THE SPATIAL DATA INFRASTRUCTURE FOR AIR QUALITY MANAGEMENT

Sergio Cinnirella, Francesco D'Amore, Nicola Pirrone

CNR-Institute of Atmospheric Pollution Research - Division of Rende

Rende, Italy

s.cinnirella@iia.cnr.it

http://giida.iia.cnr.it

KEY WORDS: SDI, Air Quality, INSPIRE, GIS, GeoPortal, WebGis, Interoperable systems, GEOSS

ABSTRACT:

Atmospheric pollution monitoring and modelling applications are fundamental tools for developing environmental policies oriented to control and possibly reduce its impact on ecosystems and human health. However, discrepancies among monitoring systems in terms of spatial and temporal differences and data availability for very different ecosystems are an important limitation in environmental management. In order to standardise protocols and procedure the EU directives (e.g. INSPIRE, Air Quality Directive) and international programs (i.e. GEOSS) have oriented the scientific community to develop advanced interoperable systems able to assure real time data analysis and dissemination within the scientific community as well as to stakeholders and policy makers. To coordinate national earth and cross-disciplinary systems for promoting GEOSS and to support INSPIRE implementation, CNR promoted the GIIDA project (Integrated and Interoperable Management of Environmental Data). Among the various working groups active within GIIDA, that on air quality is addressed develop an interoperable system for air quality information management and dissemination. The system that has been developed in the GIIDA framework is based on open-source tools compliant with European standards and oriented to develop an integrated system that facilitate data storage, data mining and visualization. In our case PostGIS, a de facto standard open source geographic database, was used for data storage. PostGIS was linked to Geoserver, a Web Map Server hosted in a Servlet Container (Tomcat), that export WMS Services with data contained in the database. GeoNetwork was used for discovering the data exported; it also offers a complete tool for adding metadata compliant to INSPIRE directive (ISO 19115, ISO 19139). Stored and discovered information is finally rendered through Intermap, a tool for WMS layer visualization exported from Geoserver. The Spatial Data Infrastructure will holds information on air quality, collected new data sets related at ground-based monitoring sites as well as at off-shore sites. It will support modelling activities and environmental assessments for different case studies aiming to evaluate the impact

of atmospheric pollution on terrestrial and aquatic ecosystems and human health.

1 INTRODUCTION

Health diseases caused by environmental pollution are a growing concern worldwide, which in recent years addressed more studies on causal links between environmental pollution and health at regional and global levels (Katsouyanni et al., 1995, Seaton et al., 1995). A greater understanding is thus needed of the links between exposure to pollution and its effect on health, as well as the long term impact on health of chemical substances, biological organisms and physical changes in the environment. Monitoring and modeling are, therefore, a crucial activity for identifying the key pressures on the environment, the condition or state of the environment, and the level of environmental quality being achieved by society (Roca et al., 2001). Monitoring and modelling are inevitably challenging, not only from a technical point of view but also due to the complexity of the problems being addressed, originating in the interaction of multiple parameters at various levels of organization (anthropogenic or biological, individual or population) and scale (from global to local). Using the information collected through monitoring provides big challenges in integrating and connecting the various information sources used and the technologies implemented. A number of initiatives already exist to link different types of environmental data for various purposes, including efforts from the EEA (EIONET), the WMO (IGACO), the OECD (ECOSANTE), UNECE (EMEP), AC (AMAP), the EU and ESA (GMES) and the EU information system (INSPIRE). These monitoring actions involve several monitoring networks, a wide range of observational platforms and the use of techniques for data harmonization (also called data assimilation), interconnectivity and linkage. Integrated health assessment is not the main purpose of these monitoring networks, nevertheless it requires both routinely available and by-project collected dataset, combined in multiple ways.

