

MurMur: A Research Agenda on Multiple Representations

Stefano Spaccapietra,
Christelle Vangenot

Database Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
Tel :+41 21 693 52 10
firstname.name@epfl.ch

Christine Parent

HEC/INFORGE
Université de Lausanne
1015 Lausanne, Switzerland
Tel :+41 21 692 34 18
christine@lbdsun.epfl.ch

Esteban Zimanyi

Department of Informatics
Université Libre de Bruxelles
1050 Bruxelles, Belgique
Tel :+32 2 650 31 85
ezimanyi@ulb.ac.be

ABSTRACT

One of the most publicized goals of modern computer society is to provide flexible access to information for everybody, anywhere, anytime. To make data management systems successful in satisfying user quest for information, an absolute requirement is the capability of tailoring the information to the peculiar needs of the targeted user. Unfortunately, this is a domain where current technology fails to support adequate functionality. Basically, existing approaches to such versatility rely on the idea that all possible views of a piece of information can be derived from a single ultimate representation. This may work in a centralized setting, but it is definitely unlikely to provide a sufficient framework in distributed, interoperable environments. This paper presents a research and development project that intends to explore the issues related to representation and management of multiple representations of the same reality. A specific focus of the project is spatio-temporal information and the support of multi-scale geographic databases, where

objects are represented at many different resolution levels.

Keywords

Data modeling, Conceptual models, Spatio-temporal databases, roles, multi-instantiation.

1. Introduction

Interoperability nowadays bridges the gap between information repositories, giving a chance to find complementary information about the same or related facts in various sources that have been independently developed. Unfortunately, semantic interoperability is not easy to achieve, as related knowledge is most likely described in different terms, using different assumptions and different data structures. Reconciling this heterogeneity to build a fully integrated database is known to be a very hard, currently unresolved problem. A simpler step is the identification of related knowledge and the provision of a mechanism that in some way materializes the relationships between different representations of the same fact. Simply stated, DBMS support of multiple representations is a necessary first step to interoperability, but to be really effective it has to be more powerful than the limited functionality currently provided by these systems, e.g., using generalization hierarchies.

One of the most evident examples where multiple representations are needed and are not supported by current technology is in cartographic applications. Map producers need to build maps of the same geographic region at different levels of abstractions (at different scales/resolutions, in technical terms). Typically, they maintain one database per scale, with no interrelationship, hence with no update propagation and no inter-database consistency.

A multiple representation GIS would allow storing all representations in a single database, and enforce the consistency of the different representations through appropriate, automatic update propagation. Beyond cartography, multiple representations of geographic data are

needed to be able to serve multi-disciplinary user communities, as the same piece of land may support analysis, planning, and forecast activities by city administrations, environmentalists, sociologists, botanists, zoologists, etc.

This paper presents the MurMur (Multiple Representations, Multiple Resolutions) project that aims at establishing and implementing a new representation paradigm allowing each information fact to be recorded through multiple, consistent, and possibly irreducible representations (i.e., not derivable from one another). Such a capability gives full freedom to designers to choose the representations they want to implement, and to users to define which representations they want to query, as well as a personalized outcome for their queries.

2. What is multiple representation and why it is needed

Databases are intended to keep an integrated and consistent set of data that provides the information needed to support application requirements from one or several user communities. These data represent real-world phenomena that are of interest to their users. While the real world is supposed to be unique, its representation depends on the intended purpose. Thus, different applications that have overlapping concerns about real-world phenomena normally require different representations of the same phenomena. Differences may arise in all facets that make up a representation: what information is kept, how it is described, how it is organized (in terms of data structures), how it is coded, what constraints, processes, and rules apply, how it is presented, what are the associated spatial and temporal frameworks, etc.

Current data management technology relies on a centralized representation paradigm, where all application requirements are integrated into a single stored representation (at least at the logical level). A view mechanism allows deriving, on demand, from the stored representation any other representation that materializes the specific viewpoint of the requesting application. View mechanisms, however, are subject to strong limitations. In relational DBMSs it may not be possible to update the database using a view, because of the inherent ambiguity of updates on views that do not rely on a 1:1 mapping between tuples in the view and underlying tuples in the database. In object-oriented DBMSs view definition is further restricted to very simple views (e.g., defined by selection operations) because the rules governing data model constructs lead to inconsistencies in case of more complex view definitions. Object-oriented or object-relational DBMSs provide some additional support for multiple representations through generalization hierarchies and the use of system-generated object identifiers. However, this is known not to be

sufficient (in terms of expressive power, user-friendliness, and practicality) to provide full flexibility in multiple representation support.

What a view-based or otherwise centralized representation mechanism can definitely not support is the case where different application viewpoints are not derivable from each other (irreducible viewpoints). Assume a hospital information system, such that patients are identified by medical teams based on a patient number inscribed on a bracelet that the patient always carries, and the same patients are identified by the administrative staff based on a social security number. If the two viewpoints do not share other information (such as name and birth date) that could provide a common identification scheme, when the patient leaves the hospital two different update operations have to be made for the medical and the administrative realms (no update propagation from one realm to the other is possible). This has evident drawbacks in financial terms (double cost for updates) and in terms of consistency of the database, that can only be guaranteed if appropriate procedures are explicitly defined by users and stored in the DBMS to be automatically triggered whenever needed.

