MULTIPLE REPRESENTATION MODELING

Christine Parent
HEC ISI, University of Lausanne, http://lbd.epfl.ch/~cparent/christine.html

Stefano Spaccapietra, Christelle Vangenot
Ecole Polytechnique Fédérale de Lausanne, http://lbd.epfl.ch/~stefano/e-index.html

Esteban Zimányi
CODE, Université Libre de Bruxelles, ezimanyi@ulb.ac.be

SYNONYMS

Multi-scale, multi-resolution, multi-granularity modeling

DEFINITION

Geodata management systems (i.e., GIS and DBMS) are said to support *multiple representations* if they have the capability to record and manage multiple representations of the same real-world phenomena. For example, the same building may have two representations, one with administrative data (e.g., owner and address) and a geometry of type point, and the other one with technical information (e.g., material and height) and a geometry of type surface. Multirepresentation is essential to make a data repository suitable for use by various applications that focus on the same real world of interest, while each application has a specific perception matching its goals. Different perceptions translate into different requirements determining what information is kept and how it is structured, characterized, and valued. A typical use case is map agencies that edit a series of national maps at various scales and on various themes.
Factors that concur in generating different representations include the intended use of data and the level of detail matching the application's concerns. The former rules the choice of data structures (which objects, relationships, and attributes are relevant) and of value domains (e.g., whether the temperatures are stored in Celsius or Fahrenheit). The latter rules data resolution from coarse to precise and impacts both the semantic and the spatial representation.

Multiple representation modeling is the activity of designing a data repository that consistently holds multiple representations for various perceptions of a given set of phenomena. It relies on a multirepresentation data model, i.e., a data model with constructs and rules to define and differentiate the various perceptions and for each perception the representations of the phenomena of the real world of interest.

HISTORICAL BACKGROUND

Support for different requirements over the same data set has first been provided by using multiple data files, each one designed for a specific application. Aiming at consistency, databases looked instead for ways to gather data into a single repository, generating the need for multiple representations. First came facilities to define application-specific subschemas, as simple restrictions of the database schema. Later, more flexibility was achieved through the view mechanism. In object-oriented terms, views are virtual representations derived from existing data. They create either an alternative representation for existing objects (object preserving views) or new objects composed from existing objects (object generating views). Each view defines a single new representation. However, views do not provide for multiple perceptions, i.e., there is no possibility to identify the collection of views that forms a consistent whole for an application.

Similarly, the concept of is-a link was borrowed from artificial intelligence to provide for various representations of the same object in a classification refinement hierarchy. However, the concept comes with a population inclusion constraint: The subtype population is included in the supertype population. This cannot cope with situations where the populations of two object types only overlap, i.e., the two populations may have specific objects not represented in the other population. Since then, the situation has basically not changed. It is only in the last decade that the need for more flexible multirepresentation and explicit support of various perceptions has been stressed by researchers.

Multirepresentation research in spatial databases can be traced back to the 1989 NCGIA program. In the early eighties, maps began to be stored in geographic databases, and that opened up many new research tracks. The specification and implementation of cartographic generalization was one of them. It relied on the idea that cartographic generalization would allow the automatic derivation of maps at different scales from a single geographic database. Realizing that this full automation was not possible to achieve increased the focus on multiple representation databases [7][10][15] . Under the pressure of delivering maps at different scales, the first researches were, in the early nineties, focussing on multi-scale data structures (i.e., data structures allowing to retrieve geometry of real-world objects for any given scale) and multiscale databases (i.e., databases in which the data for maps at different scales is stored and linked together). Later, the multiscale approach was refined and extended into the multirepresentation approach: scale indeed is not a concept relevant for databases and there is nowadays more to a geographical database than producing maps. The current scope of multirepresentation for GIS database comprises various techniques that can be used, within a single database, to automatically derive a coarser representation from another representation at finer resolution. These techniques aim at computing objects at coarser representations, for multiresolution analyses rather than for display. They include generic cartographic generalization operations (not driven by map display considerations) as well as operations performed on specific object types (e.g., selection of instances based on an ad-hoc predicate, aggregation of instances to create new objects). Some authors use the term model generalization to denote that the use of various techniques leads to the creation of a set of virtual databases for different resolution levels, and the mappings between their schemas (models in GIS terms). The automatic derivation rules (the mappings) allow update propagation from finer to coarser representations.
Currently, only a few simple multi-representation databases exist and are used to their expected potential.

SCIENTIFIC FUNDAMENTALS

Early research on multi-representation in GIS was driven by cartographers' requirements. This explains why the ability to draw maps at different scales has been for long the targeted objective, popularizing the concept of *multi-scale databases*. However, many other spatial applications that need to perform spatial data analysis require storing and managing specific representations where objects may have various geometries (derived one from another or not) and also show varying thematic characteristics (e.g., have different attributes and different relationships). Thus, the research domain evolved from multi-scale databases to *multi-representation databases.* Equally important is the capability to provide each application with a consistent set of data that corresponds to its own perception of the real world of interest, in short its own database. Therefore, support of multirepresentation should be complemented with support of multiperception.

