Experiences in using parallel systems for land information management(*)

Ettore Apolloni(1), Franco Arcieri(2), Enrico Nardelli(3,4), Maurizio Talamo(1)

(1) Dipartimento di Informatica e Sistemistica, Univ. di Roma 'La Sapienza', Via Salaria 113,
00193 Roma, Italy, e-mail: talamo@disparcs.dis.uniroma1.it.
(2) Algotech s.r.l, Via Appia Nuova 310, 00183 Roma,- Italy.
(3) Dipartimento di Matematica Pura ed Applicata, Univ. di L'Aquila, Via Vetoio, Loc. Coppito,
67100 L'Aquila, Italy, e-mail: nardelli@vxscaq.aquila.infn.it.
(4) Istituto di Analisi dei Sistemi ed Informatica, C.N.R., Viale Manzoni 30, 00185 Roma, Italy.

Abstract

In this paper we report about our experiences in research projects dealing with the use of parallel system technology for the management of land information. From these experiences we have derived a characterization in terms of user requirements, system requirements and system architecture.

The basic feature of a system dealing with planning and decision making problems in the context of land management is the possibility to support incremental constraint definition and modeling activities. In this way it is possible to decide which actions have to be done in an "on-line" way, i.e. interactively and on the basis of dynamically changing data.

We point out which functionalities have be set at the system level, and which ones can be set at the application level. According to the present user profile, we also introduce a target architecture for an on-line decision support system that is based on the view building mechanism.

1. Introduction

Decisions in land management often determine strong effects on the human activities and natural resources and, in general, on the quality of life in the concerned areas. Information Technology (IT) may provide advanced support to strategic and tactical decision making, and to the assessment of the impact of land planning decisions on population, society, and environment.

In this framework the main interest of decision-makers and policy-makers in the user organisations is in the development of an IT-based environment capable - by means of the integration of the required subsystems and tools - to support the definition of methodologies for the building of plans for land resources and for the evaluation in terms of environmental and social impact of such plans.

The user of this environment is able to create and evaluate possible *scenarios* of the impact of strategic and tactical planning decisions in land resources management. The user, working on the basis of application-specific knowledge, is provided with the possibility of the interactive definition of hypotheses and, using suitable simulation models, of testing their impact and consequences with respect to land resources in terms of social, economic and environmental aspects.

Scenarios are evaluated as a result of the interaction of mathematical models and textual/geographical databases and are presented to the user, in the form of thematic maps, by means of advanced graphical interfaces.

For example, a strategical planning activity can be carried out by (i) the characterization of the constraints to be applied to extract relevant spatial objects, (ii) the characterization of the territorial information describing a phenomenon, and (iii) the incremental application, led by an expert, of models constituting scenarios of land classifications according to compatibility and multiobjective optimality criteria.

The structure of the work is the following. In section 2 we present the profile of typical users of systems for land information management. In section 3 we describe the conceptual architecture of the system we propose as a reference framework in the field of land information management. On these grounds, in section 4 we define a parallel system architecture to allow an efficient representation and processing of such data. In section 5 we describe the real implementation of this proposal in the context of the management of a river catchment's basin. Section 6 contains conclusions and directions for future work.

2. Profile of users of land information management systems

The typical user of a system for land information management perform complex operations related to planning, decision making and scenario building.
In particular, user requirements may be so outlined:

- to obtain scenarios as results of complex queries representing a piece of geographic reality;
- to take decisions and to evaluate their consequences in

(*) Partially supported by project "Multidata" of the CNR's research program "Progetto Finalizzato Sistemi Informatici e Calcolo Parallelo" and by projects "Amusing", "Hermes" and "Ituland" of the ESPRIT's research program "Basic Research".

0-8186-6322-7/94 $04.00 © 1994 IEEE

a possibly multiobjective approach;

- to evaluate the impact of external events on the actual data (*data fusion* and *modeling*);
- to build working hypothesis and to compare them.

Then, the most usual operations that are carried out by the user can be summarised as follows:

- Definition of a view on the database. A set of spatial objects, presenting a geometric and descriptive part, which has some interest to the user is selected, while other cumbersome or deceiving objects are obscured.
- Building of the previously defined view: the selected objects are arranged so as to act as the basic datasets for the future handling, computing and transformation activities.
- On the user view, constraints are incrementally applied to extract spatial objects, both on the descriptive and geometric sides.
- Input data to mathematical models are extracted, computed or synthesized starting from the user view datasets. The user often has great difficulties in performing these operations, since neither formal nor practical tools have been developed.
- New spatial objects are created and introduced in the user view making queries, applying geometrical transformations (or functions), mathematical models and constraints.
- Lesser attention is given to the input and output procedures (digitalizing, plotting, ...), since these can be easily modeled as standard activities that do not involve the typical spatial data elaboration issues. Actually, specific I/O modules may have interest if the system must have real-time capability, such as the elaboration of monitoring data on which a spatial simulation model is run.

