Pestana, G., M. M. da Silva & Y. Bédard, 2005, Spatial OLAP Modeling: An Overview Base on Spatial Objects Changing over Time, IEEE 3rd International Conference on Computational Cybernetics, 13-16 avril, Mauritus, accepted for publication

Spatial OLAP Modeling: An Overview Base on Spatial Objects Changing over Time

Gabriel Pestana
INESC-ID
Rua Alves Redol nº 9
1000-029 Lisboa
PORTUGAL
gabriel.pestana@inesc-id.pt

Miguel Mira da Silva
IST/DEI
Av. Rovisco Pais
1049-001 Lisboa
PORTUGAL
mms@dei.ist.utl.pt

Yvan Bédard
Center of Research in Geomatics
Laval University
QC G1K 7PA Quebec
CANADA
yvan.bedard@scg.ulaval.ca

Abstract – **Enabling the decision making process to support spatial queries is not a trivial task. This task becomes even harder when using Geographic Information Systems (GIS) with Data Warehouse (DW) because these two technologies are in general used separately. In general, a GIS solution handles spatial data without considering time constrains or without requiring the analyses of geometric shapes evolving over time. Temporal maps further raise difficulties on indexing issues and on the associated query mechanisms. On the other hand, a typical DW operates with non-spatial data for different time periods. It also does not support spatial data types, such as point, lines, and polygons.**

In this paper, we propose a multidimensional spatiotemporal data model to enable spatial analysis, in a context of evolving specifications. The proposed data model addresses the problem of spatial and temporal data integration by providing information to facilitate semantic interoperability and data analysis in a spatial DW that uniformly handles all types of data. Using a practical example in the field of land parcels, we evaluate the implementation of the model.

I. INTRODUCTION

With the increasing amount of spatial data stored in business information systems, spatial data are becoming a corner stone for decision-makers to analyze business data in a spatial context. With the advent of mobile computing and location-based services, even more amount of spatial data is being collected and stored. While storing spatial data is the first step, what is more important is to look to those data from a business geo-spatial point of view.

However, traditional GIS applications still separate business data from spatial. This loosely coupled approach has a data integrity drawback because the two types of data are managed apart [1].

On the other hand, traditional DW and On-line Analytical Procedures (OLAP) tools cannot exploit spatial data because current multidimensional database technologies do not support the spatial data structures [2]. Moreover spatiotemporal query windows usually do not conform to pre-defined hierarchies. Consequently spatiotemporal data do not always have a natural hierarchy that can be used at design time to efficiently compute pre-aggregation in the DW [3].

One alternative is to follow a data-centric approach to store spatial and non-spatial data within the same database kernel. This approach preserves all of the database's core capabilities including indexes, triggers, and functionalities to work with spatial data. Emerging database products, such as Oracle Spatial [4], extended their DBMS to the spatial domain allowing the database to store spatial features as other data than integers, strings, dates, etc. inside the database [5, 6]. In this way, spatial features might be stored with their corresponding points (e.g., addresses), lines (e.g., roads), and polygons (e.g., land parcels).

Nowadays, spatial data types are already well defined [7], enabling existing applications to take advantage of the spatial capabilities provided by spatial databases (e.g., scalability, security, and replication).

However, building a multidimensional database for spatial analyses is still a challenge, mainly when the data acquisition and specifications evolve over time [8].

This difficulty drives from spatiotemporal data analysis (also known as spatial OLAP queries) requirements. This happens because databases sources differ from one epoch to another not only in data coding and structures, but also in data semantic contents and geometric evolutions. In order to permit spatiotemporal comparative studies, data must be integrated following a compatible set of temporal, spatial and semantic definitions.

When analyzing historic data, spatial OLAP queries consider many parameters in the geoprocessing workflow, making it very difficult for decision-makers to keep track of all the facts, assumption, and datasets of existing values.

The proposed data model uses a visual modeling tool, named Perceptory [9], to support spatiotemporal data modulations. We are particularly interested in modeling geometric shape evolutions over time. Our approach preserves the traditional star schema [10] while bringing new spatial OLAP capabilities into the decision process.

