICM SB RAS geoportal. From a technological point of view, it is a set of specialized web services and web user interfaces, which provide data collection from various sources of information on the state of the atmosphere in the city, such as official automated observation posts, pollution sensors of non-state environmental organizations, and so on.

Web-based data visualization and analysis tools are provided. Assessment of the atmospheric pollution and its representation on the geoportal are based on internationally accepted indicators, such as the Air Quality Index, the operational indicator NowCast, and their modifications. The prototype of the device for measuring the level of pollution of the atmosphere created at ICM SB RAS ensures registration of the concentrations of fine suspended particles PM2.5 in the air and transfer of the received data to the geoportal fully automatically via a GSM cellular network or a wireless Internet connection via WiFi. The development is based on an Arduino Nano microcontroller with a number of additional components.

Suspended particles PM2.5 are one of the main pollutants of the atmosphere, but the federal and regional systems of monitoring of the atmosphere of Krasnoyarsk have a fairly limited capability to measure their concentrations. The emergence of rather cheap PM2.5 sensors provides a potential possibility of deployment, on their basis, of a large number (from tens to hundreds) of devices in the city. As a result, a unique research network for observing atmospheric pollution by fine aerosols can be established. Such a network will help to identify problem city areas and time points in the spatial and temporal distribution of pollution; and can be a good "research basis" for further detailed analysis by certified devices, by relevant environmental organizations.

2. "Operational monitoring data" subsystem

The developed subsystem "Operational Monitoring Data" collects data from various operational observations, including data on air pollution, weather data, data from hydrological stations, etc.

Access to the data is provided using standard HTTP services via the REST Protocol as XML, allowing their use for publication and further analysis in third-party systems. To fill the observation database of the created service, a tool for collecting data from monitoring stations, observations archives, and Internet resources was developed, which includes a set of software modules for various information sources. These data are periodically downloaded from an external source, which can be represented as a flow of information in the text formats txt, csv, json, xml, html, etc. The developed software extracts data from the received necessary information and loads them using the service server API (Figure 1).

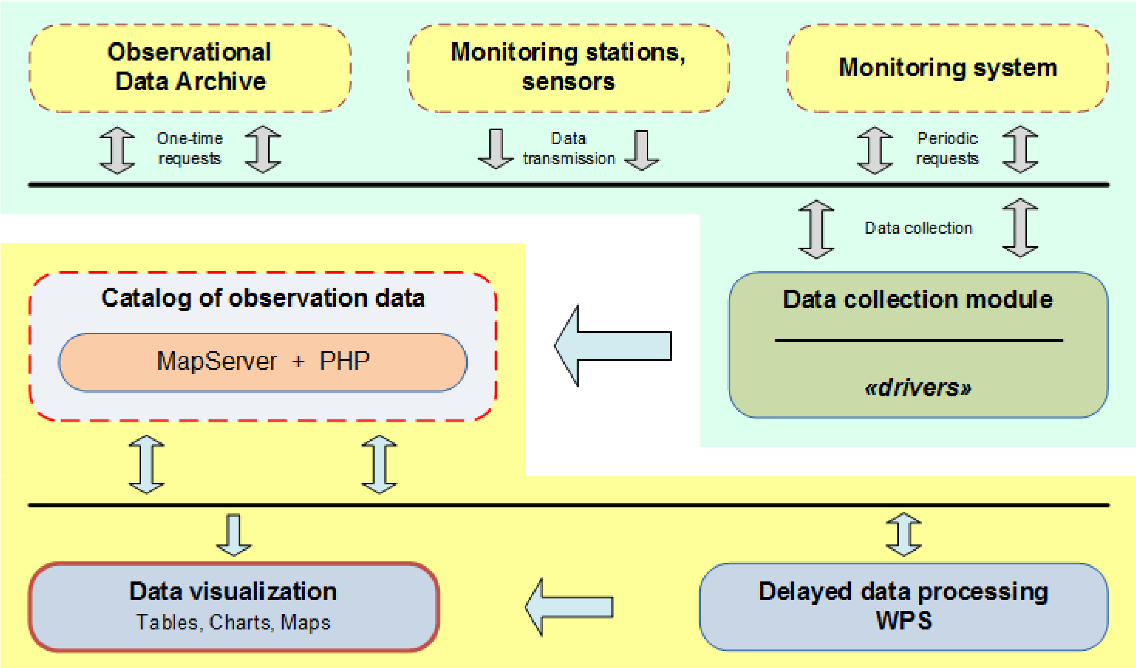

Figure 1. Block diagram of the developed software.

