Providing interoperability for air quality in-situ sensors observations using GML technology

^{1*}A. A. Alesheikh, ¹A. K. Oskouei, ²F. Atabi and ¹H. Helali

¹Department of GIS Engineering, Khajeh Nasir Toosi University of Technology, Tehran, Iran ²Department of Environmental Engineering, Graduate School of the Environment and Energy, Islamic Azad University,

Science and Research Campus, Tehran, Iran

Abstract

Geographic information provides the basis for many types of decisions ranging from simple path finding, to the sustainable management of environmental conditions. Producing such information is a time consuming and costly endeavor. Data sharing on the web is an avenue to increase the efficiency of the practices. This paper scientifically examines the new emerging technologies namely, internet, geographic markup language (GML), and observation and measurement models, to construct an interoperable repository for air quality sensors measurements. The paper also elaborates on the design and implementation of a web-based air quality information system (AQIS) for the city of Tehran. In-situ sensors measure ozone (O₃), carbon monoxide (CO), sulfur dioxide (SO₂), nitrogen dioxide (NO₂) and particulate matters (PM) in polluted metropolitans. Providing real-time air quality information can improve the decisions of the pertinent environmental organizations. Using GML for encoding sensors observations makes it possible to build an interoperable repository that is independent of platforms and vendors. Providing query possibilities based on monitoring stations, sensor names (pollutants), date and time intervals, and spatial query on the AQIS interfaces are the major functions of this system. Although standardized, it was concluded that the use of GML as data format increases the size of GML document. In addition, the developed system provides no map based results for the clients. Therefore, it is required to be improved by adding more GIS functions.

Key words: Interoperability, GML, in-situ Sensors, web-based application, AQIS, AQI *Corresponding Author, E-mail: <u>alesheikh@kntu.ac.ir</u>

Introduction

The ever-increasing population growth in urban areas amplifies certain problems, which directly or indirectly affects human daily life. Air pollution is one of the most important urban problems caused by heavy traffic, and emissions of factories that threatens human health. One of the applied solutions for mitigation and control of this problem is the use of communications, Internet and especially sensor technologies such as sensor web enablement (SWE) supported by open GIS consortium (OGC). The above-mentioned technologies provide facilities to collect, transfer, process, evaluate, and disseminate air quality condition, real-time.

In a routine scenario, the use of these technologies begins with identification of the polluted area (monitoring stations) in order to deploy in-situ sensors that are capable of measuring air pollutants. The most important air pollutants are: O_3 , CO, PM, SO_2 and NO_2 . In the next step, air quality data measured by in-situ sensors are communicated to a central station where data processing is completed. Then the extracted air quality status and related

information is disseminated through internet. The results can then be used by public and involved organizations responsible for the control of air quality.

The crucial issue discussed in this paper is the determination of a data format for encoding in-situ sensors observations in a standard and interoperable method. Data interoperability plays an important role in the Internet GIS. Encoding, storing and transferring the collected in-situ sensors observations using interoperable data format will insure data sharing and reusability in the future. The most important method to achieve interoperability is the use of standards. The leading standards in interoperability for GIS data is OGC that is supported by software vendors (GIS, database, data visualization), system integrators, computer vendors, telecommunications companies, universities, information providers, and federal agencies (De By, et al., 2001). They have joined the consortium to participate in creating software specifications and data transfer strategies, which will assist in solving the non-interoperability problems of GIS data and

software (Charvát, *et al.*, 1998). OGC has introduced several initiatives for sensor web context. Observations and measurement model (O&M) is a general model utilizing extensible markup language (XML) for encoding sensor observations (Cox, *et al.*, 2002).

Two sub-models of observation and measurement model are observation and value models. Observation model describes "observation" as a feature, which is the basic item in the OGC model for geographic information, and then determines relevant properties for that observation. These models have been implemented using XML and GML (Cox *et al.*, 2002).