From a traditional, centralized database perspective, the coexistence of irreducible viewpoints in a database may be considered as a design error. From a user perspective it is not. In current DBMSs it is up to application designers and users to cope with the situation, whenever it arises, relying on primitive system functionality, such as foreign keys in relational DBMSs or generalization links in object-oriented DBMSs, to inter-relate different representations of the same phenomenon. It is again up to users and application designers to define and enforce the appropriate consistency rules that may constrain the set of representations.

The centralized representation paradigm is even more uncomfortable when a database results from the integration of different pre-existing data sets, as it is the case in federated or cooperative information systems and in data warehouses. Such systems are more and more frequently required to support interoperation among different organizations, as well as for a single organization that needs to coalesce data from different sources, including the Web, to support its enterprise strategy. When data from various sources come together into a single data store, the situation where different representations of the same phenomena coexist is likely to happen and cannot be considered as a design error.

In summary, modern data management requires a new representation paradigm, such that multiple representations of the same phenomenon may coexist in a database, and this should be explicitly described and made known to the system so that it may manage the situation accordingly. In other words, supporting multiple representations means

that both users and the system are aware that two (or more) stored representations are describing the same real-world phenomenon. To achieve this, existing data models need to be extended with new concepts such as a multi-representation link, with a well-defined semantics (which says "this representation describes the same real-world phenomenon that this other representation"), and associated constraints and operators. Expected benefits include better real-world modeling, enhanced understanding of schema diagrams and database content, improved consistency management, automatic update propagation, and data cleaning facilities (when two representations are used to check one against the other and determine if there has been some erroneous data acquisition).

3. What is multiple resolution and why it is needed

Geographic data is a major provider in terms of systems and services for the citizen, as one of the most common concerns in everyday life is locating something you are looking for. Maps are the most natural way to provide location information, and also serve as an excellent mean of visualizing analytical data about phenomena that have a geographical correlation. Hence, map production and display is an economically significant activity, nowadays supported by GIS and DBMS databases. This includes geography-compliant maps, that show items of interest as faithfully as possible with respect to their real-world location and shape, as well as schematic maps (e.g., city transport systems, airline connections diagrams, train networks, facility management networks), where the focus is on correct connections and readability rather than on precisely locating lines and nodes.

A map is drawn according to a given scale, i.e., the actual drawing depends on the chosen scale. For instance, a rectangular building can be represented by a rectangle on a 1/10'000 map, by a point on a 1/25'000 map and have no geographic representation at coarser scales. Thus, a physical zoom-in or zoom-out operation that would only enlarge or shrink geographic representations is simply inadequate. Drawing standards change from one scale to another one, items may (dis)appear or be (dis)aggregated because their size make them (in)visible depending on the scale, their shape may be modified (made simpler or more precise), or simply the information is not available at the requested scale. Unfortunately, there is no complete set of algorithms that automatically derives a map at some scale from a map at a more precise scale, a process called cartographic generalization. Some algorithms exist and more are being investigated. Given this situation, map production systems tend to keep a separate database per scale, leading to problems such as lack of consistency and uncertain update propagation.

The ideal setting would be of course to maintain a database where geometry information is kept at the most precise scale, and all geometries at less precise scales are automatically computed through cartographic generalization. Because this derivation cannot be fully automated, and also because cartographic generalization may be a long and costly process, the alternative is to perform cartographic generalization off-line and to store its result for direct reusability. Eventually, that means that a spatial object in a database may be associated to a variety of geometric representations that are scale-dependent. Databases with such a facility are called multi-scale databases. Actually, while the scale concept is perfectly understood and relevant when talking about maps, it is not relevant anymore when talking about the geographic representation of an object in a database (scale is the ratio between the objects' size in the map and their size in the real world). When focusing on databases, it is more correct to use the term *resolution*, usually defined as the minimum size of an object to be represented. The resolution of information in the database is the resolution used at data acquisition. If different resolutions have been used for the same object, we can talk about multi-resolution objects. Moving among resolutions has a potential impact on the shape of objects (shape may be simplified in less precise resolutions), on their value (because of a corresponding change in semantic resolution according to user-defined hierarchical value domains), as well as on the existence of objects (because of a change in aggregation rules or creation of new aggregates). From the geodata application perspective, multiple resolution is certainly the most urgent multi-representation problem to solve. The potential variety of representations extends over different facets, such as:

- multiple geometries (possibly belonging to different spatial types, like area and point, or area and line) for the same object,
- multiple abstraction levels that make a set of objects coexist with the object(s) that represents their aggregation (whether the aggregation is based on geometric, temporal, or semantic criteria),
- multiple abstraction levels that result in hierarchical value domains for attributes, and
- multiple representations as for traditional databases.

4. Multiple representation and time-varying information

A wide range of applications needs to manage time-varying information for analysis, planning, and forecast, in particular for decision support systems. This includes classical database applications such as personnel management, medical records, and many others. As for geographical applications, the need for temporal support is critical in the great majority of cases. Typical examples include cadastral, risk management, and environmental applications. A map is also characterized by a given time

period of validity. Of particular importance is the manipulation of moving objects, such as cars, vessels, and pollution disasters, where the geographic characteristics of an object are time-varying. This is a currently unresolved problem for which several proposals have recently appeared in the research literature.