Multiscale databases

Multiresolution databases are still referred to by the GIS community as multiscale databases, despite the fact that scale does not apply to data storing. Scale is a concept related to the drawing of maps on paper or on screen. It is the ratio between measures on a map and the corresponding measures in the real world. Scale only characterizes an intended use of data. Instead, the level of resolution of a spatial database determines what geometries are stored. It defines a threshold such that only geometries beyond the threshold are captured and stored.

In early work by Timpf, the different map representations of the same real-world entities are interconnected using a directed acyclic graph data structure. The graph allows users to navigate among maps at different scales by zoom-in and zoom-out operations. This work later developed into a more general Map Cube Model [14], supported by a theory on the structure of series of maps of the same region at different scales. Each map is described as a composition of four components: a set of lines representing transportation and hydrology networks, the set of areas, called containers, created by the lines of the trans-hydro network, areas that are a refinement of the container partition, and objects contained in these areas. The elements stored in each of the four components (e.g., streets, land-use areas, buildings) are then organised into aggregation and/or generalization hierarchies. This forms a graph for each component, where each level of the graph corresponds to a given scale.

Stell and Worboys have also proposed a solution to link a series of maps [13]. Their database organization is called a stratified map space. Each map gathers objects of a particular region that share the same semantic and spatial granularity. Maps are grouped by map spaces, i.e., sets of maps at the same granularity, describing various regions. The stratified map space is the set of all maps spaces organized according to a hierarchy based on different granularity levels. Transformation functions allow users to navigate in a stratified map space and propagate updates.

Multi-representation databases

Work on spatial multi-representation databases has followed two main tracks: either proposing new (conceptual) data models that include explicit description of multi-representation, or proposing frameworks that organize a set of existing classic (i.e., without multirepresentation) databases into a global multirepresentation repository.

Database models for multiple representations

Several data models with specific concepts for multiple-representation modelling have been proposed. They range from simple solutions allowing users to associate various geometries to the same objects to more sophisticated solutions. They are discussed here according to the requirements for multiple-representation modelling.

A model for multirepresentation should allow one to characterize the same objects using different sets of attributes, and attributes with different values and different domains. This flexibility is supported by the MADS model [8], where multiple representations of a given phenomenon may be organized according to two strategies. In the first one, the various spatial and semantic descriptions of the same real-world phenomenon are merged into a single database construct. Each element of a description is qualified by a tag (called stamp) whose value identifies the perceptions for which it is relevant. Object and relationship types can thus have various sets of attributes depending on the perception. Attributes may bear various cardinalities or value domains according to the perception stamp; they can also contain a value that is a function of the perception stamp.

The Vuel approach [2] also offers the possibility of associating various semantic and spatial descriptions to the same real-world entities. In addition, various graphical representations, useful for drawing maps at different scales, can be defined. The data model is a snowflake model for spatial data warehousing. The fact table is composed of a specific kind of tuples, called vuels. A vuel fact is a particular representation of a real-world entity. It has three components: a geometry, a graphical description, and a semantic description. The vuel representation may vary according to these three dimensions. Moreover, the semantic dimension is a fact table itself with four dimensions: the class, the attribute, the domain of value, and the value dimensions. This allows the creation of various semantic descriptions by combining the dimensions (different classes, different sets of attributes, attributes with various domains and various values). OMT-G [3], an UML-based model, supports the modelling of multiple representations of data through a specific kind of relationship called conceptual generalization relationship. This relationship allows the definition of various views of the same real-world entities as subclasses of a shared super-class. The superclass describes the thematic attributes that are common to all the representations and it has no spatial representation. Each subclass describes its own view by specifying its own thematic and spatial attributes. The subclasses inherit the common attributes from the superclass. A presentation diagram shows graphical representations that may be associated to a class and the operations to obtain them. MRSL [5], another UML-based model, supports multi-representation through the introduction of two concepts: representation objects (r-objects) and integration objects (i-objects). All r-objects corresponding to the same real-word phenomenon are linked by a monovalued or multivalued link to a single i-object, whose role is to ensure consistency among them. Each r-object specifies a specific set of attributes and values for the same real-world object.

Multiple representation modelling is not limited to associating multiple sets of attributes or values to one object. In particular, when changing the level of detail, objects may disappear, whereas others may be grouped. Thus, in addition, there is the need to put into correspondence one object with several objects or two different sets of objects.