The developed software modules can be divided into several logical groups according to the method of data extraction from external sources:

- The data are presented in the form of web services that transmit information in structured form, for example, in the csv, json, or xml formats. Such data are easy to process and check for errors. A separate problem may be the lack of additional parameters in the web service to filter data by time, sensors, etc. In this case it is necessary to solve the problem of separating new information from that already loaded.
- There are no special web services to obtain data, but there are services of the resource necessary for its operation. For example, services for plotting or output of tabular data at a web interface, tools for data export, etc. Such services can also be used to retrieve data, but it is necessary to develop a number of tools for data analysis, including inspection of changes in the data structure, changing the internal IDs of the format of the output data, the data dimensionality, etc.
- There are no services, and the data are presented in the form of a html page of the Internet resource. To obtain such data, it is necessary to analyze the page in order to find some reference text blocks, with which one can always find the necessary information on the page. In this case, a specific text block is extracted from the page and processed further. In some cases one can work with these page fragments as xml, which makes it easier to retrieve information.

In some cases, a combination of these groups can be used, because the services may require the service information contained on the page itself. Special attention should be given to structural changes in the received data, developing a series of blocks of testing, including changing the number of sensors and monitoring stations, the integrity and completeness of data transferred, etc. For rapid improvement of the software acquisition modules should be created by means of logging text information and alerts, such as e-mail. These tools allow one to receive information about changes to a remote resource and its status constantly. By quickly making corrections to the program code it is possible to ensure the integrity and correctness of the information received.

All data obtained before import into the database are reduced to one unit of measurement, constantly checked for compliance with the data dimensionality (maximum and minimum allowable values, zero and empty values). Checks for missing data and their complete absence. When critical errors in the received data and omissions are detected, a message is generated for the developers. Depending on the configuration of the developed software, in some cases the importing process can be either stopped completely or stopped only for the current sensor, or the data import will be simply skipped for the current sensor. The data time is reduced to one time zone. Depending on the frequency of data updates on the remote service, the frequency of data collection is adjusted. However, the frequency of updating the same data from different sources may vary. The system provides for the collection of data at different time intervals.

The developed software and technological tools are high-level repositories of functions and classes, application templates directed at the end user. These provide an operational solution of the following tasks:

- Maintenance and storage of digital cartographic materials, environmental data sets;
- Navigation through cartographic resources, visualization and analysis of spatially oriented data on unified digital maps;
- Interaction with cartographic and attributive resources (databases) of third party applied information systems;
- Solution to various environmental spatial tasks using resources of the spatial data storage (spatial search, object creation, etc.);
- Provision of access to the system using modern geoinformation system technologies and interfaces.

The developed software can be characterized from the technological point of view by the following:

- The system consists of client and server parts, thereby realizing the "client-server" technology. Application of the MVC (Model-View-Controller) design pattern as the server part basis provides ample opportunities for the solution of the tasks. Using this architecture assumes separation of the application data, user interface, and control logic into three separate components. Modification of each component can be performed independently. Apart from the standard MVC elements, the key units of the system are also page presentation templates and AJAX handlers.
- Spatial data are recorded in the resource catalogue. The web application communication with the resource catalogue is performed on the basis of a service-oriented architecture using the SOAP/XML protocol. It creates a set of functions available in the form of application programming interface (API), which offers the possibility of resource searching, filtering, management, editing, copying, moving, etc.
- Two API sets were formed. The 1-st API is a common (client) interface for user applications, and the 2-nd one is for an advanced (server) interface for applications having the ability to control the resource catalogue, its objects, and relationships between objects.