By definition, keeping temporal information means that each recorded fact has multiple representations corresponding to different points in time. When introducing a temporal dimension into data management it is necessary to determine which aspects of time are relevant for the application. There are two ways of adding time to information. The usual way consists in keeping history over the past and predictions over the future by timestamping facts representing real-world phenomena. Thus, attribute values can have an associated period of validity, and objects and relationships can have an associated life cycle (i.e., when they are created, deactivated, reactivated, and deleted). The other association of time with information concerns the modeling of inter-object dynamics (e.g., causality, that object has created that other object; that object lived before that other one; this object is a snapshot of that other one). Although such relationships are rarely supported in existing systems, they are important in particular for applications related to the management, the analysis, and the understanding of natural and human phenomena.

Two complementary and independent viewpoints on timestamping are customary. The first, termed *transaction time*, consists of system-generated timestamps recording when a fact was introduced into or deleted from the database. With the second, termed *valid time*, values for timestamps are user-provided and represent the actual time of occurrence of an event or the period of time where a fact is valid in the real world. Valid time may span the past, the present, and the future. Valid time is the most common requirement for usual applications. The term *user-defined time* refers to a data type that codes valid time values, e.g., a DATE domain, which supports definition of time data items (e.g., order-date) that are relevant for the application, independently from any history keeping concern.

An aspect that contributes to multiple representations is the manipulation of data at different granularities (e.g., year, month, day, hour, minute, second). To this extent, functions are needed to convert from one unit to another.

When time-varying information is kept in a system, an important functionality is a *time-travel* facility allowing a user to visualize the state of an object or a geographical region at a particular point in time. However, this functionality is difficult to achieve for several reasons. For example, the time-varying information kept by the system for different phenomena may have different granularities.

Also, the data may not be available for specific periods in time (e.g., due to the problems in the measurement process). Thus, the need of support for imprecise and incomplete information is more critical in databases keeping time-varying information with respect to traditional databases.

5. The MurMur project objectives

The overall objective of the MurMur proposal is to extend current functionality provided by commercial data management software (DBMS or GIS) to support more flexible representation schemes, such as multiple coexisting representations of the same real-world phenomena (semantic flexibility), in particular representations of georeferenced data at multiple resolutions or scales (cartographic flexibility) and temporal representations keeping track of the evolution of phenomena (time-varying information).

As already stated, supporting multiple representations means that both users and the system are aware that two (or more) stored representations describe the same real-world phenomenon. For instance, in a relational or object-relational database the same car may be represented once as an instance of a Vehicle table and once as an instance of the Collectible table, where the two tables may have no common attribute. In a geographical multiscale database, the same building may have a rather precise geometric representation of its shape at 1/10'000 scale and a less accurate representation at a 1/25'000 scale. As for the temporal dimension, a river may be associated with several geometries to represent the evolution of its course over the years.

MurMur is specifically interested in providing a support to multiple representations that is adequate for geographical data. One of the motivations is the societal importance of geographical data: 80% of all business and policy problems involved in creating livable communities include geographical data as a key element in their decision-making process (source: NCGIA 1999 National GeoData Forum, Washington, DC, June 1999). Another motivation is that the diversity of user profiles is much broader in geodata applications (where it spans from environmentalists to city planners, cadastral administrations, sociologists, etc.) than in conventional database applications, where a given database usually serves a community of users belonging to the same organization. Numerous applications are concerned with the problem of multiple representation, such as environment management, natural risk assessment, multiscale map production systems, cadastral management and embedded navigation applications.

The project is supported by EEC under its 5th Framework IST Programme, and includes a geodata provider (the French national mapping agency, IGN), geodata users

(represented by a French public research center, Cemagref), a GIS supplier (Star Informatic, Belgium) and academics with R&D experience in geographical databases (Université Libre de Bruxelles, ULB, University of Lausanne, UNIL, and the Swiss Federal Institute of Technology at Lausanne, EPFL). From the scientific perspective, the main focus of the project is on addressing the problems due to the coexistence of various representations of e.g., the same piece of land using different resolutions and different viewpoints. Investigations and solutions will nevertheless also cover traditional, non-geographic databases.

Two applications have been selected to define the requirements in terms of modeling concepts and for testing and validating the results expected at the different stages of the project. The two applications cope with very different problems and thus provide complementary knowledge. Furthermore, these applications are representative of a wide range of geographic applications needing support for multiple representations. The first case study is a cartographic application and is mainly concerned with the problems of multiscale databases. In this application several databases at different scales have been independently created and are being maintained. The second case study is a risk management application dealing with both natural risks (e.g., avalanches and floods), and industrial risks (e.g., pollution). This application also needs to manipulate data at multiple resolutions, but here the temporal aspects are of vital importance for risk analysis and prevention.

The scientific outcome of this project will be the formal definition of an enriched spatio-temporal data model, and associated definition and manipulation languages, to allow the management of real-world objects at different spatial and semantic resolutions and at different times. The enriched model and languages shall be able to:

- associate several representations to an object,
- define topological and temporal relationships that may differ with regard to spatial resolution and temporal evolution,
- link objects corresponding to each other at different resolutions and different points in time, and finally,
- describe and query objects and their attributes at different levels of spatial and semantic resolutions, and according to different timeframes.

The requirements of two test applications will be fundamental for determining the expressive power of the data model and the associated definition and manipulation languages. The tests realized with the users involved in these applications and their feedback will ensure that the model and the languages are user-friendly and that they answer their needs for multiple representation.