Data and model simulations are, therefore, crucial to support policy makers and public participation within any environmental decision-making process as well as for a broad understanding of the environment. However, these data are not always available to the public and are not usually in a format that is understood by all the different stakeholders (Günther, 1997, Gouveia et al., 2004). Also, monitoring systems show relevant discrepancies in terms of spatial and temporal trends, as they do not cover appropriate regions, are discontinue along years and are often application-oriented. To fulfil the gap Spatial Data Infrastructures (SDIs) have been developed. An SDI is a framework of policies, institutional arrangements, technologies, data and people which enables the sharing and effective usage of geographic information (GSDI, 2004). For example, the Canadian Geospatial Data Infrastructure (CGDI) was implemented as an easy-to-use, advanced, online information resource for offering valuable benefits to decision-makers in four priority areas (CGDI, 2009), which are:

- public safety: to share maps of roads, bridges, electrical grids, water systems, buildings, and the like, to better plan for and respond to emergencies and disasters;
- public health community: to share location-based information securely to track pandemics, analyze trends, and monitor population

health;

- local population community: to connect people and communities, map the future, and realize opportunities; and
- environment and sustainable development: to better manage land and water, assess the environment, and monitor ecosystems.

In order to overcoming inconsistencies in spatial data collection, lacking of documentation, incompatibility of spatial data sets, barriers to share and re-use of existing spatial data and incompatibility of geographic information initiatives, the EU adopted the INSPIRE Directive, which was developed to enable effective sharing of geographical data locally, and across borders (EC, 2007). The overall objective of INSPIRE is to make harmonized and quality spatial information readily available to support environmental policies and policies or activities which may have a direct or indirect impact on the environment in Europe. The INSPIRE proposal lays down general rules for the establishment of an SDI in Europe based on infrastructures for spatial information established and operated by the Member States. The component elements of those infrastructures include: metadata; key spatial data themes and spatial data services; network services and technologies; agreements on sharing, access and use; co-ordination and monitoring mechanisms; process and procedures.

In addition, the Group on Earth Observations(GEO) launched a program to coordinate efforts to build a Global Earth Observation System of Systems (GEOSS) aimed to obtain benefit in a broad range of societal benefits ranging from human health to biodiversity loss throughout resources' management, climate prediction, weather forecast and ecosystems protection. GEO provided a web-based interface (GEO Portal) for searching and accessing the data, information, imagery, services and applications available through the Global Earth Observation System of Systems (GEOSS) (GEOSS, 2009).

At Italian National level, some regions have built SDI projects, with different platforms, different costs and different data (e.g. Lombardia, Sardegna, Piemonte) or are developing projects concerning spatial data and integrated information services (Abruzzo, Puglia, Calabria, Sicilia, Molise, Sardegna, Campania). Also the Department of Civil Protection and the Department of Environment of the Italian Environment Ministry have portals respectively oriented to risk prediction/mitigation and collection of a vast Remote Sensing Plan for multispectral data. Also the Comando Carabinieri per la Tutela dell'Ambiente (Environmental Protection Command of Carabinieri Corp) has developed an important project for the control of crimes against the environment, with a wide collection of spatial data, including many iperspectral coverage as well as the Department of Demanio marittimo has rebuilt and updated cadastral maps of all the Italian coasts.

Spatial datasets are, therefore, a basic part of the information systems of the public administrations and the aforementioned projects are connected with the mission of the public administration, but they are hardly ever aimed at achieving an SDI. Most of the projects use standards for interoperable data and network services and often have a web geo-portal, but there is no uniform SDI that covers all of Italy (Vandenbroucke and Beusen, 2007).

To coordinate national earth and cross-disciplinary systems for promoting GEOSS and to support INSPIRE implementation, the Italian National Research Council (CNR) promoted the GIIDA project (Integrated and Interoperable Management of Environmental Data) (Nativi, 2009). GIIDA aims to "to implement the Spatial Information Infrastructure (SII) of CNR for Environmental and Earth Observation data". It was also aimed to design and develop a multidisciplinary cyber-infrastructure for the management, processing and evaluation of Earth and environmental data. This infrastructure will contribute to the Italian presence in international projects and initiatives, such as: INSPIRE, GMES, GEOSS and SEIS.