3. Air Quality Index for air pollution assessment

Air quality is an important environmental factor that determines the health of the population and the state of ecosystems [10, 11]. For integrated assessment of the degree of air pollution in many countries of the world, unified indices are used, such as the Air Quality Index (AQI) [12].

The AQI can be seen as a tool to provide information on air pollution to the general public in a simple and visual manner [13]. The basic idea is that for each substance a scale of pollution levels is formed, consisting of several classes depending on the degree of impact on human health. AQI is calculated from the concentration indices of several pollutants: suspended particles (PM – particulate matter) less than 10 microns in diameter (PM10) and less than 2.5 microns in diameter (PM2.5), CO carbon dioxide, SO_2 sulfur dioxide, NO_2 nitrogen dioxide, and O_3 ozone. For each class of the scale of pollution levels, a color designation is also introduced (green/yellow/red/burgundy/black color means the appropriate degree of air pollution and the impact on human health: from a safe level to a natural disaster), and recommendations are formulated for the population [14].

The experience of using the AQI to assess the current state of the atmosphere in real time revealed several features and problems. The method of its calculation uses 24-hour averaging for the concentration of suspended particles, which leads to inertia of the index at small time intervals of about an hour. The sudden change in the PM2.5/PM10 in the AQI value is significantly different from the observed situation. For example, in Beijing with strong north winds, the air can almost completely clear in less than 30 minutes [15].

For these reasons, the US Environmental Protection Agency (EPA) has developed an alternative NowCast technique that is more suitable for real-time representation of the air quality index for assessment of PM2.5, PM10, and the ozone current levels. It uses an adaptive weighted averaging of hourly data depending on the rate of change of the indicator. Selection of coefficients of this method was adopted in China, taking into account sharper changes in the concentration of suspended particles, the so-called "Asian NowCast". There is also an alternative approach, when no additional averaging of hourly data is performed at all; this is "InstantAQI", and at the moment it is used in the World Air Quality Index project (http://aqicn.org).

4. Monitoring data and associated web mapping interface

The access to observation data is based on the geoportal tools, which include viewing tabular data, exporting, viewing data on maps with the ability to select time intervals, and access using commonly accepted standards. This approach can be considered common [16, 17].

The existing web-based mapping capabilities have been enhanced to view case data. The user of the web application has access to such controls as selection of one of the indicators, the time interval. With the help of additional tools one can view data with a certain time step in one of the directions. To search for anomalies, data is output in the form of an active graph of maximum values with a quick transition to viewing the data at a certain point in time.

The source of information on the atmospheric pollution level is the Departmental Information and Analytical System on the State of the Environment of Krasnoyarsk region, with information being collected since 2009. The atmospheric air quality is observed at six automated observation stations in the city of Krasnoyarsk, at which the equipment is located that provides continuous automatic measurement of the mass concentrations of oxide and nitrogen dioxide, sulfur dioxide, carbon monoxide, dust, and formaldehyde in the atmospheric air. The system collects, processes, stores, and transfers the collected information to a remote server. The concentrations of pollutants in the air are measured along with meteorological parameters (wind direction and velocity, temperature, humidity, and atmospheric pressure). Moreover, in certain neighborhoods of the city of Krasnoyarsk the air pollution (ammonia, hydrogen sulfide, hydrochloride, hydrofluoride, benzapyrene, and suspended substances) is observed everyday with the use of a mobile laboratory (on a sliding schedule).

The information collection system is implemented on the basis of software and hardware of the geoportal of ICM SB RAS. The observation results are accessed by standard means of the geoportal, which includes viewing tabular data, exporting, and viewing data on maps with the ability to choose time intervals. A distinctive feature is a flexible configuration of the visualization and web services of import/export of all available information.

