OASI: an integrated multidomain information system

M. Andretta[a], **G. Bernasconi**[a], **G. Corti**[b] **and** **R. Mastropietro**[b]

[a] *Ufficio della Prevenzione dei Rumori, Dipartimento del Territorio, Canton Ticino , Switzerland*

[b] *Department of Innovative Tecnologies – University of Applied Science of Southern Switzerland (SUPSI), 6928 Manno Switzerland*

Abstract: In the frame of a Swiss national project aimed at developing a decision support system for pollution related data, the Canton of Ticino (the Italian speaking canton of Switzerland) has developed an integrated system (OASI) for data collection, management and analysis for noise, air and traffic related data. Purpose of the OASI system is to follow the evolution of air and noise pollution in time and space, making data available and accessible; thus becoming a centre for data verification and analysis and, eventually, a tool for decision making. The OASI system has a scalable architecture as additional domains can be added in future to the same framework. OASI data are regularly and automatically collected from a network of hundreds of sensors distributed on the Ticino territory and undergo an automatic statistical, intersite and interdomain validation process before being available for further human manual validation and use. In this process, existing validated data are used to detect suspicious newly collected measures.

From the ICT view point, the system is developed in a 3-tier architecture: a data layer, a logical layer and a presentation layer. Data of different domains are collected in separated databases but can be transparently integrated for validation and analysis. One of the most challenging aspects of the project is related to the amount of measurements being collected and the size of the databases, which results in interesting storage management and system performance issues. Furthermore, the system allows the definition of users roles and data ownership. An organization may or may not allow its data to be seen or modified by users belonging to other organizations. Data can be accessed by a Java thin client that addresses scientists and researcher needs or by any Internet browser as well as via commonly used wireless devices, for use by the public. Whatever the type of client, they all share the same logic, which is implemented in the application server.

Keywords: pollution, information systems, environmental monitoring

1. INTRODUCTION

One may consider air pollution or noise pollution as two independent domains about which data could be collected and analyzed independently. However, as the environment is a complex and tightly integrated system, this view falls short when it comes to understanding the reasons of some changes in the monitored values.

In fact, it is difficult, if not impossible, to make a correct analysis of air pollution, for example, if data about traffic, locations of industries and other sources of air pollution and meteorological data are not available.

Similarly, to understand the impact of new traffic regulations on the environment, consistent collection of data about environmental parameters, which are influenced by traffic, is required.

As a further example, meteorological conditions, such as rain and snow, influence air pollution parameters (e.g. Ozone levels) or may influence noise levels and traffic.

What might at a first glance appear as independent domains, are actually tightly integrated. Hence, scientists who are supposed to investigate and analyze the environmental conditions, need tools, that provide the capability to collect, validate and analyze data pertaining to different domains in an integrated fashion.

2. THE OASI PROJECT

The idea to develop a multidomain information system arose from the considerations outlined above. The system was named OASI (an acronym for the Italian Osservatorio Ambientale della

Svizzera Italiana or Environment Observatory of Southern Switzerland).
The OASI project was actually initiated with the primary goal of becoming a source of realtime, reliable information on environmental conditions in the Canton of Ticino (Switzerland). These conditions are especially linked to the pollution induced by traffic on the A2 motorway, which crosses the entire Cantonal territory north south, and is one of the major European backbones as it links Italy to central European Countries via the Gotthard tunnel.

Furthermore, the OASI system is part of a Swiss national initiative to monitor pollution caused by traffic on the main motorways.

3. MONITORING THE ENVIRONMENT

Environmental monitoring must pursue the following goals: a) detect and track the evolution of environment loads over time and space, b) collect meaningful, precise and high quality data in order to allow a fair and effective information to politicians, citizens and researchers, c) allow the analysis of extraordinary situations and d) the effective control of strategies and decisions in environmental matters.

The methodological choices developed in the frame of the OASI project are aimed at reaching the objectives outlined above.

Among innovations, it is worth to underline the following:

- the deployment of coordinated multidomain measurement points. In the first phase, OASI was focused on the air, noise, traffic and meteo data in locations which are along the motorway as defined in the Swiss Federal project. This not only allows to follow the evolution, but also allows to detect correlations among measured parameters, and to analyze the impact of traffic on the environment, in general, and the impact of heavy vehicles traffic on the alpine valleys, especially.
- the development of a specific concept for the noise domain, which allow the permanent collection of phonic emissions and immissions. It is actually a relevant technical and scientific novelty at Swiss level. It was decided to install permanent measurement stations along the motorway and to integrate a network of mobile measurement stations, whose task is to collect data at medium-large distance from sources of noise emission, the motorway in this case. The purpose of this mobile measurement stations is to allow to determine the background noise or, from a different view point, to qualify the noise landscape of a given region.