From a development perspective, MurMur intends to develop, test, and validate a software layer (middleware) on

top of the Star Informatic's GIS. This layer will interface with users on one side and with the underlying GIS on the other side. The design and implementation of the layer shall maximize its portability to other GISs and DBMSs.

Functionality provided by the MurMur layer shall allow users to:

- define and edit a schema of a spatio-temporal database using the enriched data model,
- input data into the database and let it evolve through insert, update, and delete operations, while preserving the integrity constraints on related multiple representations,
- query the database using thematic as well as spatio-temporal criteria.

The next sections details how Murmur is organized to achieve these objectives. Basically, the project may be split in four main phases that are described hereinafter:

- Background analysis
- The data modeling approach
- The data manipulation specification
- Implementation.

6. Background Analysis

The approach that is favored is pragmatic. Application analysis aims at precisely defining the requirements of the target applications in terms of spatio-temporal modeling including multi-representation needs. These requirements will be assessed against what is feasible without overloading either the system or the user, and they will be integrated into a consistent set. Finally, they will be formalized so that a proper solution can be elaborated. The two application case studies will provide significant input to the project and avoid that the initial analysis be biased by peculiarities of a specific application. As the case studies pertain to different application frameworks, they are expected to provide complementary knowledge. In parallel, the functionality of the underlying GIS that will be used in the project will be assessed and evaluated, to get a precise definition of the gap that MurMur intends to fulfill.

6.1 Application analysis: Cartographic case study

This project step focuses on application requirements from IGN, the national map agency that is partner in the project. The agency is mainly involved in producing maps at different scales, covering all parts of the country, and providing geodata information to its customers.

The case study is based on three databases independently created by IGN, each one to store data for maps at a given scale. There are no automated procedures to propagate updates from one database to the other ones. The agency

has heavily invested in developing cartographic generalization algorithms, some dealing with individual objects, some dealing with groups of objects. Another important body of work has been done on an object-matching approach that is seen as a preliminary to a data integration phase. Both activities concur to the idea that in some future the existing databases could be integrated into a single multi-scale database.

6.2 Application analysis: Risk management case study

This part focuses on application requirements from a French regional administration whose responsibilities include urban planning, roads, and infrastructures. The administration cooperates, for the development of its geo-data applications, with the Cemagref research center. Cemagref is currently studying an application where multiple representation issues are particularly relevant. This application aims at developing a scheme for risk management in a populated valley nearby a mountainous environment. A characteristic of the application is that it has to deal with both natural risks (e.g., avalanches and floods) and industrial risks (e.g., pollution). Indeed, several industries are located in an urban setting in the valley, which opens up to the possibility that industrial equipment damaged by a natural catastrophe might generate environmental hazards to people living nearby.

The combined risk assessment application will need data from different databases, to get complementary information about land occupation, industrial hazards, geographic features (snow coverage, rivers, etc.), meteorology, etc. This is a typical venue for a federated database approach, as the data exist and is owned and managed by different administrations. Obviously, there are many relevant inter-database relationships that need to be established on both semantic and geometric features, which explains that Cemagref has already started investigating how to approach multiple representation issues.

Analysis of application requirements in this case will add at least questions about the requirements and constraints imposed by the management of data corresponding to the same reality at different times.

6.3 Technology assessment

Considering the pragmatic orientation of MurMur, it is quite natural that the current state of the corresponding technology is analyzed and its relevance to the project is assessed before starting any new development. Basic underlying technology is provided by data management systems such as DBMSs or GISs.

Regarding DBMSs, MurMur intends to focus both on object-relational systems, such as the latest versions of

Oracle and Informix, and on pure object-oriented systems, such as those supporting the ODMG standard. Object-relational systems are likely to be the leading DBMS technology for the next decades, thanks to the compromise between the relational and the object-oriented paradigm. We shall examine in particular functionalities provided by the Oracle 8i system, which to some extent supports modeling and management of spatial data through two specialized extensions: the Spatial Cartridge and the Time Series Cartridge. As object-relational systems are also close to object-oriented systems, it is relevant to extend the assessment activity to include representatives from the object-oriented family. These systems explicitly support object identity and is-a links, both features needed for multiple representation. Since these systems only provide a partial solution to the problem, we shall precisely identify where improvements are needed in an object-oriented or object-relational framework.

Regarding GISs, we shall analyze a representative subset within the many existing systems available on the marketplace. Our focus will be on those that are either widely distributed, or support the most advanced features (e.g., object-relational or object-oriented GISs).

The reference for the assessment will be the GIS marketed by Star Informatic, partner in the project. This GIS suits well to the intended developments as it already supports to some extent multiple resolution objects, a rare feature in current GISs.

7. The MurMur data modeling approach

This phase will define the data modeling component of the MurMur solution to the problem. The core of this solution is a spatio-temporal data model that responds to the requirements that have eventually been selected. The data model shall be both powerful and simple, not confusing users with too many, too sophisticated, or unintuitive constructs. It should advance the state of the art and still it should be possible to map it to an existing operational model of data management systems.

The definition of a basic spatio-temporal model called MADS [15] is available and will serve as input to this work. The existing model has already been checked against various geodata applications. The positive outcome lets us expect that it will fulfill the spatio-temporal modeling requirements of the MurMur test applications. The current data model obeys the orthogonality principle, i.e., it provides functionality for data structure, space, and time descriptions that are independent from each other. This gives maximum flexibility to application designers, who can design their data definition as they feel appropriate and without being bothered by implementation concerns.