The observation results can be viewed with the help of the existing cartographic web interface of the geoportal. The main elements of the interface are the base background map and the observation data in the form of a semitransparent layer at a given time for a certain index. The users can work with such control elements as the choice of a certain index and a time interval. Additional instruments can be used to view data with a certain time step in one of the directions (Figure 2).

The information is uploaded to the geoportal by means of a special software with a corresponding driver for processing and converting the input data and for periodical uploading of given observations via a web service in the JSON format. This format contains three sections of information: observation posts and their coordinates, a list of indices (pollutants and meteorological data), and time based values of the indices.

The load on the remote server during uploading of the archive of the 2009-2017 observation data is minimized by creating text files in the JSON format. The import of data from the database on the remote server to the geoportal server took several days. After the archive is imported, the data are downloaded once an hour, and the mobile lab data are downloaded every few days. If there are problems with importing data from a remote web service, the developed software automatically finds the date of receipt of the latest data and generates a request to download new ones. The load on the remote service is minimized by downloading the information in portions with a certain time interval until the current time is not achieved in the data.

The geoportal monitoring utility service was developed, which automatically performs recalculation of the air pollution indices (AQI, NowCast, Instant AQI). The result is stored in the geoportal database in the form of derived indicators, which makes it possible to obtain and use them through a standard programming interface. A specialized web application based on the geoportal programming interfaces (APIs) displays the current AQI values for observation posts on the map, the dynamics of the quality index for the last 48 hours. The mechanism of dynamic content generation based on templates was used to design web application screen forms.

The developed software tools allow one to view information in a pop-up window in HTML format when one requests it on a web map with a mouse click. Information can be obtained from the objects of a single layer or a group of layers depending on the current layer visualization settings in the map legend.

Special attention in the software development was given to the means of web data display. In particular, there were several modes of visualization of attribute data. The basic mechanism involves the use of a universal form of information pop-up window, in which all attribute information on the selected object layer is displayed as a normal table, regardless of its content. The alternative method uses a system of templates and external services. To activate it one must specify the need to use a template or an external service in the layer or map metadata, configure the appropriate parameters, and create and save the template to be used in the metadata. The software implementation of the template engine is based on the open source template software TWIG, which provides an effective combination of a static HTML code and the dynamic content. The templates allow one to change the order and form of attribute data for features in the map layers, including different styles for color, font settings, and so on.

The developed software also provides the possibility of using online mapping services for combining pollution monitoring data with third-party background maps. The system has about 40 different base maps and satellite mosaics from such data providers as OpenStreetMap, Google, Yandex, Bing, Rosreestr, GeoMixer, ArcGIS, etc.

Along with the traditional way of using the base map as a background image for thematic geospatial data, the current version of the software also provides the ability to work with special base maps, which are placed on top of the user's thematic data. Such layers usually contain labels of geographical objects, a small number of lines (roads, rivers). In a number of applications the considered functionality has proved to be very relevant.

Freely distributed technologies and software became the technological basis of the development. The main development language is PHP (http://www.php.net). This language has a sufficient set of methods and libraries for analyzing XML, JSON, and CSV files. HTML processing uses XML parsing libraries, as well as common methods for working with text. During data processing, data types, their format, compliance with templates, etc. are checked. The Javascript code and cookie arrays are also analyzed to extract key parameters of the resource, such as encryption keys, session ID, list of sensors and their characteristics. PostgreSQL DBMS is used for data storage (http://www.postgresql.org). It is a freely distributed object-relational database management system, the most developed open DBMS in the world, and it is a real alternative to commercial databases with an additional module for work with PostGIS spatial data (http://www.postgis.org).