A typical working session consists of two main steps:

- building a scenario, starting from the spatial object database, by means of queries on the database;
- customising the obtained scenario, by means of the application of constraints and transformations.

As a result, these operations produce a new scenario and a set of new spatial objects, satisfying the user needs.

Then, this kind of user wants to query the spatial objects with regard to their properties, eventually creating new objects as a result; the user is not interested in knowing (at least he could not know how to declare it) which transformation the production of new spatial objects needs. This new user profile introduces a strong requirement for the GIS system; actually, the data cannot be considered like static information, but dynamic; the queries expressed by the user can induce some updates in the database, and these updates must be managed in a coherent way, solving the problem of the consistency maintenance.

The remarks previously made about the user profiles make evident the fact that the activity of the new kind of user of land information management system is characterised by an higher level of abstraction from the implementation details than it was in the past.

Actually, when an application becomes even more complex, it is important to hide irrelevant details, focusing the attention of the user on the whole problem. It is possible to assert that many teams, involved in the research and modeling activities, focus up to 50% of their attention on computational issues that are irrelevant to the scope of the research or of the modeling [AAENT93]. It is clear that significant increases in efficiency are possible for applications where computation is a major activity. Since these users may then focus their resources on the solution of the whole problem, rather than on irrelevant computational issues, the efficiency should increase markedly. The geometric transformations of the data are important, but they do not constitute the real objective of the modeling.

Just as in the field of the programming languages with the passage from the procedural languages to the object oriented ones, or from the low level languages, where there are not control structures, to the structured languages, in the field of the spatial data management now it is important to know *what* we have to do, but we do not need to know *how*.

An analysis of the actual situation in the land resources management field was accomplished to understand the range of application areas, to gain knowledge of which activities are performed during the land management, to know the types of data required and/or used during these activities, to consider tools currently used to carry out land planning, their limitations and required improvements, to realize which are the most used methodologies in land management. This analysis involved the tools actually used by the users, the adopted methodology, the typical user profile, the data currently used and the high system functionalities available to the final user. The following areas were investigated: economic area (bank services, extractive activities, energy planning), network area (traffic, roads, technological networks, transportation, utilities location), social area (education, user information, leisure, telematic services), natural resources (waste, environmental monitoring, environmental planning, accident prevention), housing and built-up environment (urban improvement, town planning, heating plants).

The general results of this analysis can be summarized as follow:

- there are models for some planning activities; in practice, just simple models are used to aid land planning; the main problems opposing to the use of models are the lack of resources and the complexity of real world;
- no unified formal methodology is available; formal methodology use is closely related to the model availability for the planning activity; since there are

several areas where no model exists, no formal methodology is used at all;
- however, there is a set of practical methodologies driven by heuristics tailored to each particular condition of one planning activity in one city;
- data required for carrying out planning activities seems to be more uniform across different planning activities and different cities. In general, cadastral, topological, statistical, descriptive and legislative data are required; the data usually come from public files of data which are available for other purposes;
- a graphic characterisation supported by maps is the most expressive representation used to present planning results to other people (decision makers, public, ...).

3. Reference conceptual architecture

Systems for the management of spatial data are a relevant example of complex applicative environments, due to the structural variety and plenty of relationships implicitly existing among data. These relationships are generally barely exploited, though they may be treated inside the well formalized theory of geometry.

Users of such systems solve complex territorial problems following a typically incremental and iterative way. In fact, many times the end-user interaction is mainly aiming to identify which classes of data are of interest and which relations exist (either explicitly or implicitly) among them; moreover, patterns of manipulations of data and relations cannot be defined a priori since they depend on the expert's working methodology. This means that the user interaction pattern is based on a *identify-relate-apply* cycle, where firstly the classes of data of interest are identified, then relations among them are established, and finally some expert-defined methodology is applied. Therefore, the user performs navigation actions through raw data and partially structured data, browsing through them to establish a schema of interest for his/her task; then he/she carries out experiments and processing actions required by the specific (user- and task- dependent) working methodology. This approach of deriving user views and refining them is common to a wide class of new applications, such as scientific, hypermedia, and financial applications [Z93].