GML is written in XML format for the modeling, transport and storage of geographic information including spatial and non-spatial properties of geographic features (Percival, 2002). GML is developed as a data exchange interface standard to achieve data interoperability between different systems and is considered as the standardized means of geographic information exchange in XML encoded files. XML and American standard code for information interchange (ASCII) based format uses descriptive tags to store data independent from any proprietary vendor specific formats (Prins, 2003). In general, GML-based data repositories containing in-situ sensors observations have the following advantages to the other alternatives:

- GML has been standardized and accepted by the major GIS software vendors, academia, private and governmental organizations as web data format that can meet all the interoperability needs of the GIS community.
- GML is open and non-proprietary, thus, it can be easily shared. In order to establish data sharing among the data repositories, a data conversion processes is required, which in many cases limits the users.
- GML is based on XML, supported by world wide web consortium (W3C), and inherits all XML characteristics. XML is extensible, flexible and selfdescribing language. The self-describing feature can solve problems of complex data transferring between software applications, in particular over the web.
- The number of softwares and products supporting the GML are increasing. Hence, creating data repositories based on GML will insure the success of data interoperability.
- XML (and thus GML) is a simple text format and

for dealing with GML data repositories, a simple text editor is adequate. This characteristic makes XML platform-independent, since text format can be understood by all platforms.

- GML encoded data can be easily integrated and distributed with other data.
- GML enables the user to take part in the spatial web; otherwise our communication with the world will be blocked.

The following is a case study on building a webbased AQIS and interoperable data repository for air quality in-situ sensors observations using GML technology.

Materials and Methods

Tehran, the capital of Iran, has become one of the most polluted cities in the world. Mobile sources are responsible for more than 70% of Tehran's air pollution. Moreover, 37% of country's industrial units are located at the metropolitan area or its vicinity. Reports indicate that more than 4,500 people are dying every year in Tehran due to air pollution (AQCC, 2005). Efforts are being made, but at best, they may only succeed in holding pollution at current levels.

This research is a step taken towards providing a standard based information system that allows the authorities to tackle the air pollution. The aim is to use the latest technologies and standards for creating an interoperable data repository. At present, the major organization involved is the (AQCC) of Tehran. This company was founded in 1993 by Municipality of Tehran.

In order to improve, control and manage the air pollution in Tehran, AQCC has taken benefits of eleven monitoring stations (Figure 1). These stations cover the areas that are more prone to air pollution. Every station has been equipped with in-situ sensors that measure the amount of O_3 , PM, CO, NO₂, and SO₂ pollutants hourly. After measuring the amount of air pollutants, the data is transferred to a central station in Tehran for further processes. Central station checks the data for errors, and by computing air quality index (AQI) specifies the air condition. AQI is an environmental index that is used to report on air pollution status.

The latest information provided via related insitu sensors and data encoding and processing steps are done through binary and proprietary software.

For these cases, this paper intends to provide and improve present functionalities of AQCC by introducing the development of a system that utilizes GML technology and observation and measurement models which are supported by OGC. In addition, it is desired to develop a system that makes use of document object model (DOM), active sever page (ASP) with VBScript providing dynamic web pages and possibilities for doing query based on in-situ sensors observations encoded in GML. It is believed that the system would facilitate data analysis, interpretation and management of the air pollution. It is obvious that providing a system that is based on Internet, communication and sensor technologies with near real-time information on air condition can be instrumental in tackling air pollution problem.

The following elaborates on the design and implementation phases of this system. The AQI is a composite index computed from ambient levels of CO, SO₂, NO₂, PM and O₃ (USEPA, 2005) which need to develop for Tehran air conditions. In the final step, the air condition based on AQI scales (good, moderate, unhealthful, very unhealthful or hazardous) is reported to the public and decision makers. At present, AQCC makes use of an exclusive website (www.aqcc.org) in order to publish air quality related information via Internet. This web site takes benefit of static web pages containing tables and charts that show the air status based on Figure 1.