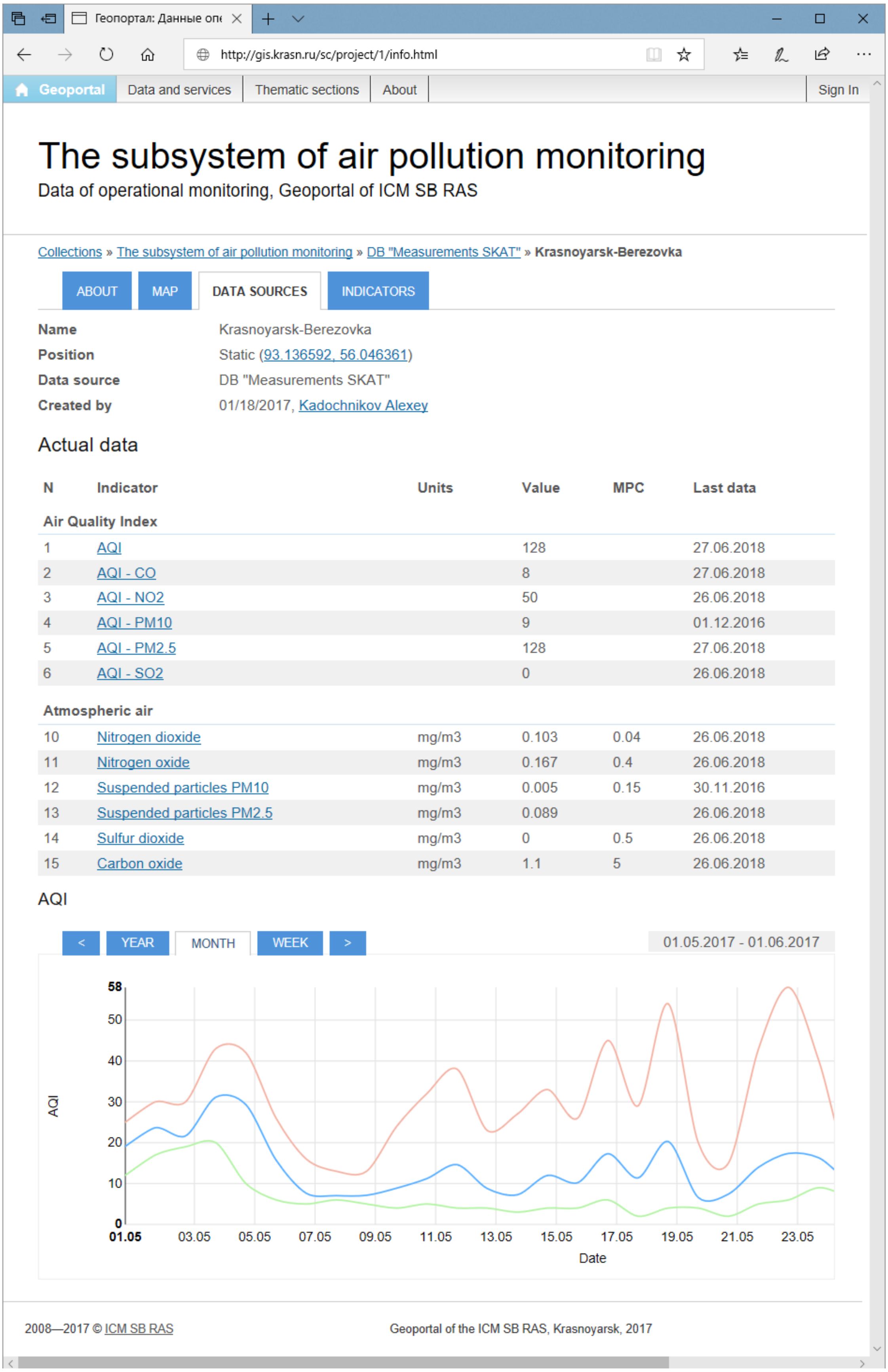

Figure 2. Rapid assessment data at the geoportal of ICM SB RAS.

5. Device for particulate matter concentration measurements

A device for recording the concentrations of fine suspended particles in the air has been designed on the basis of commercially available electronic components. The system hardware is implemented on the basis of the Arduino Nano microcontroller platform.

The main measuring element of the device is the PM7003 module, which is a universal digital sensor for measuring the concentrations of suspended particles in the air (PM1.0, PM2.5, PM10). PM7003 is manufactured by Plantower company [18]. The recorded data are displayed on the screen and automatically transmitted to the "Operational Monitoring Data" subsystem through the GSM cellular network or the WiFi Internet connection.