Finally, data acquisition procedures that map the relevant existing test data to the model constructs shall be defined.

7.1 Data model specification

The specification of the MurMur data model will be formally stated, using a simple set-based formalism. This is to avoid ambiguities in its interpretation, which is particularly important in a consortium-based project.

The model shall support powerful conceptual data structures, capable of describing the complexity of real-world phenomena. To that extent it will include objects with complex attribute structure, generic relationships, aggregation links, and multi-instantiation links.

For space description, the model shall allow spatial features to be associated with objects, links or properties, depending on application needs. In particular, it shall provide facilities for the explicit description of spatial relationships and associated constraints. It shall support both the discrete (or, vector) view and the continuous field (or raster) view of space.

For time description, the model should primarily deal with valid time support, although transaction time may be considered. Time features may be associated to objects, relationships, or properties, depending on application needs. They will cover the object (relationship) life cycle semantics, as well as the history track semantics. Thus, it should be possible to query the existence of an object (relationship) at a given time, its status with respect to the life cycle, and the history of selected properties. Finally, through combination of space and time specifications, the model shall support description and management of moving or shape-changing objects (e.g., a flood, avalanche, rescue team).

While the above features provide basic services for conceptual spatio-temporal modeling, more concepts shall be added to adequately support multiple representation objects, with specific emphasis on multiple resolution objects. This should lead to the definition of new types of links, and the associated management rules (e.g., to express that if object A is included in object B at some scale, the topological relationship also holds at a different scale).

7.2 Data acquisition specification

This step will define the data acquisition procedures that map relevant data from the two test applications to the model constructs of the MurMur data model. Test data is currently stored in GISs supporting their own data modeling concepts and rules. A relevant subset shall be extracted from these data stores and input into MurMur to be able to demonstrate MurMur's functionality. We do not intend to develop a reverse engineering tool that would do

the job automatically. Although such a tool would be highly desirable, its design and implementation would be a major overhead distracting MurMur from its specific goals. We will rather design ad-hoc procedures and run them interactively, as needed with respect to the requirements that we want to analyze.

A complementary data input stream will come from object matching processes that are available and currently used by IGN to prepare for manual data integration. Such input takes the form of correlated pair of objects that have been identified through mostly geometric matching algorithms. Inputting such information into MurMur should be prepared through the elaboration of an appropriate data structure for storing an integrated view of the correlated objects. Of course, this implies that also the schema of the source databases is communicated to MurMur. This elaboration will be done manually and input to MurMur using the schema editor interface.

8. Data manipulation specification

Assuming that a database can be installed and data acquired according to the new MurMur data model, the following phase intends to make the data accessible to applications. A reasonable query formulation interface will be defined, based on an underlying formal language that will be defined first. As for the design of the MurMur data model, the elaboration of the associated query language will stem from the analysis of application requirements, in this case formulated as a set of typical queries and services needed in our two case studies. MurMur will also consider services other than data retrieval, such as an intelligent zoom facility and a time-travel facility.

8.1 Query language specification

An analysis of application requirements in terms of data manipulation will be done with the involvement of final users of our test cases, and with strong interdependencies with application analysis. This will provide guidance to define an appropriate spatio-temporal query language that covers multiple representation features, and to specify a supporting software interface.

While the most desirable result, in terms of user friendliness, would probably be a visual interface offering multi-paradigm query formulation, this is a goal that is quite difficult to achieve. A substantial amount of research is still required to identify a visual interface that has full expressive power, is flexible enough to adjust to various query formulation approaches (e.g., sketches, by example, navigational, through voice and gestures) and remains simple enough to be understood and effectively used. This project will be limited to providing a simple, yet powerful interface, combining direct manipulation with textual or menu driven specifications, that will be able to take into

account all requirements through sufficiently sophisticated but easy-to-use techniques.

Typically, such a visual query editor will first display the schema diagram of the database that the user wants to query. Next, the user will select the data items (objects, relationships, attributes) of interest for the query and, similarly, specify the structure of the result (s)he is looking for. This mainly calls for point, click, and drag operations. Specification of the selection predicates will be the hardest challenge to meet, as predicates may be quite complex, with sometimes no intuitive formulation, and include spatio-temporal components for which there is no well-established visual formulation approach. Moreover, as there is a significant variety of operations that are specific to spatio-temporal features (e.g., buffering, minimum bounding rectangle), this is where menu-driven specifications are likely to be used.

The query language itself will be formally defined as a set of algebraic operators, whose properties and rules will be precisely stated. While the kernel of the language will consist of well-known operators, tailored to MurMur's data structures and spatio-temporal features, additional operators and rules will be defined to take into account multiple-representation features.

8.2 Intelligent zoom facility

A zooming operation moves from a display of some area or object at a given resolution to a display of the same area or of the same object at a different resolution. A physical zoom operation is like zooming with a photocopier: it's a purely geometric operation. An intelligent zoom is one that displays, at the target scale, objects or fields that are relevant at that scale and using the representation defined at that scale. For instance, zooming-in on an object displayed as a single building may reveal that it is a "fictitious" building actually representing two separate buildings, that have been merged at the less precise scale because they were too close to represent them separately.

8.3 "Time-travel" facility

When the temporal dimension is taken into account, different representations of the same area or object correspond to different points in time. Thus, the time-travel facility will allow a user to visualize the information about some area or object at a particular point in time. However, this functionality is known to be difficult to achieve, in particular because the information about an object may not be known at the point in time the user is interested in. Thus, interpolation functions are necessary.