- the option to record noise as an acoustic picture, so as to allow later direct comparisons based on the actual perceived noise rather than on physical parameters only.

- the joint management of measured data and data about pollution sources (cadastre, streets, industries, railway, antennas, etc.) makes it possible to analyze and correlate activities and their side effects, computed data and measured data.

The goal of the OASI project is not only limited to the building of a tool for observing and recording. It is conceived to get a snapshot of the environmental situation, on the one hand, and to provide support for analysis, validation and exchange of data, on the other.

In fact, data collected in the OASI information system is available to a variety of users depending on their competences. Especially, they are available to authorities of the Cantonal and Federal Administration, that are in charge of working out strategies and making decisions aimed at protecting the environment. In this context OASI becomes both a decision support system and a tool to verify if objectives have been reached.

4. SYSTEM FUNCTIONALITIES

4.1 Data Collection

Data is automatically collected from a variety of sources for the different domains and must be validated before being used for actual research and reporting.

4.2 Data Validation

Given the huge amount of data collected, it is not feasible to leave the entire validation work to scientists or operators. Some automatic support for detecting suspicious data must be provided by the system.
For this purpose a unified quality assessment and data validation system has been developed, that can be applied to all continuously sampled data, independently of the domain. On top of this, the system allows each organization to perform human (i.e. manual) validation on their own data, taking advantage of their long term experience in a

certain environmental domain. The quality assessment system is built as a succession of automated procedures with increasing complexity, and ends with the manual validation module included in the OASI application.

The automated quality control procedures monitor all the incoming data from automatic stations. If requested, early alerts on possible malfunction of the monitoring network can be sent to the responsible operator and to the network maintenance team by e-mail or sms. Statistical algorithms perform interparameter, intersite and interdomain tests and flag suspect data sets, thus implementing a unified flagging system. At the end of each automated procedure, appropriate flags indicate the status (quality) of the data. Therefore, the status of data is determined by the number of executed procedures and by the results obtained for each test.

The automated procedures are conceived to be at the same time a safety net, for early identification of network problems, and a valuable support for the later manual validation of the data, helping the operator in the difficult task of distinguishing instrumentation problems from unusual, but physically plausible situations.

The fact that data from different domains are simultaneously measured at the same location could lead to a complete change in validation paradigms. As an example, spectral characteristics of traffic noise could be used to confirm certain meteorological events, like rain or wind gusts.

4.3 Integrated Data Analysis

Scientists can access the data repository via the OASI application in order to carry out the key part of their work: performing integrated data analysis.

Users can select locations, parameters, time intervals and diagrams types. Multiple curves can be displayed and compared on a single diagram. Data belonging to different domains can be shown on a single screen, for comparison purposes.

4.4 Users Authorizations

Users of the system may belong to different organizations with different responsibilities and different tasks to perform. Data hosted in the database may also belong to different organizations, as well as pertaining to the different domains, as we have seen already. Therefore, the system provides a mechanism to grant access to data that accounts for this.

Each user has a login name and password and is assigned one or more roles within the system. A role may span multiple domains. Each role has a set of functions associated to it, that the holder of that role can carry out on the data. Read-only or read-write permissions may be granted to each function. The structure is very flexible and allows to configure new roles with grants as needed. User grants will be reflected in the OASI application layout: some menu entries or buttons will be enabled or disabled based on the user profile. This solves the problems of allowing or disallowing certain operations to certain users. The logic for granting access to data is based on data ownership: each data is generated by an instrument. An instrument belongs to an organization. Users belong to organizations. An organization can grant access to its data to users belonging to other organizations.

This architecture makes it possible for an external organization to join the OASI framework and add its data to the system, with the choice to allow or to prevent access to them to all or to a subset of users.

4.5 GIS Integration

Environmental data are obviously geographically referenced, hence an environmental information system must provide some sort of integration with GIS applications.

In the case of the OASI system, the integration is twofold: data can be accessed via a GIS application (ESRI) and via GIS-like functionalities built into the application and conceived to perform simple tasks that would hardly justify the need to install a full blown GIS application.

Georeferenced data as well as GIS maps are stored as Oracle Spatial data types. This allows both the integration with a GIS and to perform georeferenced queries from the application making use of the Oracle Spatial extended functions.

From within the OASI application map, GIS layers can be retrieved and displayed. Additional layers can be built using OASI specific data (e.g. the map of air domain sensors' locations).

The map can be browsed to navigate more detailed data.