The sensor PMS7003, according to its developer, ensures the accuracy of measurement of PM $\pm 10\mu g/m^3$ in the range of 0-100 $\mu g/m^3$ and ±10% in the range of 100-500 $\mu g/m^3$. In paper [18] several such sensors were tested, along in one place. The aim of the experiment was to study the accuracy of the sensor readings and the potential level of error caused by the lack of effective calibration of inexpensive sensors. The measurement results confirmed compliance with the specified characteristics.

Similar studies of the measurement accuracy were initiated by the authors in Krasnoyarsk. The data of the developed device were compared with the results of observations at certified city posts of the official services. Preliminary results show a discrepancy in the results in the range of 5-15%. This result provides a basis for further detailed analysis of the capabilities of the equipment in question, given its extremely low cost, and the associated prospects for its widespread use. In the future, it is expected to perform a thorough comparison of the results, calibration of the device.

An internal microprogram for the microcontroller was developed using the Wiring language. There are several modes of operation of the device with different periods of display and data transfer to the server (time of the screen update / data transfer): 5s/off, 5s/20s, 5s/20m, and 20m/20m. In modes with a long period between measurements, there is a temporary shutdown of the internal fan of the dust sensor to save operating life. The application is constructed on the basis of a task scheduler, which makes it possible to unload the main program cycle and make logic in the form of separate tasks with a controlled launch period. At the stage of preliminary study of the capabilities of the selected components, the device is assembled and debugged on a breadboard. Later the prototype is assembled in a 100×60×25 mm compact package. The device needs an external power source connected via a standard microUSB connector (Figure 3).

Also, a web GIS monitoring user interface was designed and developed on the basis of the program interfaces of the geoportal of ICM SB RAS, a set of the previously developed hardware and software. The web application forms are designed by using the mechanism of formation of dynamic content on the basis of templates. The templates can change the order and form of the display of attribute data on the objects in the map layers, including various elements of the style design: color, font parameters, the order of the layers on the map, etc.

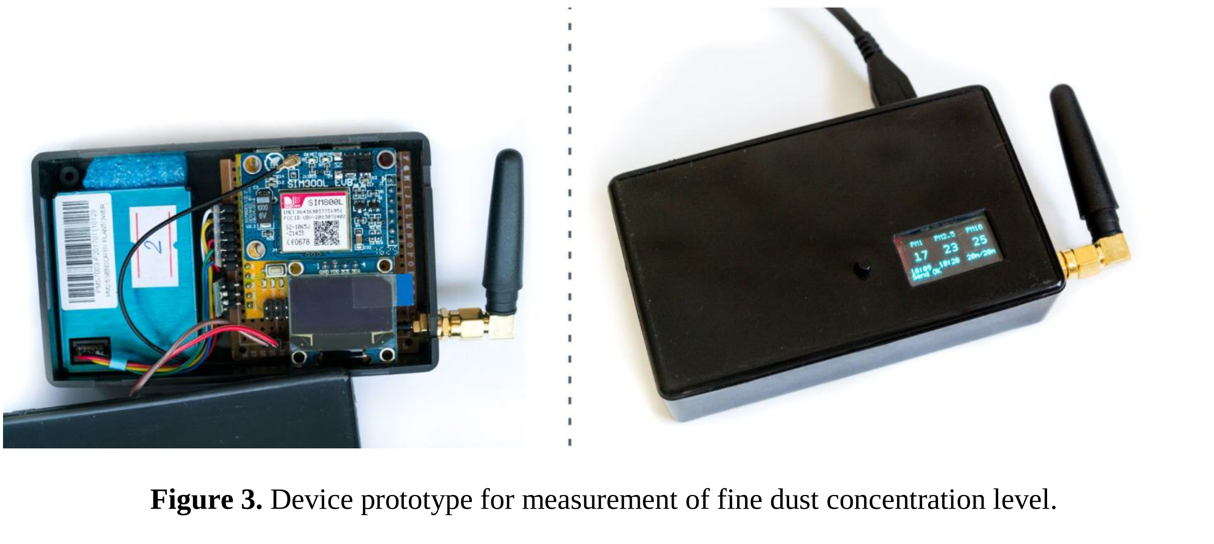

Figure 3. Device prototype for measurement of fine dust concentration level.