9. Implementation

A prototype shall implement the data model, the query language, and other services designed during the previous

phases of the project. Within this project, the goal is to implement the prototype as a software package integrated within Star Informatic's GIS. However, implementation will be designed for facilitating in the future the development of a MurMur front-end to other GISs or DBMSs.

For instance, the translation of MurMur structures into Star Info structures will be performed in a two-steps approach. The first step simplifies a MurMur schema through a set of predefined schema restructuring operations. The goal of this step is to obtain a MurMur schema that uses only simple modeling constructs, such as those found in existing software. The second step rewrites the simplified MurMur schema using the specific constructs of the target system. The advantage of this decomposition of the translation process is that the modules realizing the first step may be reused for a number of different target systems. This improves portability and code reuse and enhances the evolution capabilities of the product. A similar concern for portability will rule the development of the schema and query editing functions.

9.1 Schema editor

The outcome of this task shall be a schema editor, i.e., a visual interface tool that supports schema definition and editing using the MurMur constructs. EPFL's laboratory has already designed such a schema editor for traditional databases [3]. The MurMur tool will conform to state-of-the-art technology for visual schema editing. It shall also be able to store the schema definition into the underlying GIS. To that extent, mapping algorithms will be developed, such that no semantics is lost while moving between MurMur and the underlying GIS. Mappings shall be bi-directional, as data extracted from the GIS will have to be adjusted to MurMur specifications.

9.2 Query editor

This task represents the specification and implementation of the query tool, including map-based querying facilities such as those usually supported by current GISs. As already stated, interactive query formulation will be supported by displaying the schema diagram and allowing users to build a query expression by direct manipulation (pick, drag, and drop) of schema elements. Expressions involving complex operations will be supported through complementary menu-driven specifications; these will be context dependent for providing a user-friendly interface. Expression of complex predicates will be eased by textual facilities.

A textual query syntax and parser will allow the tool to respond to queries from application programs.

10. Related work

Multiple representation is the necessary consequence of both the subjectivity of perception and the diversity of interests with respect to real-world applications. Traditional relational systems support multiple representation through the view mechanism. As little semantics is embedded in data structures, data items can be rearranged in any desirable way using relational algebra operators and the new structure stored as a view definition. Object-oriented database systems follow the same idea, although they cannot support the same flexibility in data restructuring as relational systems. The semantics embedded in composition structures and in generalization links leads to unavoidable restrictions in the extent of the view definition mechanism. Beyond the easiest solutions (i.e., not allowing whatever cannot be easily managed by current object-oriented systems), there has been an increasing number of proposals to support multiple representations through multi-instantiation, i.e., allowing an object to be simultaneously represented in several classes. Some proposals aim at supporting multi-instantiation at the object level, so that objects can be dynamically re-grouped to form new classes, while keeping their membership in their original classes [16, 19]. Other approaches deal with the issue at the schema level. An analysis of the object life cycle determines the set of relevant possible representations for an object. These representations, called *roles*, *aspects* or *deputy objects*, allow to formulate that an object may play several roles during its life span [1, 7, 10, 17, 18]. To allow data sharing, mechanisms such as inheritance or delegation have been proposed.

Even if only a few of these database mechanisms have been investigated for spatial databases, they all can be used for spatial multiple representations. For example, view definition can still be used to support multiple points of view [2], but multi-representation needs cannot be fully satisfied using such a purely deductive approach. As already mentioned, cartographic generalization is a complex process that actually remains partially interactive (i.e., needs human intervention) [14]. As a consequence, a view mechanism cannot express the mapping between the input and output of a cartographic generalization process: input and output data are different representations that are not deducible one from the other. As cartographic generalization processes are usually very long, memorizing the result avoids re-calculating it for each request. In [25] this approach is termed the *representation-oriented view* of generalization.

A significant part of the research in multiscale databases was inspired by the largely hierarchical nature of transitions between scales. Several authors [6, 11, 23] proposed to keep the representations at different scales, obtained from

an interactive cartographic generalization process, by storing spatial objects in a hierarchical data structure, where levels correspond to increasing detail and the various representations of the same object are linked together. Objects can then be used to compose a map at a particular scale. In [24] is proposed the map cube model whose hierarchical data structure supports and describes the behavior of pure map objects that represent the same real-world entity in map series. This data structure is a combination of three different types of hierarchies: an aggregation hierarchy for aggregated objects, a generalization hierarchy for more generic object classes, and a filtering hierarchy to filter objects. The map cube model has been implemented and used in several case studies. According to [13], the classification and aggregation hierarchies play an important role in linking the definition of spatial objects at several scales. He proposes a spatial data model for multiscale approaches that contains topologic structures for the representation of spatial objects and classification and aggregation hierarchies. The impact of spatial and semantic resolution on data representation from both the modeling and querying points of view is analyzed in [20, 21]. The authors propose a model allowing to represent a particular kind of data: zones that fit into each other. The associated language allows database querying of thematic attributes with hierarchical domains without exact knowledge of the data abstraction level. A system helping to process and reason with spatial datasets that are heterogeneous with regard to semantic and spatial resolution is proposed in [22]. The granularity of a representation specifies the levels of detail with respect to which the data is registered. The authors distinguish between spatial and semantic granularity. Their multi-representation system, called *stratified map space*, consists of a granularity lattice and for each granularity a *map space*. The schemas corresponding to the semantic granularity and the datasets, which have a representation at this spatial resolution, are associated to the map space.