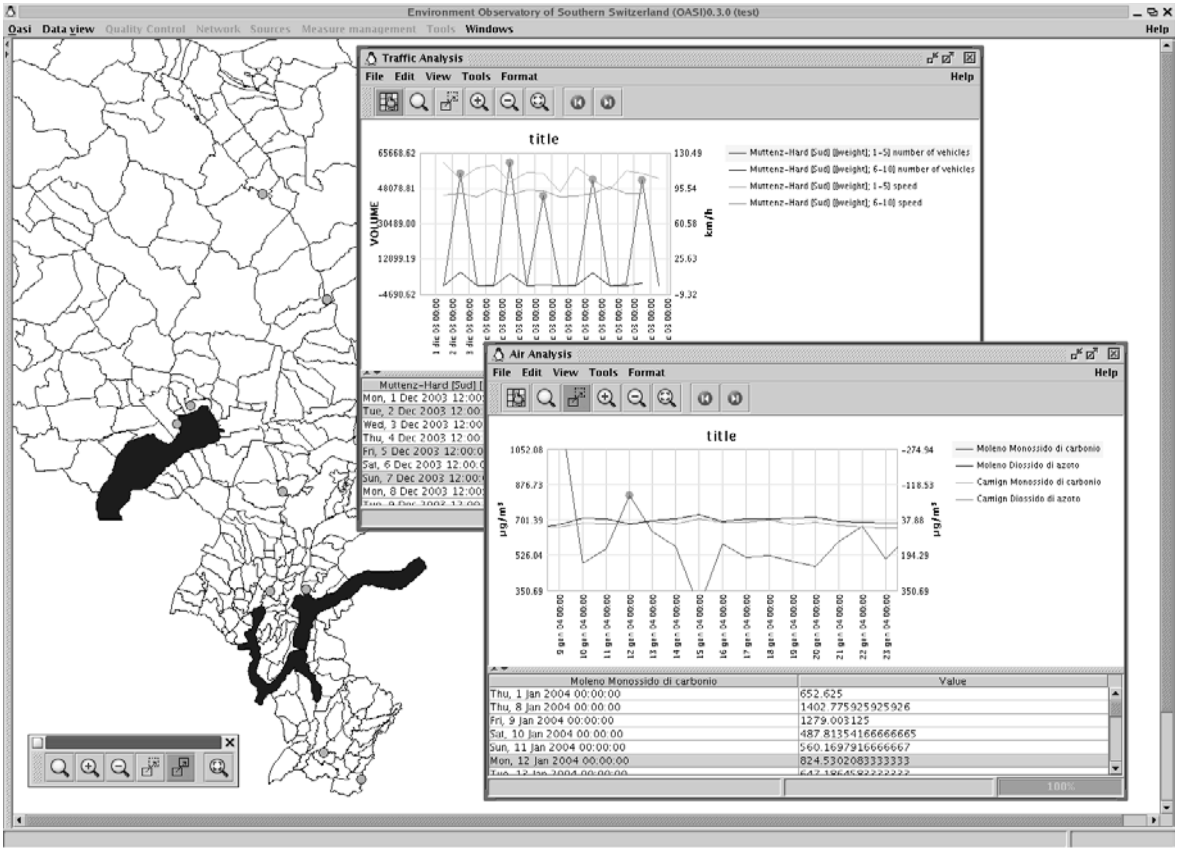

Figure 1. OASI application screenshot

5. SYSTEM ARCHITECTURE

5.1 Database Architecture

Although the OASI information system only deals with pollution related data, these data actually belong to different domains which are dealt with by different Cantonal departments and offices. Each office is consequently responsible for the management of the data of its domain of activity. This organizational aspect would be a driver towards a solution where there is an independent database for each domain. On the other side there are data, administrative data, that are common to all the domains. Moreover, the requirement standing behind the OASI information system is to have an integrated environment where data can be analysed in an integrated fashion and the correlation among events and among different types of pollution can be given evidence.

Based on these somewhat conflicting requirements, a mixed architecture was designed.

Data can be grouped according to the following taxonomy:

- data that must be shared by all domains in read/write mode
- data that must be shared in read mode by multiple domains
- data that are private to one domain but have the same data structure in all domains
- data structure and data that are specific to one domain

As a result of this, there is a shared schema which includes data structures that have to be shared among all domains. And there is a schema for each specific domain. Shared structures are directly visible and are referenced by structures in each domain making domains apparently independent from one another.

Today, the available domains are: air, noise, traffic and meteo. Further domains could be added in future and plugged into the existing infrastructure leveraging the design work already done.

Currently the database is composed by over 200 tables for the four domains supported today.