The rest of the paper is organized as follows: Section 2 refers to related work on Spatial DW. Section 3, presents an overview of stereotypes and describes Perceptory graphical notation for spatial data modeling. Section 4, presents the problem we are trying to solve. Section 5, describes our proposal to solve the problem. Section 6 presents a practical example, and Section 7 concludes the paper, outlining how the model will be extended.

II. RELATED WORK

Although the multidimensional model is widely known it is seldom used in spatial data modeling. A multidimensional model is usually represented as a star schema [10]. The star schema is centered on a main table F (called fact table) which stores a multidimensional array of measures, M, while dimensional tables D1, D2,…, Dn store the details of those measures. These dimensions usually represent business perspectives, such as customers, products, and time. A row in F has the form <Di[ ].key, M[ ] > where Di[ ].key is the set of foreign keys to the dimension tables and M[ ] is the set of measures. A dimension is usually organized in hierarchies supporting levels of data aggregation as well as multiple inheritances.

A multidimensional spatial data model also has facts and dimensions providing the descriptive characteristics of spatial features that bring spatial OLAP analysis to life. This means that spatial measures could also store in the fact table. In the same way the spatial dimensions (in addition to the usual descriptive attributes) could store attributes of a spatial data type, enabling spatial operations such as overlap, contain, intersect, merge, or split.

Most of these issues are not new and some have already been individually addressed in other papers. In [11] the authors present some pioneer work in this area by proposing a spatial DW framework to select spatial objects for materialization. They also propose the idea of spatial measures defined as pointers to regions in space. The use of a spatial star-schema is also proposed by [12], with methods to process arbitrary aggregations. In [20], the building of spatio-temporal topological operator dimensions (STTOD) is proposed. These cases, however, do not cover ad-hoc queries. An ad-hoc query not confined by the hierarchy would still need to access the fact table, even if the entire spatial DW was pre-aggregated.

In [13] the research was on methods to achieve the best data approximation within a pre-specified range of data. The focus was on finding the minimum partition of a region. In [17] a spatial DW modeling framework is presented. This work distinguishes different types of measures and dimensions. It highlights important aspects when implementing a spatial DW with commercial off-the-shelf (COTS) tools, namely how to implement a multidimensional spatial data model without paying significant syntax or user interface penalties. Several other proposals [2, 11, 12, 14] focus specialized indexes, and aggregation techniques to manage high volumes of spatial data as well as Spatial OLAP benefits [21].

III. PERCEPTORY SPATIOTEMPORAL NOTATION

In this section we present the significance and convenience of representing spatial data in a multidimensional model, with a focus in spatial dimensions and spatial measures.

In the spatial data model design process we have to consider how to model spatial features, their geometric shape evolutions over time, and how to generalize/aggregate those features into higher levels of abstraction.

Spatially-aware languages have become a recent research focus in academia, industry, and standardization bodies such as the OpenGIS Consortium [18] and IS0 TC211 [19], to extend UML for spatial database development. Following this trend we use Plug-ins for Visual Language (PVL) notations [9] to facilitate spatial and temporal modeling.

The use of stereotypes (icons) as a graphical notation to depict geometry shapes and temporalities simplifies the modeling diagram. Table 1 presents the basic PVL graphical notation defined by Perceptory [9].

Perceptory is an extension to UML based on stereotypes that adds support for spatiotemporal properties that are aligned with the ISO-standards for geographic information.

Table 1. The three Perceptory basic pictograms for a 2D universe.

Grammar Notation	**Description**
Point, 0-dimensional shape	, the pictogram records a date and a period (begin/end dates)
Line, 1-dimensional shape	,
Polygon, 2-dimensional shape	,

The selection between or depends on the temporal granularity defined into the repository for each class, attribute and geometry. It then becomes possible to keep the various values that object features (attributes) have during their lifespan. The temporal pictograms are useful for instance to model spatial features with instantaneous existence like "Fire", or features with a durable life like "Replanting Rights".

Pictograms may be mixed to represent more complex geographical notations, as presented in Table 2.

Table 2. Examples of the PVL Graphical Notation.

