ENVIRONMENTAL DATA AND VENICE LAGOON

RENZO ORSINI, FRANCESCO DALLA LIBERA, ALESSANDRO RONCATO

Dipartimento di Informatica, Università di Venezia

Riassunto.

In questo lavoro viene presentato lo stato di avanzamento del progetto "Modellazione, analisi e visualizzazione di dati ambientali'", nell'ambito del Secondo Programma di Ricerca del CORILA, programma nato per approfondire le conoscenze sui processi che maggiormente contribuiscono alla contaminazione dell'ambiente lagunare veneziano. Vengono descritte le principali problematiche che si incontrano nella gestione di dati ambientali, e i primi risultati del progetto, volti alla definizione di una base di dati ambientali. Vengono infine presentati gli approcci adottati all'interno di alcune linee di ricerca attualmente in corso di sviluppo e di interesse nell'ambito dell'area delle basi di dati..

Abstract.

The current state of the project "Modelling, analysis and visualization of environmental data" is presented, born in the framework of the Second Research Program of CORILA, which is aimed to the increase the knowledge on the pollution processes of the Venice lagoon. In the paper we describe the problems posed by the management of environmental data, as well as the first results of the project for the definition of an environmental database. Moreover we present some of the approaches of the research line currently explored which are of interests in the area of databases.

1. Introduction.

The project "Modelling, analysis and visualization of environmental data", jointly developed by people at the Department of Computer Science of the Venice University "Ca' Foscari" and at the National Institute of Applied Sciences (INSA) of Lyon, is driven by a set of different needs:

- allow the systematic collection of data arising from scientific researches in very different areas, like environmental sciences, idraulics, architecture, social sciences, etc.;
- develop tools and methods to integrate as much as possible such data, also through integration or interchange with other heterogeneous information systems on the Venice lagoon;
- develop tools to allow the use such data both by researchers of the different areas, as well as people responsible of the governance of the territory, in a simple way, through the use of web interfaces, graphical visualization systems or GIS systems;

- develop tools and methods for the analysis of data and the discovery of new knowledge through techniques of data mining, image analysis, spatio-temporal reasoning, complex systems, etc.

The objective of this extended abstract is to give a framework of the project, to present the main results obtained both by this project as well as by previous work devoted to a definition of an environmental database, to discuss some of the research lines currently under development, and to describe a few problems which will be attacked in the rest of the research.

2. The project

A large number of studies currently performed on the Venice lagoon are producing a huge amount of scientific environmental data which are characterized by the complexity, the spread and the heterogeneity. [Smith *et al.*, 1999*;* Bill *et al.*, 1999a,b; Blair *et al.*, 2002b].

In order to use such amount of information to increase the knowledge on the processes which mainly contribute to the pollution of the environment of the lagoon, so that it would be possible to make careful and useful interventions on the territory, it is essential that: 1) the maximum number of interrelations would be exploited among such heterogeneous data; 2) tools be developed to analyse and extract different types of knowledge, to apply not only to single datasets, but on large set of related data; and 3) tools be made available in order to simplify the access to the interrelated data and to the analysis on them.

The project has the objective of producing tools to model, integrate, analyse and access environmental data currently collected by research projects of by external sources, through three workpackages which are devoted to the three aspects cited above.

The first workpackage tackles the problem of modelling the data and integrate them through three different point of view: 1) by defining data structuring mechanisms which can be used to model heterogeneous data in a unified framework; 2) by studying ontologies available in the environmental area, with the production of tools for creating and using them by researchers of the field, and 3) by providing tools for the integration of environmental data available from different data sources and organizations.

In the second workpackage we study different techniques for the analysis of environmental data. A first activity is based on data mining techniques for the extraction of frequent patterns and recurring rules for spatio-temporal referred environmental data and on methods for the reasoning on the extracted knowledge. Another activity concerns the use of techniques for pattern recognition for images, applied to images of buildings, in order to extract significant features and acquire knowledge on their age and state. Another activity concerns the modeling of the lagoon with the use of the theory of Complex Systems, in order to evaluate some of its properties, like efficiency and weak points.