5.2 Application Architecture

The OASI application is an Internet (web) application based on the 3-tier model. The OASI application is written in Java and is self-contained. It is actually composed by a client side and a

servlet-based server side. The entire application logic is server side and is run within the application server.
The client side is thin and implements just GUI aspects of the application.
The server side modules are used also to support the accesses via the web site which is available, with limited functionality compared to the application, for simplified public access.
A client, be it a Java stand-alone application or an Internet browser, resorts to a number of Java servlets, which in turn access the database through JDBC (Figure 2).
Servlets are capable of recognizing their caller, thus returning data in the appropriate packaging (i.e., HTML for browsers, serialized Java objects, tunneled into HTTP, for Java stand-alone clients).
The code which is, in terms of function, closer to the database is mostly stored and run within the Data Base Management System in the form of stored procedures. Hence, presentation, application logic and data are actually decoupled in three tiers.

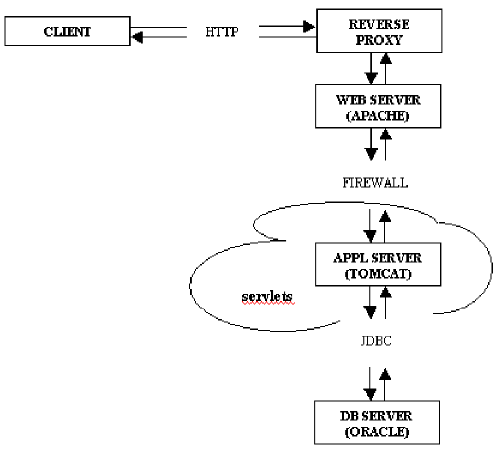

Figure 2. OASI architecture

6. TECHNICAL CHALLANGES

6.1 Database size

The amount of data collected and stored in the databases is one of the challenging aspects of the project. To give an example, the number of vehicles passing by a single measurement station daily averages to 40'000. For each vehicle a number of parameters are measured (speed, length, gap from the previous vehicle, category, …). This means that the system collects an average of 70 millions measures per measurement station every year. This data end up being turned into an equivalent number of table rows. Similarly, storing the noise spectrum for each noise measurement, causes several millions rows to be stored in the data base.

These Very Large Databases are challenging both for the impact on the system and ultimately application performance, but also from a management view point. For instance, running a full database backup requires a considerable amount of time but also noticeable investments in terms of media and infrastructure. In fact, if the size of the backup does not fit a single tape, either a person must be available to replace the tape, or the backup subsystem must be replaced with a tape library that can provide some degree of automation.

Additionally, as data can be queried anytime via the web server access, the database management system cannot be stopped to run a cold backup. The backup must be run while the system is running and is accessed by users.

Analysing the data life cycle, especially the measurements that are those which take up by far the largest portion of database space, it can be observed that after data loading, measurement undergo a semi-automatic validation process and once validated or marked as invalid are not changed any more. This means that measurements older than a few weeks will be kept in the database for years without undergoing any further modification, nevertheless they will be daily backed up with fresh data. To overcome this inefficiency and reduce the daily backup window the tables which contains measurements are partitioned by range, keeping the measures of a each year in a dedicated partition. Following this approach there will be only a single partition per table which will be daily modified and will require a backup. The other partitions can be kept in read-only tablespaces and will not require regular backup.

6.2 System Performance

The large amount of data has an impact also on the system performance perceived by users. While researchers who use the OASI application are aware of the amount of data they are dealing with and so are ready to wait for a few seconds for some computations, internet users behaviour is known to be less tolerant. If the web page is not displayed in a handful of seconds, the user will assume the site is unavailable and will go away.

In order to support fast response to the query submitted via web, which are by the way a reduced subset of all those that may be submitted via the application, it was necessary to store the statistics measure used via the web into materialized views that are automatically updated on a daily basis. This allows very fast response as complex computations against several gigabytes of data are run in batch mode and data are ready to be displayed during interactive activity.

7. CONCLUSIONS

The OASI system currently manages 6 so called *superlocations*, which are permanent and collect data about air, noise, traffic and meteo parameters along the A2 motorway between Chiasso and Basel, and 7 locations where air and meteo parameters only are collected. These locations collect averaged data with a frequency of half an hour. In future the integration of additional locations for the collection of data about the quality of the air is foreseen, as well as the extension of the system to other domains, namely to that of non ionizing radiations.

In terms of functionalities, access to the system via various kinds of mobile devices, from old GSM phones via SMS or from more modern Java-enabled devices via simplified yet powerful interfaces, has already been developed as a prototype.

From the methodological and operational directions described previously, a great built-in potential of the OASI project can be derived, which is not limited to environmental aspects, but can also cover administrative and management aspects.

Due to the its interdisciplinarity, the whole system will have to evolve and adapt to new, upcoming requirements and expectations.