Information model

Information model defines model elements and their relationships. Here, the main aim is to develop an information model based on the need analysis, rules and terms defined in observation and measurement models introduced by OGC. This model is developed in a way that meets AQCC requirements, although it can be used for other conditions and required changes. Figures 2 and 3 show the developed information model using unified modeling language (UML) for supporting AQCC requirements. As the diagram shows this model makes use of observation collection as the root element. Observation collection element is used to record heterogeneous observations measured by multiple kinds of sensors. Observation collection consisting of ordered, homogeneously typed observations is an observation array (Cox, et al., 2002). Target property in this model is used to describe stations. Here the station location is meant as sensor location and the relevant result. Based on this model, location element carries sub-element called "point" that contains coordinates element and reference system attribute to record location of monitoring stations. Observable is defined as phenomenon that is subject to observation (Cox, et al., 2002). In this model, an observation is defined to have a result to record the values of measurements, and a time stamp property for recording time and date of measurements.



Figure 1: Tehran map and air pollution monitoring stations distribution



Figure 2: The developed UML class diagram showing AQIS information model



Figure 3: The legend of class diagram

GML document creation

The first step in creation any kind of GML document is the schema document creation. Rich observation schema (introduced by OGC) is the main schema used, which contains all the components to construct GML document elements to record observations and measurements of sensor-based applications. Therefore, GML document of interest is based on aforementioned standard schema supporting the application. The main task to create GML document of AQIS is to identify and use elements required in recording of AQIS observations and measurements based on the model shown in UML diagrams and using rich observation schema.

In this phase XMLSPY software was employed to create the GML schema, and finally GML document. XMLSPY provides easy to use solution. Figure 4 shows the developed GML document for AQIS. The developed document is used to enter the collected in-situ sensors observations.

System architecture

In this section, the design and implementation of

an internet based system that will provide facilities to disseminate air quality information encoded in GML are elaborated. The client-server architecture is selected for implementation. Client-server architecture is a common form of distributed systems where the software is split between client and server tasks (Groot, et al., 2000). In a distributed system, a user does not see which computer is performing, and which one functions. Based on this architecture, a client is a process or computer that requests the services of another computer or process (server) using some rules (protocol). The server processes the request and sends the result back to the client. The exchange of information is done through message exchange according to the protocol (Groot, et al., 2000). Figures 5 and 6 show the developed UML class diagram containing client-side and server-side components for AQIS.

Based on Figure 5, server-side class contains functions supporting the client-side requests. The main functions include:

• Initiate DOM document function, which is used to initialize DOM object that enables system



Figure 4: XMLSPY grid view showing the developed GML document of AQIS using standard schemas

applications access and manipulate GML document.

- Do spatial query function, which is used to do search sensors located in a specified area for a certain date.
- Compute AQI function, which is used to determine AQI and assessment of air quality.
- Do advance search function, which is used to do search based on station, pollutant, date and hour.

As the diagram shows, the server-side search script makes use of air quality data encoded in GML to locate the client request. The, server-side script employs DOM programming techniques to manipulate the GML data. Here, XML parser includes MSXML that is in the Microsoft's Internet Explorer.

In this process, client-side class provides clients facilities to make request and submit it to server-

side scripts to further process. As the diagram shows, three pages (interfaces) have been dedicated to do so (spatial query page, AQI page and advanced search page).

Implementation and user interfaces

ASP with VBScript language, and DOM programming techniques are used to implement the system. In order to support the user requirements different kinds of interfaces are developed. Figure 7 shows the developed main page of AQIS and its options.

Options of the main page of AQIS provide the following facilities for the users:

I. Search by station, pollutant, date and time option, which can be used to obtain air quality information based on the combination of station and pollutant at different dates and hours.



Figure 5: The developed UML class diagram for air quality information system



Figure 6: The legend of UML class diagram

- II. Air quality index (AQI) option, which can be used to get information about air condition. AQI is an environmental index which is computed based on the amount of SO₂, NO₂, PM, and O₃ pollutants.
 III. Spatial query option, which provides facilities to
- users to get air information within a specified area. This area is defined by using relevant coordinates in universal transverse mercator (UTM) map projection.

Figure 8 includes two combo boxes for selecting station and pollutant, two fields to determine the start and end date, and finally two fields to specify time intervals in terms of hour. Providing information for these fields and submitting information to server side script results in further process and sending response page towards the client. Figure 9 contains a text field to specify the date of interest and a combo box for entering the name of station that the user is interested to get information about air status based on AQI.