GIIDA was divided in seven main thematic areas/domains: Biodiversity, Climate Changes, Air Quality, Soil and Water Quality, Risks, Infrastructures for Research and Public Administrations, Sea and Marine resources following the main research areas of the CNR. CNR-Institute of Atmospheric Pollution Research (CNR-IIA) leads the Working Group on Air Quality, which developed i) a specific Web Portal; ii) a thematic catalog service; iii) a thematic thesaurus service; iv) a thematic Wiki; v) standard access and view services for thematic resources such as: datasets, models, and processing services; vi) a couple of significant use scenarios to be demonstrated. This paper describes the SDI on air quality implementation as follows out from GIIDA project.

2 AIR QUALITY AND SPATIAL DATA INFRASTRUCTURE

2.1 Introduction to Air Quality Management

The goal of air quality management is to protect and enhance air quality for preserving human health and ecosystems ((Pirrone and Mahaffey, 2005)). To accomplish the goal numerous regulations and standards, a broad suite of management tools, and several monitoring networks to track progress have been established (e.g. AERONET, EMEP, GAW). All of these components depend on robust and up-to-date scientific and technical input, which includes an understanding of relationships between air pollutant levels and impacts on human health, ecosystems, atmosphere composition and materials. At this level a SDI is fundamental as repository but also as interface between science and management as well as policy.

2.2 Role of SDI in Air Quality Management

A SDI oriented to Air Quality Management is a framework that should collect air quality information at both dataset and metadata level, store them in a database and make them accessible through a web-interface. The idea is to connect heterogeneous databases at different locations and access them from a simple web interface such that they appear as one, virtual database for the end user. Through a data catalogue (metadata) the user's requests can be directed towards the appropriate database server(s). The results of such a query can then be downloaded from a central server or emailed to the customer. Large collections of air quality observation and simulation data can be made available through the cooperation of as many data centres as possible, and the physical location of these datasets becomes irrelevant. The only real problem to be overcome is the different data policies, but this is changing owing to a European Directive (Directive 2003/4/EC) on public access to environmental information (EC, 2003).

2.3 Metadata and Interoperability in Air Quality Management

Storing and organizing data needs to contextualize data themselves in order to give information on collection methodology, lineage, spatial and temporal domain, copyright, context of use, etc. Such documentation is called metadata. Metadata helps organise and maintain data and provides information about an organisation's data holdings in catalogue form; avoids duplication of effort by ensuring the organisation is aware of the existence of data sets; helps users to locate all available geospatial and associated data relevant to an area of interest; builds upon and enhances the data management procedures of the geospatial community; promotes the availability of geospatial data beyond the traditional geospatial community; advertise and promote the availability of their data and potentially link to on line services (GSDI, 2004).

Though air quality data are important inputs to simulation models and decision-making systems, they are highly distributed and heterogeneous, and thus difficult to access in a coordinated manner. The first step to make them interoperable is to produce a spatial and temporal metadata registry that enables the science community to more easily use and exchange data and services. Metadata are very important for interoperability and systems integration, because with metadata is possible to define a common standard between systems. Use metadata is the main step for building complex geo-system where interoperability is a common target. The second step is to produce data in common formats following internationally accepted content standards by government agencies, academic institutions, and commercial companies around the world. The central management and storage of air quality data and metadata enable the interoperable data access at all levels of this framework by integrating in a central database data retrieved in most common format, including netCDF, ESRI© shapefile and Microsoft© Excel.

2.4 Users of SDI in Air Quality Management

From the perspective of an End User, a SDI should organize information and resources and distributes them, providing their services via a single access point, often online. The resources managed by a SDI should be Geo Portals or catalogs of Geographic Web Services oriented to users. In both cases, through the access points to the SDI, the end user should find the resources searching among metadata.

Users accessing an air quality SDI can be a human operator or a Service Client, often a Web Service Client. In the first case, the SDI is used by operators that can be data readers or users that create data and metadata. Among them, the decision makers in public administration and researchers interested in the Air Quality are fundamental. In order to match the requirements for this kind of users is important create tools to search geographic data, necessarily based on metadata, allowing the detection of geographical data based on keywords and the geographic location. The data visualization during the search phase has no a fundamental impact on the user (reference), but a minimal visualization tool for performing research based on geographic localization is needed. After this phase, visualization is a critical requirements in Air Quality management, and the visualized products should be different as result of a different use of data. Resulting metadata must indicates access points related to data use.