Linking the representations at different levels of detail allows to characterize the semantics of multi-representation and the associated constraints. Such links may carry information about the transitions between scales. Those inter-level connectivities [12] allow to propagate updates. In [4] a multiscale database is built from two monoscale databases where the corresponding representations at different scales are linked with scale-transition relationships conveying the sequence of generalization operators.

In [9] is presented a deductive knowledge-based system, Geodyssey, that provides a solution for the design and the implementation of multi-scale, multiple temporal, multiple representation spatial databases. Each real-world object within an object directory records references to its scale- and time-dependent geometric representations. Real-world

entities are described by means of classes organized within a classification hierarchy, spatial and temporal relationships, and metadata. The system incorporates a reasoning process to maintain consistency and propagate updates. Moreover, in the query interface, simplified representations of objects may be derived if no suitable representation is found. Several aspects of this proposal have been implemented using deductive and object-oriented database techniques.

An important issue is the problem of maintaining consistency among the objects in a database with multiple representations. Topological information is generally considered as first-class geographic information [5] and thus, the topological relationships among objects at different scales allow to compare multiple representations and to assess their consistency. In [8] is studied the evolution of topological relationships between two objects through the change of scale. The work introduces the concept of R-Like relationship to get consistent answers when querying sets of scale-heterogeneous data.

11. Conclusion

A first innovative aspect of MurMur is a new representation paradigm, the multiple representation approach, which is the core of the MurMur proposal. To enhance the limited multi-representation functionality of current data management systems, researchers have investigated multi-instantiation mechanisms, i.e., allowing an object to be simultaneously represented in several classes. Proposals exist for multi-instantiation at the object level (objects may be dynamically re-grouped to form new classes), or at the schema level (an analysis of the object life cycle determines the set of relevant possible representations, or roles, for an object).

MurMur will follow the latter stream and take it to the implementation phase. The innovative scientific contribution is a comprehensive approach that also takes into account inheritance and polymorphism mechanisms as well as data manipulation facilities. Moreover, our approach to roles does not require that a primary role is identified, a constraint that may not be consistent with the real world to be represented but that is nevertheless quite common in the proposed approaches as it makes design and implementation of the approach easier.

A second innovative aspect is that MurMur specifically contributes in advancing the state of the art in multi-resolution geographical databases. The resolution concept applies to spatial, temporal, and thematic characteristics of information. In the spatial dimension, the multi-resolution aspect relates to cartographic data presentation at different scales. In the temporal dimension it is usually understood as related to time granularity. In the thematic dimension,

multiplicity of resolution levels may materialize through definition of classification hierarchies, aggregation of objects (based on thematic, spatial, or temporal criteria), and definition of hierarchical domains for attribute values. Most of the work in multi-resolution has been done in the GIS domain, where researchers have focused on multi-scale aspects.

MurMur intends to develop a more comprehensive approach, building on similarities between the different facets of multi-resolution. In particular, MurMur will be able to provide more intelligent links among representations. Another specific target will be an elaborated scheme for aggregations, to apprehend nxm links where a group of n objects in one resolution level corresponds to a group of m objects in another resolution level, an area where little research results are known. Geometric and semantic resolution models will be dealt with.

The new paradigm will be embedded in a conceptual model for spatio-temporal data, based on the existing data model MADS [15]. The conceptual model builds on the orthogonality principle, i.e., the thematic, spatial, and temporal dimensions are independently organized so that choices in one dimension do not constrain choices in another dimension. Orthogonality provides design flexibility, hence an expressive power that is beyond those of current data models. Inclusion of multi-representation functionality will make this model one of the most advanced conceptual models for spatio-temporal data. The model is expected to play a major role in fulfilling the current gap between user expectations and requirements for user-friendly systems, and the tools that are on the marketplace.

Several tools will be implemented in order to assess the adequacy of the new multi-representation paradigm with respect to the users' need.

First, a visual schema editor will allow the definition of spatio-temporal schemas through direct manipulation of the modeling constructs on the screen. While visual schema editing is a well-established technology, its usage for spatio-temporal databases is limited to very few tools. However, these tools inherit the limitations of the underlying model, such as lack of orthogonality and limited data structuring capabilities.

Next, a visual query tool will combine direct manipulation with textual specifications and menu-driven query formulation. Developing a fully multi-paradigm query tool (e.g., including sketches, voice and gesture interactions) for spatio-temporal databases is beyond the scope of MurMur, as it would be a full project in itself due to the inherent complexity of spatio-temporal queries, in particular for

disambiguation. Still, the MurMur query facility is expected to provide significant advances to the state of the art in query tools.

Furthermore, the tools developed in the MurMur project will constitute an initial answer for supporting the process of integration of existing information sources in distributed and interoperable environments. Although MurMur does not intend to develop a full data integration approach (again, this would be a research project in itself), MurMur plans to support acquisition of information about correlated data objects in different databases. This kind of information is typically produced by object-matching processes applied to separate geographical data sources. MurMur will consider this information as a possible input to create a multiple-representation database. From this input MurMur will build an integrated view of the correlated objects and databases. This may be seen as a minimal mediation facility, whose purpose is to match user requirements on a simple, incremental basis. Such an approach is innovative for geographical databases.