Figure 10 provides two text fields for specifying date interval and the coordinates of an area based on UTM map projection. Submitting this page towards the server will produce available air quality sensors information that is located within the specified area. To provide a better view of the system functionalities, the following result pages (Figures 11 to 13) are created based on information specified in Figures 8 and 9.

Discussion and Conclusion

The objective of this paper was to discuss the interoperability issue and most importantly, OGC initiatives namely sensor web enablement (SWE) and GML technologies providing interoperability in the network. An application for GML, through the design and implementation of (AQIS) was presented. The main advantage of this system is the use of GML as a data format to encode air quality in-situ sensors observations and measurements. The use of GML makes our system repository, most importantly, interoperable, platform independent, and vendor neutral. These characteristics are the most important criteria in the development of the webbased systems in today's web world.

Despite of all the advantages of GML, it is worth mentioning that it still is in its infancy, and one can hardly find GML-based system on the web. However, it holds promise to provide interoperability between heterogeneous systems on the web. Based on the obtained experience during this research, it was realized that GML needs advanced knowledge in XML and related topics. working with XML has its own difficulties too. The advantages of applying



Figure 7: AQIS main page

http://locar.ost/Search-dake-po-st-nour.asp - Mic	psort lakernet taplorer
Address (2) http://locathost/Search-date-po-st-hour-asp	▼ 2°60 Links
Air Quality	Information System
A	ivanced Search Page
	Main Page
This page makes it possible to do advanced in name of Station and Pollutant,Date and How along with relevant statistical parameters (m Choose station of Interest: Aghdesieh	earch based on combination of different parameters. With specifing intervals for system, system will provide tables of measured values aximum and average).
Choose Pollutant of Interest: 0 -	
Enter Dates of Interest: From 2003-01-01	To 2003-01-02
note: Date Format: Year-Month Day - Example: 2003-0	1.15
Enter Hours of Interest: From 12	To 19
note: Hour Format: Hour - Example: 14	

Figure 8: Search by station, pollutant, date and time page



Figure 10: Spatial query page

dress 🕢 http://localhost/computePSI.asp	<u>।</u> २७
Air Quality	Information System
<u>air quality in</u>	ndex (AQI) Page
	Main Page
nakes use of CO and PM pollutants values to compr ou need to enter date of interest, choose a station a inter Date of Interest: [2003-01-02]	ite final AQI. In order to understand the air status of stations nd click Submit button. Format:Vear Month Day Example: 2003-01-02
nakes use of CO and PM pollutants values to comp ou need to enter date of interest, choose a station a Enter Date of Interest: [2003-01-02 choose Station of Interest: $\sqrt{4a}$	ate final AQI. In order to understand the air status of stations nd click Submit button. Format:Year Month. Day Example: 2003-01-02

Figure 9: Air quality index (AQI) page

11	ir Quality Result of	Information of Advanced S	on System
	Report fo Name	or Aghdasieh : of Pollutant:	Station CO
low	Date	Hour	Measured Value
1	2003-01-01	12	10
2	2003-01-01	13	7.8
3	2003-01-01	14	8.4
4	2003-01-01	15	8.8
5	2003-01-01	16	10.9
6	2003-01-01	17	12.4
7	2003-01-01	18	13.6
8	2003-01-01	19	17.1
9	2003-01-02	12	10
10	2003-01-02	13	7.8
11	2003-01-02	14	9.4
interest and	5555 CL 65		20

Figure 11: System output based on Figure 8 user entry

iss e http:/	(nocalhost/search-date-po-	st-nour.asp		Links "
Α	ir Quality Result of	Informati	on System	-
	Report fo Name	or Aghdasieh : of Pollutant:	Station CO	
Row	Date	Hour	Measured Value	1
1	2003-01-01	12	10	
2	2003-01-01	13	7.8	
3	2003-01-01	14	8.4	
4	2003-01-01	15	8.8	
5	2003-01-01	16	10.9	
6	2003-01-01	17	12.4	
7	2003-01-01	18	13.6	3
8	2003-01-01	19	17.1	
9	2003-01-02	12	10	
10	2003-01-02	13	7.8	
11	2003-01-02	14	9.4	
the second se				

Figure 12: System output based on Figure 9 user entry

XML and its derivatives such as GML are for greater. Moreover, GML documents contain relevant metadata about data. This feature results in increasing file size of GML considerably in comparison with other formats, but this issue may produce some challenges for web-based systems and requires for more attention.