The architecture for such a kind of systems is therefore centered around the module (view builder) supporting user interaction with the database by means of view building capabilities (figure 1).

As we have seen above, building a view may result in deriving new objects, that is extracting knowledge that before was implicitly contained in the data and it has been now materialized. Database is updated as a consequence of this process, which therefore needs to be defined in a way the ensures the maintenance of the overall data consistency. This requires the definition of an interaction model based on a suitable supporting mechanism to allow consistency maintenance.

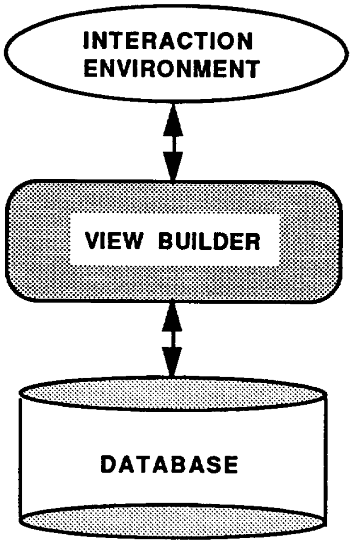

Figure 1.

For geographical data such mechanism can be found in geometry which, thanks to the property of object identityness which characterises its instances (i.e. each spatial items), can act as a kind of fix-point during data manipulation and play the role of a universal key. Geometry allows to maintain those cross-reference links among different views which ensure consistency in the values of the descriptive part of those entities which are connected through implicit relations involving the spatial component. In this way it becomes possible to derive in a correct and consistent way not only the spatial component of new pieces of information and the relations which are involved, but also its descriptive counterpart.

Hence, geometry is the leverage allowing to work safely with implicit knowledge in geographical information. The implementation of such an approach requires the definition of primitive operators which allows to extract knowledge in an efficient and consistent way.

In order to give a profitable support to this kind of users, the system should allow a *high level* language to specify mixed topological and descriptive queries, and to create new spatial objects. This language ought to hide the topological relationship analysis and identification process induced by the query, realising the process of abstraction in the specification of the operations. It ought also to maintain the consistency of the spatial database.

4. A parallel system architecture for the efficient management of land information

Through an iterative approach developed along a number of italian and european research projects we have defined and realized a full system prototype of a working environment supporting land planning and environmental risk assessment based on a parallel architecture [AD89, GNT91, AN92, GINT92, AEN93, H93]. The approach is based on the integration of a set of tools available on the market or in research prototypes, and it is strongly driven from end user applications needs. The focus is on the environment protection issues; the end user is a multidisciplinary expert in environmental issues, and his methodology requires the iterative construction of land scenarios with respect to different working hypothesis. A *scenario* is the representation of a situation which exists just now or may exist in the future as a consequence of events happening just now.

Such a IT-workbench for land expert mainly contains spatial databases, tools for the management of mathematical models, tools for spatial analysis and tools for the definition and application of qualitative rules. The databases can be a snapshot of a land area - i.e. the static data describing persistent information on the land area, such as geological data -, the result of real time data collection from a remote sensing system, an hystorical database maintaining information about the past.

Different case studies, with different user organizations, are used to steer the development of the system prototype, which is based on the interaction between tools managing mathematical models and spatial and descriptive (alphanumeric) databases; the aim is to allow the user to interactively declare the required parameters to tune a mathematical model, and to allow the system, starting from a cartographic database and using these parameters, to automatically build the input to the mathematical model. The system permits the application of rules on a scenario, giving the possibility to characterize land areas with respect to different event and risk hypotheses. In this way, the end user exploits the rule to build a *what if* scenario or a risk scenario. Due to the complexity of the real world it is fundamental to have *flexible* languages to define models and rules.

The system prototype has to support the expert in the decision processes, taking into account the fact that the expert needs an answer within a fixed time: afterward, the result can be useless. In this sense, the prototype is a *real time system*. For this reason, parallel computing techniques are used to support performance requirements deriving from scenario building activities. Different case studies dealing with disasters in the natural environment, such as flood or pollution phenomena, have been developed.

The whole system has two main components:
- the remote data collection and fusion system
- the DSS; an interface permits to access the following modules:
    - an Information Server,
    - a model management system,
    - a rule base system.

The Information Server is a system whose main task is bearing data files in a standard format for internal use. For example, it produces input data for mathematical models; another typical task is to perform data collection and fusion from remote sensing devices, data validation, filtering and final data storage. The remote sensing device network, the model system, and the rule system interact with this system to send or receive data, which are stored or extracted from a global repository. Processing of rules and execution of mathematical models is done on the parallel machine, by manipulating suitable views extracted from the Information Server. The user directly access the Information System by means of a specific Extended-SQL language, that permits the manipulation of thematic maps.