References

- [1] A. Albano, G. Ghelli, R. Orsini, Fibonacci: A Programming Language for Object Databases. *Very Large Data Bases Journal*, 4(3): 403-444, 1995.
- [2] C. Claramunt, *Un modèle de vue spatiale pour une représentation flexible de données géographiques*. Thèse de doctorat en informatique, Université de Bourgogne, 1998.
- [3] Y. Dennebouy et al, SUPER: Visual interfaces for object + relationship data models. *Journal of Visual Languages and Computing*, 6(1):73-99, 1995.
- [4] T. Devogele, *Processus d'intégration et d'appariement de bases de données géographiques: Application à une base de données routière multi-échelle*. Thèse de doctorat en informatique, Université de Versailles, Institut Géographique National, 1998.
- [5] M.J. Egenhofer, E. Clementini, P. Di Felice, Evaluating inconsistencies among multiple representations. *Proceedings of the Sixth International Symposium on Spatial Data Handling, SDH'94*, p. 901-920, Edinburgh, Scotland, 1994.
- [6] A. Franck, S. Timpf, Multiple representations for cartographic objects in a multi-scale tree: An intelligent graphical zoom. *Computers & Graphics*, 18(6), 1994.
- [7] M. Gentile, *An object-oriented approach to manage the multiple representations of real entities*, EPFL PhD Thesis no 1490, 1996.
- [8] T. Jen, *Formalisation des relations spatiales topologiques et application à l'exploitation des bases de données géographiques*. Thèse de doctorat en informatique, Université Paris XI Orsay, 1999.
- [9] C.B. Jones, D.B. Kidner, L.Q. Luo, G.L. Bundy, J.M. Ware, Database design for a multi-scale spatial information system. *International Journal of Geographical Information Systems*, 10(8): 901-920, 1996.
- [10] Y. Kambayashi, Z. Peng, Object deputy model and its applications. *Proceedings of the Fourth International Conference on Database Systems for Advanced Applications, DASFAA'95*, p. 1-15, Singapore, 1995.
- [11] D. Kidner, C. Jones, A Deductive Object-Oriented GIS for Handling Multiple Representations. *Proceedings of the Sixth International Symposium on Spatial Data Handling, SDH'94*, p. 882-900, Edinburgh, Scotland, 1994.
- [12] T. Kilpeläinen, Maintenance of topographic data by multiple representations. *Proceedings for the Annual Conference and Exposition of GIS/LIS '98*, Forth Worth, Texas, 1998.
- [13] M. Molenaar, Composite objects and multiscale approaches. *An introduction to the theory of spatial object modeling for GIS*, p. 161-191. Taylor & Francis, 1998.
- [14] J.C. Müller, J.P. Lagrange, R. Weibel, F. Salgé, Generalization: State of the art and issues. In J.C. Müller, J.P. Lagrange and R. Weibel, editors, *GIS and Generalization: Methodology and Practice*, p. 3-17. Taylor & Francis, 1995.
- [15] C. Parent, S. Spaccapietra, E. Zimanyi, Spatio-Temporal Conceptual Models: Data Structures + Space + Time, *Proceedings of 7th ACM Symposium on Geographic Information Systems, GIS'99*, Kansas City, Missouri, 1999.
- [16] M.P. Papazoglou, B.J. Kramer, A. Bouguettaya, On the representation of objects with polymorphic shape and behavior. *Proceedings of the 13th International Conference on Entity-Relationship Approach*, p. 223-240, Manchester, UK, 1994.
- [17] B. Pernici, Objects with Roles, *Proceedings of ACM Conference on Office Information Systems*, p. 205-215, Cambridge, Massachusetts, 1990.
- [18] J. Richardson, P. Schwarz, Aspects: extending objects to support multiple, independent roles. *Proceedings of the ACM SIGMOD International Conference On Management of Data, SIGMOD'91*, pages 298-307, Denver, Colorado, 1991.
- [19] D. Rieu, G.T. Nguyen, A. Culet, J. Escamilla, C. Djeraba, Instanciation multiple et classification d'objets, *Actes VIIèmes Journées Bases de Données*

Avancées, France, 1991.

- [20] P. Rigaux, M. Scholl, Multi-scale partitions: Applications to spatial and statistical databases, *Proceedings of the 4th International Symposium on Advances in Spatial Databases, SSD'95*, p 170-183 LNCS 951, Springer-Verlag, 1995.
- [21] M. Scholl, A. Voisard, J.-P. Peloux, L. Raynal, P. Rigaux, *Systèmes de Gestion de Bases de Données Géographiques, Spécificités*, International Thomson Publishing, 1996.
- [22] J. Stell, M. Worboys, Stratified Map Spaces: A formal basis for multi-resolution spatial databases. *Proceedings of the 8th International Symposium on Spatial Data Handling, SDH'98*, p. 180-189, 1998.
- [23] S. Timpf, A. Franck, A multi-scale DAG for cartographic objects. *Proceedings of Auto Carto 12*, p. 157-163, Charlotte, North Caroline, USA, 1995.
- [24] S. Timpf, *Hierarchical structures in map series*, PhD thesis, Technical University Vienna, 1998.
- [25] R. Weibel, G. Dutton, Generalizing spatial data and dealing with multiple representations. In P. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, editors, *Geographical Information Systems: Principles, Techniques, Management and Applications*, vol. 1, 2nd edition Geoinformation International