References

- AQCC, Air Quality Control Company web site report. 2005, available at: <u>http://www.aqcc.org/</u> <u>clean_air.htm</u>
- Charvát, K. and A. Limpouch, The use of topoL for forestry in the Czech Republic. Proceedings of the ^{1,7 th.} International Cartographic Conference, ICC Barcelona, Spain. 1995
- Cox, S., A. Cuthbert and R. Lake, and R. Martell, (eds.). OpenGIS implementation secification #02-009: OpenGISâ Geography Markup Language (GML) Implementation Specification, version 2.1.1, 2002
- De By R. A., R. A. Knippers, Y. Sun, M. C. Ellis, M. J. Kraak, M. J. C. Weir, Y. Georgiadou, M. M. Radwan, C. J. van Westen, W. Kainz and E. J. Sides, Principles of geographic information systems, An introductory textbook, Version of 25^{th.} January 2001, ITC, Netherlands. 232, 2001

ess 🛃 http://	locahost/searcharwindow.a	sp	•	e 60	Norton AntMirus	.
	Air Quality Result of Size of Window: 20	Information f the Spatial 530506 3951023 5312 03.01.01 to 2003.01.03	Que	sten ery	n	
	Stat Locati Pollut	ion Name: Azadi on: 530506 395122 ant/Sensor Name: (24 Co			
Row	Date	Hour	N	leasure	d Value (ppm)	
Row 1	Date 2003-01-01	Hour	N	feasures 5.2	d Value (ppm) 2	
Row 1 2 2	Date 2003-01-01 2003-01-01 2003-01-01	Hour 0	b	feasure: 5.2 4.6	d Yalas (ppm) 2 6 2	
Row 1 2 3	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01	Hour 0 1 2	ħ	feasured 52 4.0 3.0 3.1	d Value (ppm) 2 6 7	
Row 1 2 3 4 5	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01	Hour 0 1 2 3 4	N	deasured 5.2 4.6 3.3 3.3 2.9	d Yelat (ppm) 2 6 7 3	
Row 1 2 3 4 5 6	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01	Hour 0 1 2 3 4 5	2	бевзилес 52 46 33 33 25 25 21	4 Value (ppm) 2 6 7 3 9 5	
Row 1 2 3 4 5 6 7	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01	Hour 0 1 2 3 4 5 6	b	feasurei 52 46 33 33 25 25 25 25 36	d Value (ppm) 2 6 7 3 9 5 6	
Row 1 2 3 4 5 6 7 8	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01	How 0 1 2 3 4 5 6 7	b	feasure 52 46 33 33 25 25 25 36 7	d Value (ppm) 2 6 7 3 9 5 6	
Row 1 2 3 4 5 6 7 8 9	Date 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01 2003-01-01	How 0 1 2 3 4 5 6 7 8	b	feasure 52 46 33 25 25 25 25 36 7 0	d Value (ppm) 2 6 7 3 9 5 5 6	



- Groot, R. and J. Mclaughlim, Geospatial data infrastructure. Concepts, Cases, and Good Practice, Oxford university press. 2000
- USEPA, Air Quality Index A Guide to Air Quality and Your Health, United States EPA-454/R-00-005, US Environmental Protection Agency, Washington, DC 20460, 2005, available at: <u>http://www.epa.gov, Air</u> <u>and Radiation</u>
- Larry, k.. The official XMLSPY handbook. 0-7645-4964-2 Wiley publishing, Indiana, 330, 2003
- Percival, G., OGC AS 12, The OpenGIS Abstract Specification Topic 12: OpenGIS Service Architecture Version 4.3, available at <u>http://www.opengis.org/</u> <u>techno/specs.htm, 2002</u>
- Prins M., Is GML only for Internet GIS, CARIS Geographic Informations Systems BV. In Directions Magazine, 2003