In the second strategy of the MADS approach, the various descriptions of the same phenomena belong to separate object types. They can be linked by inter-representation links that are either traditional associations or multi-associations. Multi-associations are binary relationships that, contrarily to association relationships, do not link two objects but two groups of objects. A multi-association is needed whenever the real-world entities are not represented per se, but through two different decompositions; e.g., a decomposition of a road in segments according to the number of lanes, and one according to crossroads. The other modelling approaches only support correspondence links of kind association, and the supported cardinality of the link varies: in MRSL, a i-object can be linked to r-objects through 1:1 and 1:N links thus providing support for the 1:N and N:M correspondences. In Vuel, corresponding objects can also be linked through 1:1 or 1:N inter-representation links. However, there is no support for N:M inter-representation links. OMT-G does not support inter-representation associations between objects.

Not only objects may need multiple representations, also relationships do. This is only supported by MADS. In MADS all characteristics of a relationship may have various representations: its semantics (e.g., topological, aggregation, or plain), its roles, and its cardinalities. For instance, a relationship type can be a topological adjacency relationship in one description and a near relationship in another one with a more precise resolution.

In the spatial context, as data from one representation may often result from the derivation of the same data represented at another resolution, the representations of the same real-world entity are not independent and one may expect to be able to state constraints between these representations. Consistency constraints in databases are maintained through the definition of integrity constraints. Some constraints, such as cardinalities, are embedded in the concepts of the model – in particular some constraints are inherent to the multiple-representation concepts – while other constraints need to be defined in the application. MRSL is the only model proposing specific multirepresentation constraints: three kinds of rules can be associated to an i-object and its linked r-objects: consistency rules, which can be object or value correspondences, matching rules, and restoration rules. Matching rules specify how to match objects representing the same entity. They can be attribute comparison, spatial match operations, or global identifiers. Restoration rules are used to restore consistency between an i-object and its r-objects when needed.

Finally, considering that a multirepresentation database contains several representations of the same real-world phenomena, it is important to associate metadata to the representations to identify the application(s) they are relevant for but also in order to know which representations together form a consistent whole for the application. This important requirement is fulfilled by MADS through the concept of perception stamp. In

MADS, a perception stamp is a vector of values (e.g., a viewpoint and a resolution) that identifies a particular perception, and all elements of the database (types, properties, instances) are stamped for defining for which perception they are relevant. In Vuel, the designer can define views that are compositions of vuels. Each view defines a particular perception, thus providing a functionality similar to perception stamps.

Architectures for distributed representations

Instead of proposing new concepts allowing users to integrate multiple representations of the same real-world phenomena into a unique multirepresentation database, other proposals followed a less intrusive approach. Capitalizing on the fact that there already exist many spatial databases, these approaches create a multirepresentation framework out of a set of existing classic (i.e., describing a unique perception and resolution) databases. There are two main kinds of proposals: the first one focuses on the *definition of links between objects* in corresponding databases, the second one aims at building *federated database management systems*.

In the first category, the work of Kilpelainen [6] was one of the first proposals tackling multiple representations from a database point of view. It supports bi-directional links that allow one to propagate updates in both directions and perform reasoning processes in the form of generalization operators.

In federated spatial databases, users access a set of databases through a single integrated schema, which describes virtual multirepresentation objects. During query processing, multirepresentation objects are dynamically constructed by merging all the corresponding monorepresentation objects that exist in the various databases. There have been several proposals for spatial database integration [4]. Particularly interesting are those that build the integrated schema using multirepresentation concepts, e.g., [5], based on MRSL, and [12], based on MADS. Using MRSL, each r-object in the integrated schema holds an UML tag that identifies the corresponding source database. Using MADS, perception stamps can fulfil the same functionality.

KEY APPLICATIONS

Cartography

As they cannot automatically derive maps at different scales from a single detailed database, national map agencies have to create several databases, one per scale. For them, multirepresentation modeling is crucial for two main reasons:

- To propagate updates [1]: the cost of updating can be lowered by entering updates only once in a database and propagating them, at least semi-automatically, to the other databases.
- To enforce consistency [11]: multi-representation databases play an important role in order to enforce consistency between the same data described at different levels of details. In addition, integrating existing databases to create a multi-representation database allows one to detect inconsistencies between the databases.

Multi-scale analysis:

Multirepresentation databases can benefit many applications that need to analyze data at different levels of details or defined for different viewpoints. For example, a fire monitoring application may need very detailed data on current fires (to direct the action of fire brigades as precisely as possible), only need medium-level resolution data for records of past fires, and use low-level resolution data for generic organization of fire management activities.

Other candidate applications are those relying on spatial data warehouses, using spatial OLAP and spatial data cubes to perform multi-dimensional analysis. An example is traffic accident monitoring applications, e.g., for analysis of the number of deadly accidents according to multilevel criteria (by road, region, department, or state). Multirepresentation storage of spatial data is needed in order to drill-up and drill-down the cube [2].