Data producer has different needs from data reader, as he should create data and metadata. It is important to use a unique standard and a single access point to metadata, in order to manage standard and access policy. Often the access point for metadata is a Web Application, with metadata editor and an end point for metadata catalog service.

Data and metadata entered must be reached even by non-human clients, typically Web Service Clients. This is an important requirement to ensure the scalability of the system towards systems of systems such as GEOSS or GIIDA.

3 DEVELOPMENT OF THE SPATIAL DATA INFRASTRUCTURE FOR AIR QUALITY THEME UNDER GIIDA

3.1 Data Analysis and state of the art

Data and state of the art analysis is crucial to understand what kind of SDI to create. In our case, data are related to pollutants holding geographical information (i.e. latitude, longitude, elevation, pollutant concentration or emission). Data are gathered from monitoring systems and numerical simulations running on a parallel computing cluster. Simulations outputs are often in netCDF format, which is a binary format used for its flexibility and widely used as a container for scientific information (reference). Both measurements and simulations must be integrated into the SDI by considering work methodologies. Some data are or have been collected in geographical containers (e.g. ESRI© shapefile or NetCDF) and come from previous studies that have produced geographic data. They represent an essential background for the startup the Air Quality SDI.

3.2 SDI Architecture and used technologies

The Architecture of the Air Quality SDI aims to provide geographic services for integration into a Service Oriented Architecture (SOA) as GIIDA and GEOSS. SOA is the main orientation of further development within the environmental sector. The core of the system is represented by the PostGIS geographic database, which holds vector information (Fig. 1). PostGIS, an extension of PostgreSQL DBMS, is a *de facto* standard in the panorama of open source software for storage of vectorial geographic data, but its current limit concern raster data storage in comparison with not-free competitors (e.g. Oracle).

Data are stored into the database with a dedicated tool, the Data Management System (DMS), developed at CNR-IIA (http://giida.iia.cnr.it). The DMS receives as input the results of numerical models or measurement data, reads the content and persists them in the central database. Data input is in the format of XML meta-language, netCDF or shapefile. The DMS was developed with a modular architecture that allows to change the functionality to meet futures requirements. The goal is to use the DMS as an information platform for the management of vector data coming from measurements and numerical models.

As shown in Figure 2, the DMS enables the persistence of vector data using Hibernate as Object-relational mapping (ORM) directly connected to the PostGis database. The system has a Web interface that allows interaction with the end-user. We are studying the way to integrate the DMS with used models (e.g. Weather Research and Forecasting Model, WRF) in order to minimize human interactions and enhance integration and interoperability.