The rule base and the model base, together with a basic and powerful spatial data manipulation language, permit the end user to incrementally build the flood impact scenarios (i.e., the personal view) by means of queries, application of rules and models. The model base system manages the instantiation and running of mathematical models.

In particular, it permits the user to:
- define the definition characteristics of the output map;
- define the function that implements the model on each cell of the land; i.e., the user gives the DSS the formulas to compute the values of physical parameters on that area, such as geology, slope, and so on.

The rule base system manages the definition and application of rules. The application of rules yields to potential scenarios. Rules can be:
- of topological nature, or
- of *what if* type.

The topological rules are applied to the results of models and basic cartography, in order to obtain risk scenarios; the *what if* rules are embedded in the system, i.e. the user cannot introduce new rules or edit them.

In more general terms, our proposal for a system architecture dealing with land information management in the framework of parallel system technology can be described in the following way.

The base assumption is that it is expensive to develop decision support systems for parallel machines and many times work has to be started again and again even for slightly different applications.

The solution to this problem is to define a high-level "workbench", designed for a given classe of applications, that allows to its end-users to flexibly instantiate a solution for the specific problem which has to be solved. Consider that in the field of land information management problems have to solved not just once, but many times, with slight changes, to identify the best solution according to a multi-objective function. Such a workbench has two fundamental parts:

- the design of a parametric data structure for the efficient representation of the way data are organized and related in the considered class of applications;
- the definition of a language that allows to instantiate the parametric data structure on the specific problem to be solved.

Once both parts have been defined and implemented end-users can work at a very high level, by focusing on initial conditions, relative weight of constraints, tuning of state evaluation functions and so on, with no need to bother with code.

In the next section this general proposal will be made more concrete through the description of a real project based on the described approach.

5. A parallel decision support system for the management of a river's catchment basin

An important example of application of the architecture described in previous section is the development of an on-line Decision Support System (DSS) for the assessment of flood impact in the Tiber catchment basin [ABT93]. The system is developed in co-operation with the Tiber Basin Authority ("Autorità di Bacino del fiume Tevere" - ABT).

The basic features of this system are:

- the capability to store a large quantity of heterogeneous data;
- the capability to perform scenario building according to available information, a rule base, and a model base.

In order to achieve the prefixed goals, the ABT made up a hydrological data bank, upon which hydraulic models are run for forecasting ends. The hydrological data bank is continuously updated with data coming from a remote sensing device network. The end-user wants to on-line foresee flooding events, estimate risks, and assess consequences of intervention decisions, according to a multi-objective evaluation functions.

A suitable parallel data structure - the *polyhedric tree* - was developed to support the efficient processing in a parallel environment of a large class of iterative network-flow problems. Its main characteristic is that it bears an analysis of the event in several detail levels.

The evaluation of the risk maps requires to perform heavy spatial operations; in this system their computation is improved by using the parallel environment. For instance, a risk map with respect to a fixed target might be so defined in natural language: *A high-risk industrial area is a region in which the water level is greater than 30 cm and the industrial site is active and the produced substances are toxic, whereas a low-risk industrial area is a region in which water level is greater than 30 cm and the industrial site is active and the produced substances are not toxic*.

It is clear that the topological predicates of this simple risk map definition can be independently evaluated on distinct processors, and only successively combined together.

We consider a fundamental feature the possibility to support several alternate hydraulic models; in order to achieve this goal, we defined a reliable and flexible data structure, which permits the execution of user-selected models. By this point of view, this tool can be considered as an experimentation and comparison environment for hydraulic models.

We show the use of the rule definition language with an example. Suppose a landslide, a road and a electric line maps are available. The landslide map shows the areas subject to landslide risk, without - for semplicity reasons - any further specification; in the road map the highways and other roads are represented; in the electric line map both the wire path and the pylon positions are stored. The end user might define a *electricity_risk map*; in this map a region subject to landslide risk, in which the water level is higher than 30 cm and a electric pylon is settled and a highway crosses the landslide area, is considered as a *high electric risk area*, whose code is 2; if there are no electric pylons, the area might be defined as a *low electric risk area*, whose code is 1. The flooding event map is represented with a extended relational table, in which the descriptive attribute *level* specifies the water level, and the spatial attribute *geo* describes the geometry of the related region; likewise, the road map is represented with a extended relational table in which the descriptive attribute *type* characterize the road type, and the spatial attribute *geo* the related geometry, and so on.The language grammar requires the name of the risk map to be specified, as well as the water level range in the flooding map; optionally, topological operators on whichever map can be linked by the AND operator.