FUTURE DIRECTIONS

Work in progress explores the use of multirepresentation capabilities in support of modularization of knowledge repositories. In particular, the semantic web community is developing various approaches to turn huge ontologies that are being built in several knowledge domains into smaller sets of more manageable ontological modules. Existing approaches follow both the integrated direction (a single ontology is modularized) and the distributed direction (various existing ontologies are interconnected within a global knowledge sharing system). A forthcoming book on Ontology Modularization [9] is due for publication in 2008.

CROSS REFERENCES

Data Models, Database Design, Multidimensional modeling, Geographic Information System, Spatial Data Types, Topological Relationships, Field-Based Spatial Modeling, Topological Data Models, Semantic Modeling for Geographic Information Systems.

RECOMMENDED READING

[1] Badard, T., Lemarié, C.: Propagating updates between geographic databases with different scales. In Atkinson, P. and Martin, D. (eds.), *Innovations in GIS 7: GIS and GeoComputation*, chapter 10, pp. 135–146. Taylor and Francis, London, UK (2000)

[2] Bédard, Y., Bernier, E.: Supporting Multiple Representations with Spatial View Management and the Concept of VUEL. *Proceedings of the Joint Workshop on Multi-Scale Representations of Spatial Data*, ISPRS WG IV/3, ICA Com. on Map Generalization, Ottawa, Canada (2002)

[3] Borges K., Davis, C.A., Laender, A.: OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica*, 5(3), pp. 221–260 (2001)

[4] Devogele, T., Parent, C., Spaccapietra, S.: On Spatial Database Integration, *International Journal on Geographical Information Systems*, 12(4): 335–352 (1998)

[5] Friis-Christensen, A., Jensen, C.S., A., Nytun, J. P., Skogan, D.: A conceptual schema language for the management of multiple representations of geographic entities, *Transactions in GIS*, 9 (3): 345–380 (2005)

[6] Kilpelaïnen, T.: Maintenance of topographic data by multiple representations. *Proceedings of the Annual Conference and Exposition of GIS/LIS*, pp. 342–351, Forth Worth, Texas, November 10-12, (1998)

[7] Mustière, S., Van Smaalen, J.: Database requirements for generalisation and multiple representations. In: Mackaness, W.A., Ruas, A., Sarjakoski, T., *Generalisation of geographical information: Cartographic modeling and applications*. Elsevier (2007)

[8] Parent, C., Spaccapietra, S., Zimányi E.: *Conceptual Modeling for Traditional and Spatio-temporal Applications. The MADS Approach*. Springer (2006)

[9] Parent, C., Spaccapietra S., Stuckenschmidt, H.: *Ontology Modularization*, Springer, (2008, to appear)

[10] Sarjakoski, L.T.: Conceptual models of generalisation and multiple representation. In: Mackaness, W.A., Ruas, A., Sarjakoski, T., *Generalisation of geographical information: Cartographic modeling and applications*,pp. 11–36. Elsevier (2007)

[11] Sheeren, D., Mustière, S., Zucker, J.D.: How to Integrate Heterogeneous Spatial Databases in a Consistent Way? In: Gottlob, G., Benczúr, A.A., Demetrovics, J. (eds.), *Proceedings of the 8th East-European Conference on Advances in Databases and Information Systems*, pp. 364–378, Budapest, Hungary, September 22–25. Lecture Notes in Computer Science, vol. 3255, Springer, Berlin, Heidelberg (2004).

[12] Sotnykova, A., Vangenot, C., Cullot, N., Bennacer, N., Aufaure, M-A.: Semantic Mappings in Description Logics for Spatio-Temporal Database Schema Integration. In: Spaccapietra, S., Zimányi, E. (eds.), *Journal on Data Semantics III*, pp. 143–167, Lecture Notes in Computer Science, vol. 3534. Springer, Berlin, Heidelberg (2005)

[13] Stell, J.G., Worboys, M.F.: Stratified map spaces: A formal basis for multi-resolution spatial databases. In: Poiker, T.K., Chrisman, N. (eds.), *Proceedings of the 8th International Symposium on Spatial Data Handling*, pp. 180–189, British Columbia, Canada. Taylor and Francis, London, UK (1998)

[14] Timpf, S.: Map Cube Model: A model for multi-scale data. In: T. Poiker, T.K., Chrisman, N. (eds.), *Proceedings of the 8th International Symposium on Spatial Data Handling*, pp. 190–201. British Columbia, Canada. Taylor and Francis (1998)

[15] Weibel, R., Dutton, G.: Generalizing spatial data and dealing with multiple representations. In: Longley, P., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*, vol. 1, 2nd edition, pp. 125–155. Wiley (1999)