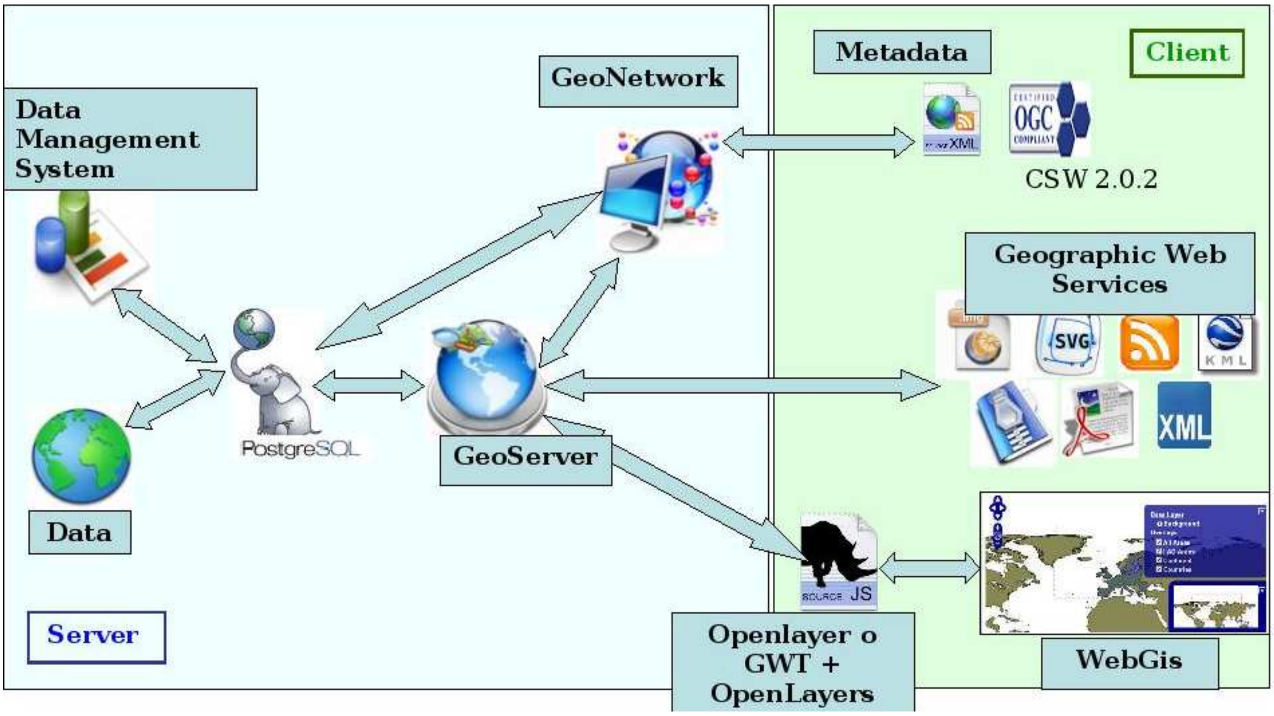

Figure 1: The Architecture of the Air Quality Spatial Data Infrastructure

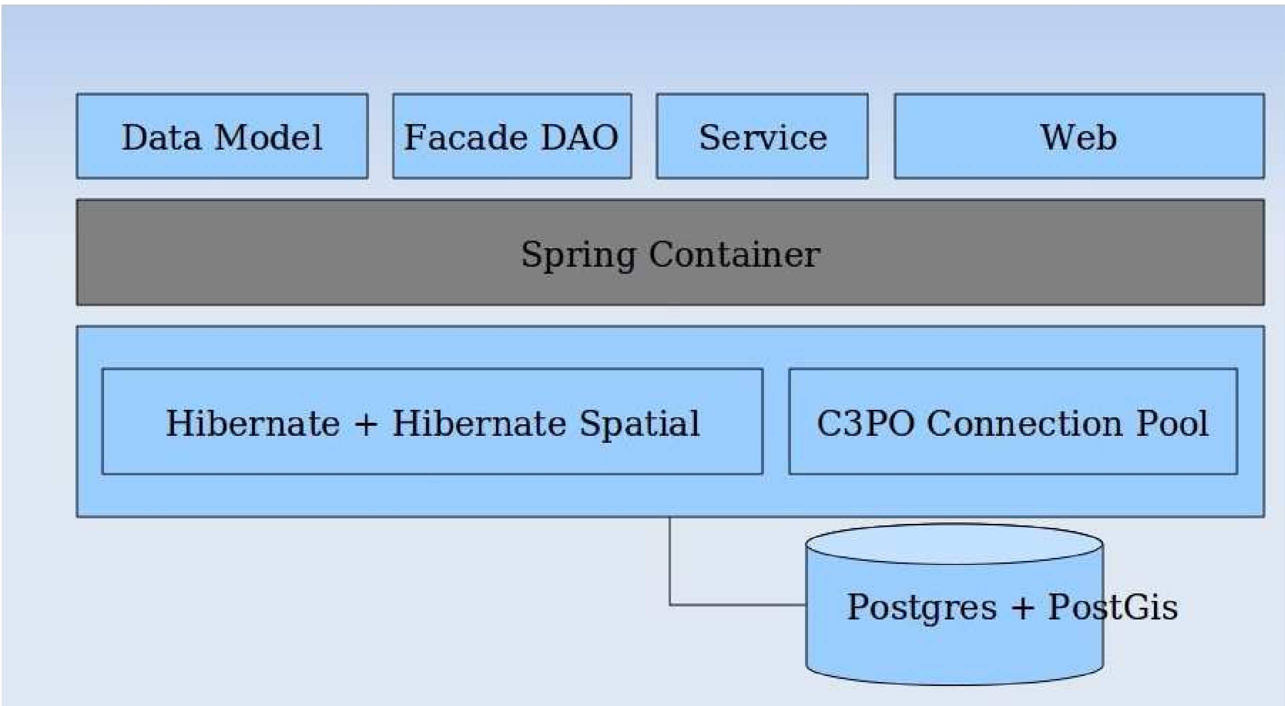

Figure 2: The Data Management System (DMS) Architecture, developed at CNR-IIA

ESRI© shapefiles are *de facto* a standard for spatial data and represent the product of previous works but they are often distributed in several personal computers making data redundancy and not interoperable. As consequence, these vector files have been converted and stored in the database as our goal is to store all geographic data in the central DBMS. A problem remains for raster data that currently are not supported by PostGis. In the SDI they are handled directly by Geoserver and stored in the File System. Nevertheless, we are studying ways to store raster data on the PostGIS database by using mechanisms similar to those used for vectorial data (WKTRaster).

Geoserver is a map server that exports the data by creating Geographic Web Services, complaint with OGC (WMS, WFS, WCS). These services can be used directly by end-users in complex systems SOA compliant, or geoportals built with Web technologies. The OGC services are also used by GeoNetwork for integrate metadata with maps and other geographic data. GeoNetwork is a tool used by Air Quality SDI for the creation and management of metadata based on INSPIRE directive (http://cartoserver.iia.cnr.it/). The metadata is stored in central database, as illustrated in Fig. 2, which thus represents the central unit of storage both for data and metadata. The metadata managed by GeoNetwork are exported via the CSW 2.0.2. This service is the basis for integration of Air Quality SDI in complex systems such as GIIDA, as explained in **??**.

3.3 Metadata Editor: Standard and Metodologies

3.4 The Client: WebGis and DesktopGIS

