RIVELA: database for the research on Venice and the lagoon

Pierpaolo Campostrini[1], Caterina Dabalà[1], Stefania De Zorzi[1] and Renzo Orsini[2]

[1]CORILA
  San Marco 2847, I-30124 Venezia, Italy
  E-mail: campostrini@corila.it

[2]Computer Science Department, University of Venice
  via Torino 155, I-30172 Venezia, Italy

Abstract

CORILA is running a complex Research Programme on the Venice Lagoon, involving several research institutions of different fields. The integrated and interdisciplinary management of information constitutes a challenging goal of the Programme. RIVELA is a relational database of which the main aim is to store and make available the data and results of the present Research Programme, but it is also open to consider previous and different data sets collected on the Venice Lagoon, even for different purposes. The vision is to achieve, in a near future, a distributed database made of many archives, each independently maintained by different administrations, with a complete data sets visibility and browsing from any position. The data model specifically built for RIVELA required to interview all the data producers, coping with the different ways of storing the relevant information, in order to find a data structure effectively able to handle data of various types and origins, but not too complicated, to persuade to use the system. The interdisciplinary use of data is of paramount importance in the environmental research, and specifically in the Venice case. The scientists' co-operation from the first stage has been considered fundamental to achieve the project goal. The RIVELA relational database structure is made of two parts, static and dynamic. All the interactions with the database occur via the web. The data input phase requires first a CSV file preparation on the user's computer, following easy rules, then to connect to the CORILA web site to send the file. A consistency check is automatically performed before the actual database population. The access to data occurs via the web, through guided interactive searches and GIS queries. The query result is produced in a tabular format, which can be shown either in the web page, or downloaded for further elaborations. The result visualisation on GIS maps will be provided too.

Keywords: Environmental database; GIS; Web interface; Data model; Venice Lagoon.

Introduction: the research on the Venice Lagoon and the many environmental data archives

The Consortium for the Researches on the Venice Lagoon System (CORILA) is an association of the University of Venice, the University of Padua, the Venice University Institute of Architecture (IUAV) and the National Research Council of Italy (CNR). It was established in 1998 to promote and to co-ordinate the scientific research on Venice and its Lagoon. The current Research Programme (2000-2004) has a value of € 10 million and is mainly funded by the Italian Special Law for Venice, via the Ministry for Teaching, University and Research. The programme involves an international network of 70 research institutions and it is organized in 11 research lines of different disciplines, grouped in four thematic areas: Economy, Architecture, Environmental Processes and Data Management.

The Venice Lagoon represents a very complex and delicate environment of unique cultural and natural value. The urban and lagoon environments require continuous interventions, both for maintenance and for remediation of errors made in the past (*e.g.* pollution in the sediments due to past industrialisation) and for coping with global changes problems (mainly sea level rise). Moreover, in order to continue to exist as a living city, Venice should maintain and develop economic activities (*e.g.* port, industries, fishery), while it is subjected to an increasing tourist pressure. Therefore, from the management needs, several specific and general questions rise, how to maintain all these activities and interventions in a 'sustainable framework' for the safeguarding of Venice: some of these questions are challenging the scientific community.

In fact, since many years, a wide research in several branches is occurring on the Venice Lagoon system, producing a huge amount of data, reports, publications. However, in the past, there was not a systematic effort in the scientific research to put together all the pieces of information, making available tools which would permit an easy access to all the data. Nevertheless, very long historical series of data have been recorded, by single private or public institutions and in the last years some of them have been made available, mainly through the efforts spent by a private academic institution, *Istituto Veneto di Scienze, Lettere ed Arti*, based in Venice (*www.istitutoveneto.it*).

Independently from the scientific community, the Ministry of Public Works, who has the competence by law for the Venice Lagoon safeguarding, through its local office *Magistrato alle Acque*, is running large monitoring activities and is maintaining the larger data set: however, presently the access to data is strictly regulated and possible only after a specific authorisation, while an easier access is foreseen in the near future. Other Administrations collect data for specific purposes (*e.g.* air or water quality control, tide level forecast, fishery health checks, etc.), but there is not at present a real structured information exchange, even where this would permit relevant mutual aid for the administrations themselves and would improve the phenomena comprehension by the scientific community.