The third workpackage is devoted to the development of models and tools for the analysis of continuous data, that is data which model continuous phenomena and which are derived from continuous fields. Such a data are known only on certain points (sampled data). The problem is the building of spatio-temporal interpolations in order to

know the phenomenum in each point, and to store not only the sampled points with their coordinates and attributes, but also the spatial and temporal interpolation methods.

2. The database RIVELA

It is well known that the most difficult aspect of structuring an environmental database is the necessity of managing lot of data deriving from completely different areas, like geology, idrology, biology, chemistry, physics, etc. [Michener *et al.*, 2002]. In fact the environment is a really complex system, and each discipline which studies it treat different data with different approaches. The management of the complexity and variety of such data require a database which can accommodate the different research groups and which allows the integration of those data in a single coherent but flexible structure, and, at the same time, which facilitates, through the exploitation of implicit relationships among the data, the discovery and analysis of complex phenomena which span onto different areas.

At the CORILA the relational database RIVELA ("Ricerche su Venezia e la Laguna", Researches on Venice and Lagoon) has been defined and implemented which integrates both biological and chemical-physical data, with the relevant geographical and temporal aspects, as well as any information concerning their acquisition [Campostrini *et al.*, 2001, Campostrini *et al.*, 2002].

The database consists of two main components: a *statical* one, with auxiliary information, and a *dynamic* one, concerning field or laboratory data. In Fig. 1 the main sections of the database are shown, and the dynamic component is in the grey area.

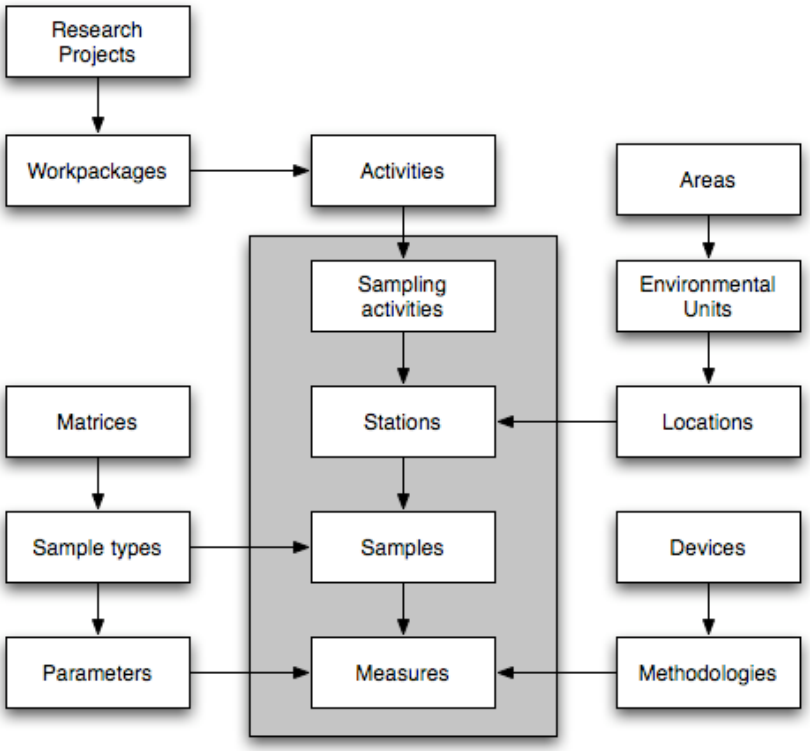

Fig.1 The structure of the RIVELA database

The static part maintains the following information:

- Administrative and scientific data concerning the research activities (Research Projects, Workpackages, Activities)
- Geographical location of data (Area, Environmental Units, Locations)
- Information on types of data acquired (Matrices, Sample Types, Parameters)
- Methodologies and Devices of data acquisition

The dynamic part contains four fundamental entities: Measures, Samples, Stations and Sampling Activities. A measure is the value of a parameter which derives from a certain sample, which has exact spatial coordinates (Station) and a temporal location. Samples, on the other hand, are classified according to their type (Sample types), which depend on the environmental matrix.