In Air Quality SDIs, are important three types of clients: 1) Web based clients, oriented to specific products and projects related to Air Quality theme; 2) DektopGIS oriented to researcher and operators in data elaboration ; 3)Web Services, compliant with standard OGC and designed for system integration and interoperability. The DesktopGIS systems can access information in two ways: by connecting directly to the geographic database or at Geoserver services, using WMS. Connecting directly to the database, the user of DesktopGIS that has the credentials to do so, can modify and work on the data. These changes will be made available directly to all services that use the data, given the centrality of the database in our architecture.

The web-based systems are used by end-users by means of geoportals, in which geographic data are used to create high level services and specific projects and are related to particular areas of interest such as monitoring the ozone precursors emission from facilities, or simply view a list of information layers contained in the Air Quality SDI. The web based systems are therefore the ideal tool to realize information systems that reach a wide range of users.

The technologies used for the realization of these systems are based mostly on javascript frameworks as Openlayers, which allow to develop GIS applications with AJAX support. Openlayers is a very effective tool that allows a connecttion directly to a WMS service in order to develop WebGis effective and easy to use. These methodologies currently represent the most used technique in the area of opesource WebGis products and offer excellent results. However, new methodologies for creating Web-based applications are available in the marketplace that allow to overcome the development approach based on Javascript, as building, reusing, and maintaining large JavaScript code bases and AJAX components can be difficult and fragile. For our purposes, GWT a technology integrated with Openlayers, which directly uses Java or similar Software Engineering techniques, was adopted.