Generally speaking, the problem of data archiving and sharing is presently felt as one of the most relevant for properly assessing the safeguarding of Venice, not only by scientists, but also by administrators and lay people. In this framework, the CORILA's database RIVELA is intended to be a first step, mainly oriented towards the scientific community, but it is expressly built open to a wider integration with other archives, for realising, in a near future, a really accessible-to-all distributed information system of the Venice Lagoon.

The need of a common data model for interdisciplinary research

RIVELA is a relational and flexible database for storing and managing, in permanent and secure archives, information from the researches on the Venice Lagoon, independently from the actual research executor. RIVELA is designed to facilitate the provision of the research results to the scientific community, the decision makers and the general public.

Presently, it is populated mainly by the data produced by the first CORILA's Research Programme 2000-2004. It is well-known that the main problem in planning, structuring and organizing an environmental database is the management of the mass of data which derive from different disciplines (Michener and Brunt, 2002). In fact, the complexity and the richness of the research programme ask for a carefully designed data management, both for long term archiving of data and for the immediate and interdisciplinary data usage by researchers of many different sciences. Therefore one of the most important objectives of RIVELA is the definition of a common data model able to integrate and manage interdisciplinary data (*e.g.* the interaction between chemical parameters, information about the biota and physical data).

The development of such a database required, in the planning phase, the lengthy task of organizing the existing knowledge, to obtain an accurate description of environmental matrices, sampling localities, parameters to be measured, types of sample, data acquisition methodologies, and a lot of other ancillary information, to be stored in the database in addition to the main data. Thus, the classification of records in RIVELA is the result of a complex interaction with the research groups to collect and define, in a common structure, the different ways to store the relevant information.

The RIVELA database structure

The RIVELA relational database consists of two main components (Fig. 1): a *static* part, relative to the results of auxiliary support data, and a dynamic part, relative to surveys performed in field or laboratory activities. The database engine is Oracle 9i.

The static part contains the following information:
- Research Groups administrative data and list of current activities (Research Projects, Work Packages, Activity);
- Geographical location of data (Zones, Environmental Units, Localities);
- Data types (Matrices, Types of Sample, Parameters);
- Data acquisition methodologies (Method, Apparatus).

The dynamic part contains four interconnected fundamental entities: *Measurements, Samples, Stations* and *Sampling Activities*. A *Measurement* is the value of a parameter deriving from a certain Sample, which has a precise spatial location (*Station*) and a temporal location. The Samples are, in turn, classified according to type (*Sample type*), depending on the environmental matrix.

RIVELA applications

To ensure the maximum flexibility of accessing RIVELA, all the applications have a web interface. This allows the researchers to work remotely both for loading and for querying the data. Moreover, RIVELA is designed to offer both textual and GIS interfaces. A web application allows the visualization of the static part of the database, and the suggestion to the database administrator of modifications or additions to the following lists: Parameters, Environmental Units, Localities, Substratum, Unit of Measurement.

The other two main web applications available are for data input, and for data search and extraction through guided queries.
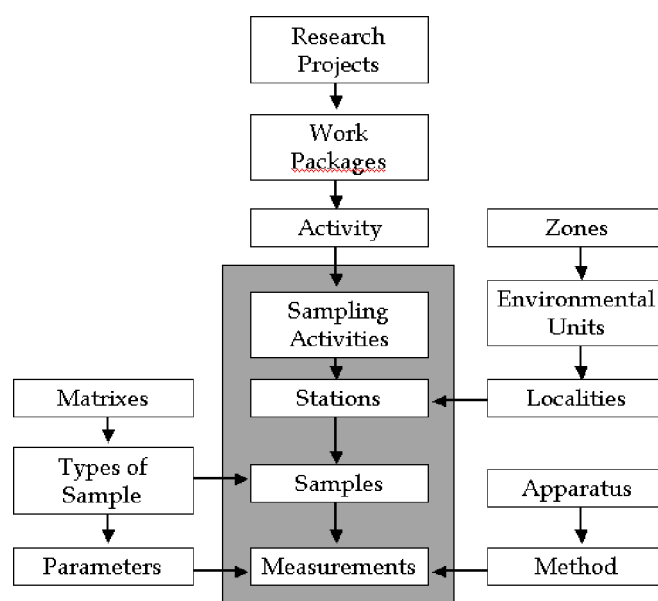