The database is enriched by a set of web applications which allow the researchers to insert dynamic data, verifying their consistency with the database, search the data also as datasets through guided queries, or visualize the data with a GIS system.

Finally, note that the current version of RIVELA is more complex than what has been described here, by maintaining semi-structured data, as scientific reports, publications, etc., as well as data collected by researchers of other areas (such as historical, economical or architectural data).

3. The use of Ontologies

In the context previously described, the use of ontologies and thesauri is considered important for the following reasons:

- enhance the sharing of the information among the researchers of different areas (and also in the same area), through normative descriptions of data;
- increase the diffusion of standard ontologies (or metadata) to allow a better classification both of semi-structured data (as technical reports) and of structured data (as the collected datasets) in order to improve the retrieval of the information;
- facilitate the integration of Rivela with other databases through tools and methodologies based on ontologies;
- build a solid information infrastructure for the development of tools of data analysis, model building, etc.;
- experiment tools for the cooperative building of onthologies.

A first analysis of the state of the art has been described in [Casagrande *et al.*, 2005], which has found a few scientific ontologies which could be adopted in the project.

Another of the objectives of the project is the development of a web application to allow to the researchers to cooperate to the building and use of a specific ontology.

4. Database integration

In the project a number of integration aspects must be addressed, and several research problems are tackled. In particular:

- the building of tools to access in a uniform and transparent way both Rivela and other databases internal and external to CORILA;
- the development of methods and tools to integrate into the data model of Rivela data currently maintained in external data sources;
- the building of tools which allows the development of scientific analysis software parametric with respect to the data source, and capable of operating both on databases with different structure that on an integrated view of heterogeneous databases.

An approach currently under development which looks promising is based on the concept of "database interface" [Orsini *et al.*, 2005]. This approach is based on a declarative language which allows the definition of an interface between a database and one or more "external" applications which are not defined on its schema. The interface is used to specify the "minimal" requirements by the applications in order to work correctly with a database. Such requirements are described using the SQLI language, which is used both to give a partial specification of database tables and columns (data structure constraints), and partial specifications of table instances (value constraints).

An important objective of this approach is the possibility of porting an application over a database on which a very restricted set of operations is allowed, typically only a limited access, without the possibility, for instance, of defining user views to adapt the database to the application. The main objective is the isolation of the application from the database, by modifying appropriately the interface when one of the two elements changes.

Conclusions.

The project "Modelling, analysis and visualization of environmental data" has been presented, its first results are been described and several promising approaches are discussed. Other activities of the project are in their first stage and for this reason are not yet ready for public discussion.

One important objective of this project is to stimulate the increase of the research on environmental data, both in the area of data modelling and management, as well as in the area of data analysis and visualization. On these arguments a workshop is planned in the next months.

References.

Michener W.K., Brun J.W., 2002. Ecological Data (Design, Management and Processing).. Balckwell Science Ltd..

Campostrini P., Dabalà C., De Zorzi S., Orsini R., 2001, Rivela, (Database for Research on Venice and the Lagoon): An instrument for Environmental Research, in "*Scientific Research and Saveguarding of Venice: CORILA Research Program, 2001 Results*", Venezia.

Campostrini, P., Dabalà, C., De Zorzi S., Orsini R., 2002, Rivela: Database for the research on Venice and the Lagoon. In "*The colour of Ocean Data. International Symposium on Oceanographic Data and Information Management*". Brussels, Belgium.

Casagrande A., Orsini R., Dalla Libera F., 2005, Stato dell'arte su formati, linguaggi e modelli per la definizione e lo scambio dei dati delle Scienze della Terra. Rapporto Tecnico Dipartimento di Informatica, Università Di Venezia.

Orsini R., Roncato A., 2005, Interfaces for Databases: Motivations, Definitions and Tools, Rapporto Tecnico Dipartimento di Informatica, Università di Venezia,