The preceding example rules are translated as follows:
electricity_risk:
flood.geo(flood.level>30cm) AND
INTERSECT(landslide.geo,road.geo) AND
INCLUDE(landslide.geo,electric_pylon.geo) AND

road.geo(road.type=highway) ≥ 2*;*
flood.geo(flood.level>30cm) AND
INTERSECT(landslide.geo,road.geo) AND NOT
INCLUDE(landslide.geo,electric_pylon.geo) AND
road.geo(road.type=highway) ≥ 1*.*

Now the end user chooses a flood routing model, and joins a database in which the model parameter values are stored. When the flood ruoting model is run, the electricity risk map is evaluated according to the polyhedric tree data structure, i.e., its evolution in time at different detail levels.

Moreover, the Extended-SQL spatial data manipulation language provided by the Information Server permits the end user to perform further investigations about the flood event. For instance, the end user might be interested in crossing it with a population map, or in the area of the high electric risk region:

SELECT AREA(electricity_risk.geo)
FROM electicity_risk
WHERE electicity_risk.risk=2;

Of course, the end user can define new risk maps, e.g. a population or a road risk map, by defining suitable qualitative rules on the related map attribute, or redefine a pre-existent risk map by alterating its logical rules.

Moreover, the defined map manipulation language allows to the end-user to define new maps by applying mathematical functions to numerical attributs of regions. So, for example, a new map can be generated by the end user, where for each region defined in the input map, the value of a numeric attribute of the output map is computed according to the formula (max_height - min_height)/area.

6. Conclusions

We have discussed user requirements and a conceptual reference architecture in the context of the use of parallel system technology for the management of land information. This proposal derives by a number of user-driven research projects.

The proposed system allows to its users to carry out planning and decision making activities for land management by supporting modeling activities and incremental constraint definition through mechanisms for the definition of views and the maintenance of their consistence.

The use of a parallel architecture allows more efficiency in the execution of models and sets of rules used to build scenarios, while geographical information (both spatial and descriptive) is managed by a spatial DBMS which act as a central repository, ensuring overall consistency.

Acknowledgments

We are indebted to Luigi Barella and Stefano Ercoli for their contribute to the design and implementation of the many prototypes developed within the HERMES, AMUSING, and ITU-LAND projects.

References

[ABT93] E. Apolloni, G. Batini, M. Talamo, 'Parallel alarm models: a logic framework for the assessment of flooding impact in the Tiber river basin', Proc. of the AM/FM European Conference IX, 'Spatial management in a Europe without borders', Strasbourg, Oct. 1993.

[H93] ESPRIT Project HERMES 5405, Final Report, August 1993.

[GNT91] M. Gargano, E. Nardelli, M. Talamo: "Abstract Data Types for the logical modeling of complex data", Information Systems, Vol. 16, n° 6, 1991.

[AN92] F. Arcieri, E. Nardelli, "An integration approach to the management of geographical information: CARTECH", 2nd ACM-IEEE International Conference on System Integration (ICSI'92), Morristown, New Jersey, USA, June 1992.

[AAENT93] E. Apolloni, F. Arcieri, S. Ercoli, E. Nardelli, M. Talamo, "Un modello di riferimento per l'interazione con sistemi per la gestione di dati geografici" (in italian), in Proc. of SEBD '93, June 1993, Gizzeria Lido, Italy.

[Z93] S.B.Zdonik, "Incremental database systems: databases from the ground up", Proc. of SIGMOD 93, Washington, D.C., USA, May 1993.

[AD89] F. Arcieri, P.Dell'Olmo: A Data Structure for the Efficient Treatment if Topological Join, 4th Intl. Symp. on Computer and Information Sciences, Turkey 1989.

[AEN93] F. Arcieri, S. Ercoli, E. Nardelli: "The geoSQL language for the manipulation of geographical data", International Conference on Computers and Information (ICCI'93), Sudbury, Ontario, Canada, Maggio 1993.

[GIN92] S.Giordani, C.Iorio, E.Nardelli:, M.Talamo: "Lotech: an interactive problem solver for location", International Workshop on Advanced Visual Interfaces (AVI'92), Roma, Maggio 1992, in T.Catarci, M.F.Costabile, S.Levialdi, (eds), Advanced Visual Interfaces, World Scientific Series in Computer Science, 36, World Scientific Press, Dicembre 1992.