Fig.1. Simplified scheme of the RIVELA structure. The dynamic part is in gray framework.

Data input

The use of web application for data input, even if largely suggested, presently is not so frequently implemented in scientific data bases (Bobak, 2000; Michener and Brunt, 2002). The presence of many users of different institutions represented a strong incentive to use this approach.

In order to feed RIVELA, the scientific coordinator of a research unit, properly identified with username and password, accesses an appropriate web page containing both a series of fields to be filled in manually and buttons to send one or more files in CSV format, which contain the data previously prepared on his own computer. Information on the methodologies used must also be inserted, and auxiliary data like bibliographies, images, etc. can be provided too.

Three formats for data files are foreseen, according to the following types of measurements:

- a set of measures relative to a unique sample of a unique station in a unique sampling activity;
- a set of measures relative to a set of samples and measurements related to a unique station via a unique sampling activity;
- data relative to a unique sampling activity with many stations and many measurements.

In the data files, data are organized in tabular form, where, as a general rule, the rows represent different values of a set of parameters measured in a specific sample and/or at a particular time; the columns represent all the values taken for each parameter during measurements of the samples or over time. So, the number of the columns is variable, as each column contains all the measured values of each single parameter. The file can be prepared with any program managing tables, like MS Excel or MS Access, or any text editor, or can be produced by appropriate software converting data from some measurement device, and saved in the standard CSV format. Each file sent to the database is processed by an automatic loading program, which carries out a set of controls of validity (presence of the coordinates in the correct reference frame, correct name of the parameters etc.) and indicates by e-mail any errors, both to the user and to the administrator. In fact, this control is considered of paramount importance to guarantee the data quality (Brackett, 2000; Michener and Brunt, 2002).

The structure scheme of the data input is represented in Fig. 2. An important feature of the input system is that the CSV file containing the input data is permanently stored in the database too. Moreover, all data in the database maintain an information which can be used to trace back to the original input file. These two facts make possible the following fundamental feature of RIVELA: *it is always possible to correct the database whenever an error is discovered in an input data file*, by 'undoing' the relative insertions, and 'redoing' the new correct ones.