6. Conclusions

The above-developed software and technologies solve the problem of automatic collection of information on pollution in the atmosphere of a city which comes from various sources in a centralized database. The available information on the air pollution is presented in a simple visual form based on the Air Quality Index. A working prototype of the device for measuring fine dust air pollution and feeding automatic data into the monitoring system based on the available components was created.

References

[1] Ashitko A and Man'shina I 2014 System for monitoring the state of atmospheric air in the city of Kaluga *Vestn. Kaluzh. Univ.* **22** 5-9

[2] City quality monitoring system CityAir. Available at: http://cityair.ru/.

[3] Azyomov D 2016 System for monitoring atmospheric air of St. Petersburg *Okruzh. Sreda Sankt-Peterburga* **2** 8-14

[4] Karpov V, Panarin V and Goryunkova A 2012 Hardware-software system for estimating the atmospheric air pollution *Izv. Tul. Gos. Univ. Tekhnicheskie Nauki* **2** 173-82

[5] Wiemann S, Brauner J, Karrasch P, Henzen D and Bernard L 2016 Design and prototype of an interoperable online air quality information system *Env. Model. Soft.* **79** 354-66

[6] Savu T, Jugravu B and Dunea D 2017 On the development of a PM2.5 monitoring network for real-time measurements in urban environments *Revista de Chimie* **68** 796-801

[7] Hu L, Yue P, Zhang M, Gong J, Jiang L and Zhang X 2015 Task-oriented sensor web data processing for environmental monitoring *Earth Sci. Inform.* **8** 511-25

[8] Zaldei A *et al.* 2017 An integrated low-cost road traffic and air pollution monitoring platform for next citizen observatories *Trans. Research Procedia* **24** 531-38

[9] Yakubailik O, Kadochnikov A and Tokarev A 2015 Applied software tools and services for rapid web GIS development *15th International Multidisciplinary Scientific GeoConference SGEM 2015*. (SGEM2015 Conference Proceedings, book 2 vol 1) pp 487-94

[10] Gulia S, Nagendra S, Khare M and Khanna I 2015 Urban air quality management – a review *Atm. Pollut. Res.* **6** 286-304

[11] Ghodousi M, Atabi F, Nouri J and Gharagozlou A 2017 Air quality management in Tehran using a multi-dimensional decision support system *Pol. J. Env. Stud.* **26** 593-603

[12] Real-time Air Quality Index (AQI). Available at: // URL: http://aqicn.org/.

[13] Qiu X *et al* 2015 Development of an integrated policy making tool for assessing air quality and human health benefits of air pollution control *Front. Env. Sci. Eng.* **9** 1056-65

[14] Wang H *et al* 2015 Design and demonstration of a next-generation air quality attainment assessment system for PM2.5 and O-3 *J. Env. Sci.* **29** 178-88

[15] Zhu S, Lian X, Liu H, Hu J, Wang Y and Che J 2017 Daily air quality index forecasting with hybrid models: A case in China *Env. Pollut.* **231** 1232-44

[16] Qin R, Lin L, Kuang C, Su T, Mao X and Zhou Y 2017 A GIS-based software for forecasting pollutant drift on coastal water surfaces using fractional Brownian motion: a case study on red tide drift *Env. Model. Soft.* **92** 252-60

[17] Gulliver J *et al* 2015 Development of an open-source road traffic noise model for exposure assessment *Env. Model. Soft.* **74** 183-93

[18] The Plantower PMS5003 and PMS7003 Air Quality Sensor Experiment. Available at: http://aqicn.org/sensor/pms5003-7003/