3.5 Integration of Air Quality SDI in GIIDA

Interoperability and systems integration has become important in geographic systems and SOA architectures become popular even in geo-systems area . The Air Quality SDI was created having in mind the more high level system, GIIDA. The integration of Air Quality

SDI with GIIDA occurs primarily through the use of OGC CSW protocol that supports the ability to publish and search collections of metadata about geospatial data, services and related resources. It is therefore an integration based on common use of metadata standards according to the INSPIRE directive. The Air Quality SDI exports metadata entered through Geonetwork, using the protocol CSW. The end point for CSW is the catalog GI-CAT that collects data from Geonetowork. The information carried from CSW are linked to GIIDA which can acts as a collector of geographic information. Similarly, other participants to GIIDA can incorporate information coming from their SDIs using CSW protocol, creating thus a complex system of systems that offers integrated services to users.

3.6 Conclusion

One of the most limiting factor, when developing tools and scenarios for air quality, are discrepancies among monitoring systems in terms of spatial and temporal differences and data availability. Fostered by European directives (e.g. INSPIRE, Air Quality Directive) and international programs (i.e. GEOSS), which have oriented the scientific community to develop advanced interoperable systems able to assure real time data analysis and dissemination within the scientific community as well as to stakeholders and policy makers, CNR-IIA developed a framework based on open-source tools compliant with European standards and oriented to develop an integrated system that facilitate data storage, data mining and visualization. The Spatial Data Infrastructure holds information on air quality, collected new data sets related to ground-based monitoring sites as well as to off-shore sites. It supports modelling activities and environmental assessments for different case studies aiming to evaluate the impact of atmospheric pollution on terrestrial and aquatic ecosystems and human health. Further work is necessary to construct a more user-friendly interface for managing information, which facilitates users control and interaction with data input, publication and retrieve. This interface will act as a point of contact that enables interaction between the users' and the SDI on Air Quality.

REFERENCES

CGDI, 2009. Canadian geospatial data infrastructure.

EC, 2003. Directive 2003/4/ec on public access to environmental information.

EC, 2007. Directive 2007/2/ec establishing an infrastructure for spatial information in the european community (inspire).

GEOSS, 2009. The global earth observation system of systems (geoss) 10-year implementation plan.

Gouveia, C., Fonseca, A., Cmara, A. and Ferreira, F., 2004. Promoting the use of environmental data collected by concerned citizens through information and communication technologies. Journal of Environmental Management 71(2), pp. 135 – 154.

GSDI, 2004. Developing spatial data infrastructures: the sdi cookbook. Technical Report GSDI Cookbook, Version 2.0, Global Spatial Data Infrastructure.

Günther, O., 1997. Environmental information systems. SIGMOD Rec. 26(1), pp. 3–4.

Katsouyanni, K., Zmirou, D., Spix, C., Sunyer, J., Schouten, J., Ponka, A., Anderson, H., Le Moullec, Y., Wojtyniak, B., Vigotti, M. and et, a., 1995. Short-term effects of air pollution on health: a european approach using epidemiological time-series data. the aphea project: background, objectives, design. Eur Respir J 8(6), pp. 1030–1038.

Nativi, S., 2009. The giida (management of the cnr environmental data for interoperability) project. Geophysical Research Abstracts 11, pp. 3425.

Pirrone, N. and Mahaffey, K., 2005. Dynamics of Mercury Pollution on Regional and Global Scales. Elsevier, New York.

Roca, J., Padilla, E., Farr, M. and Galletto, V., 2001. Economic growth and atmospheric pollution in spain: discussing the environmental kuznets curve hypothesis. Ecological Economics 39(1), pp. 85 – 99.

Seaton, A., Godden, D., MacNee, W. and Donaldson, K., 1995. Particulate air pollution and acute health effects. The Lancet 345(8943), pp. 176 – 178.

Vandenbroucke, D. and Beusen, P., 2007. Spatial data infrastructures in italy: State of play 2007. Technical Report GSDI Cookbook, Version 2.0, SPATIAL APPLICATIONS DIVISION CATHOLIC UNIVERSITY OF LEUVEN RESEARCH & DEVELOPMENT.