Guided queries

RIVELA is a relational database, and therefore all the usual ways of accessing the data are available to users: for instance, by using directly the SQL query language, or through "individual productivity" tools, such as spreadsheets, like Excel, or databases, like Access or FileMaker, to extract sets of data in tabular format. Moreover, two additional access and extraction ways have been provided: a guided interactive search and a GIS query tool (Fig. 3).
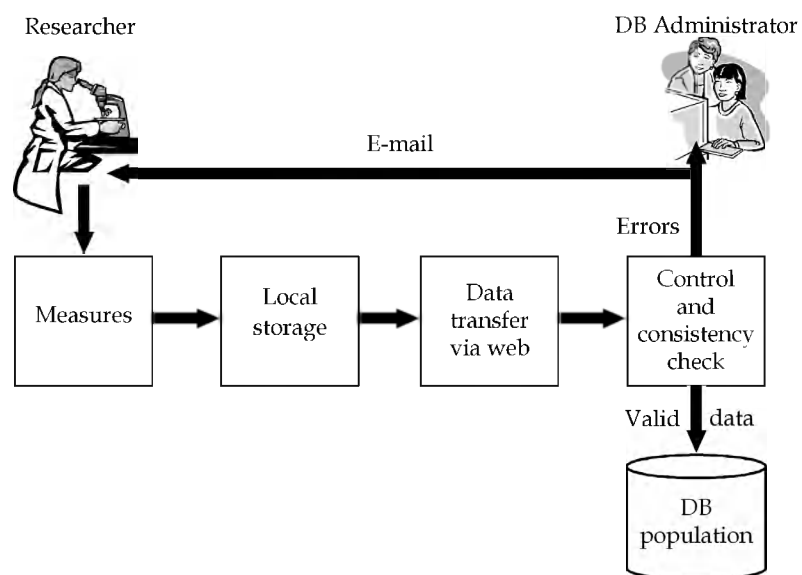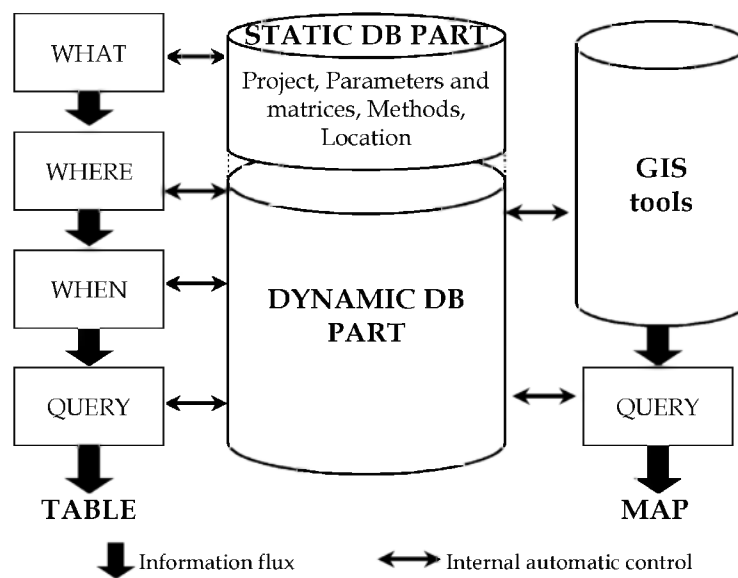
Fig.2. Scheme of RIVELA data input.

Fig. 3. Scheme of a guided query.

In formulating an interactive interrogation, the user can choose between different query modes:

- by writing a free text to select a parameter;
- by choosing among all the parameter types, listed and divided in different matrices;
- by choosing among all the parameter types, classified by the corresponding type of sample.

Then, the user is guided towards:

- selection of time and location;
- selection of attributes required to be shown on the report table.

The spatial location can be indicated either by selecting a locality, or by specifying directly the geographical coordinates of the area of interest.
The attribute selection page allows, by check-boxes, an easy-to-read representation of the report table in the screen window.

Finally, the query result is produced in a tabular format, which can be shown either in the web page, or downloaded, in CSV format, for further elaborations. On the web page, each record is always linked to its ancillary information, which can be explored by traversing all the relations among the data.

Integration with a Geographical Information System

A proper and effective correlation between the environment data, the landscape and the geographical representation is of emerging importance in all information systems used in environmental applications (Hof and Bevers, 2002) and it has been considered essential in RIVELA, where for each stored piece of information, a specific temporal and geographic location is retained. This allows the 'database engine' to be used by a number of GIS applications.

A set of base maps, stored in the database, is provided as the default layout: these are active maps and allow the user to select any geographical object.

The Venice Lagoon is displayed at different scales, to allow users to find easily the relevant information.

All information stored in the database is in principle displayable overlaid on the Venice Lagoon maps. Presently, RIVELA and the GIS tools are integrated only in the local CORILA network, where the GIS applications run on specific workstations and make use of the data stored in the database server. The GIS applications are based on Intergraph GeoMedia© systems family. The system is organized in a client-server architecture.

Many data classifications and correlations can be visualized (Fig. 4), and queries can be expressed by selecting an area with an input device.
The GIS-web interface, that will allow to display the content of RIVELA on the georeferenced maps, is presently under construction. This interface will provide the user with different types of result visualisations, modelling and statistical analyses.
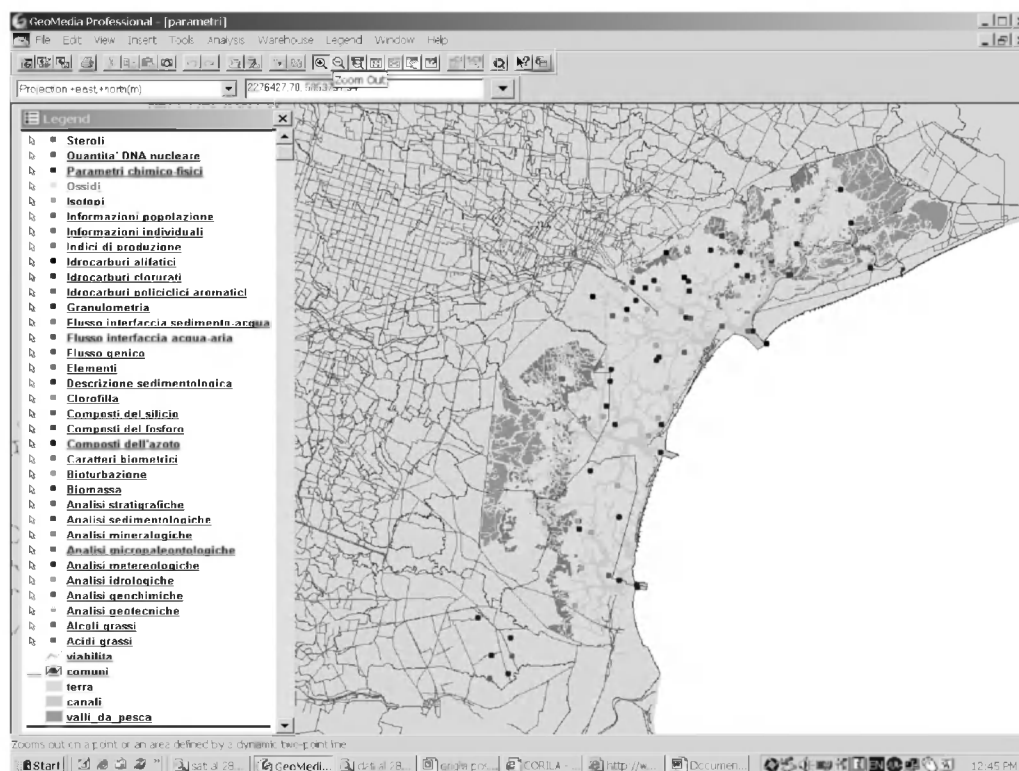
Fig. 4. Example of a GIS query: acquisition sites of the CORILA research program.

Conclusions

RIVELA is a flexible and open archive for scientific researches on the Venice Lagoon, growing with the CORILA Research Program. It is developed in strong connection with the final user, *i.e.* the CORILA researchers and the scientific community. While the use of a relational structure allows the use of standard tools for managing the database, the web interface allows scientists to remotely load and query the data in a very simple manner.

The aim of RIVELA is to become an open archive able to store, in a standard format, the information related to the scientific researches, developed on the Venice Lagoon. Its main goals are:

- to avoid any data loss;
- to permit interdisciplinary data analysis;
- to allow wide dissemination and easy access to data.

The researchers will find in RIVELA a data warehouse, able to give added value to one's own data, accessing larger data pools and using advanced instruments, both for analysis and visualization.

RIVELA is under continuous development: next works will concentrate on the web GIS

interface, on a data engine to search information into textual documents, and on an interface for statistical and spatial analyses.

Acknowledgement

References

Bobak Angelo R.. 2000. Database distribuiti e multi – database. Jackson Libri
Brackett Michael H. 2000. Data resource quality: turning bad habits into good practices. Addison – Wesley
Hof John and Michael Bevers. 2002. Spatial optimization in ecological applications. Columbia University Press
Michener Willam K. and James W. Brunt. 2002. Ecological data (Design, management and processing) – Blackwell Science Ltd