Pictograms	**Description and Examples**
	Complex shape (e.g., an hydrographical network composed of 1D river and 2D lakes)
	Alternative Shape (e.g. Parcel having a 0D shape if < 1 hectare or a 2D shape if > 1 hectare)
	Multiple shapes for an instance (e.g., polygonal municipalities having a non-derivable point located downtown
	Derived shape (example for a municipality centroid derived from other geometric information, i.e. the municipality polygon)

On the other hand, geometric evolution of objects involves shape modifications and/or displacements. As presented in Figure 2, All types of parcels have their geometry and a derived attribute "area" while only Vineyard Parcels have geometric evolution. Both types of parcels have an existence.

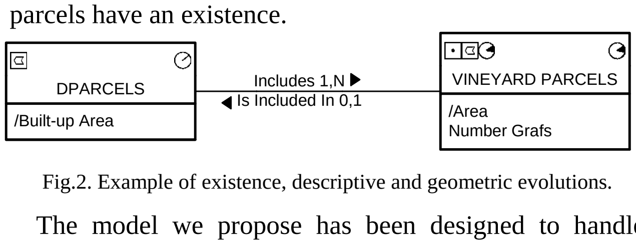

Fig.2. Example of existence, descriptive and geometric evolutions.

The model we propose has been designed to handle these aspects, namely to depict geometry changes over time.

IV. THE PROBLEM

To integrate and model spatiotemporal data, several inter-related problems appear. For instance, in a spatial DW the topological relationships can include more than two spatial dimensions [14]. This n-ary topological relationship is not a trivial task because it establishes explicit and multiple hierarchies in dimensions.

Explicit hierarchies are used for aggregating data to the right level of detail in exploratory analyses that use roll-up/drill-down operations [15, 21]. Conversely, support for multiple hierarchies means that multiple aggregation paths are possible. These concepts are important to enable better handling of the imprecision in queries caused by partial containment in spatial dimension structures (e.g., districts would, though approximately, roll up to cities).

Indeed, a huge amount of work has been done during the last few years to formally define the notion of imprecision of spatial data, in particular in the reasoning research area [16]. In this paper, we do not follow this approach, although we take into account domain semantic dependency in the general design of spatial granularities.

Furthermore, the spatial measure (e.g., shape and location) may change over time [8]. The spatial dimensions may also be volatile, i.e., the regions at the finest granularity may evolve over time. Considering vineyard evolution for each Wine Region as an example, it is logical that vineyards cultivation area change according to buy/sell operations, or that refunds payments depend on the damages caused by destructive burning or by extreme meteorological conditions.

This dynamic behavior further complicates the development of spatial data models to support spatial OLAP queries because it requires some elaborate integration of spatial and temporal structures to model the existence and evolution of spatial objects over time.

V. PROPOSAL

We propose a multidimensional spatial data model that explores the spatial and temporal functionalities provided by a spatial DW. Our data modeling approach uniformly integrates both spatial and non-spatial data and differentiates itself from traditional data modeling.

By "integrates," we mean not only storing spatial and non-spatial data in the same database, but also to persistently store topologies concerning each geometric change over time. This approach benefits spatial OLAP queries such as: "show me the evolution on the geometry and total number of objects in the regions intersecting a query window qs during a time interval qt".

We consider that the query window (qs) defines a fence or polygon in the space to request information about features contained in, adjacent to, or overlapping a specified area [4]. Therefore, a typical spatial OLAP query corresponds to a query window that searches all features within a defined area or spatio-temporal context.

On the other hand, temporal queries (qt) may involve relationships such as adjacency, connectivity, or containment and distributive functions (e.g., count, sum, union) of objects for each timestamp.

In our data modeling approach, we extend the traditional star schema to the spatial domain by inserting standard spatial data types and spatial processing capabilities into the spatial DW. We also exploit spatial databases native aggregate functions to summarize geometry objects and reduce the complexity of the underlying queries.

We are mostly engaged in using spatial OLAP queries to evaluate, for instance, the amount of damages caused by a fire on a land parcel or more typically to respond to the problematic of monitoring an oil leakage on the sea.

Spatial Dimensions

We have adopted Han's definition of spatial dimension [2], whose primitive level and its entire high-level generalized are also spatial. For example, decision-makes may want to analyze refund payments to Grape-Growers concerning fire incidents, aggregated by county or by wine regions. This requires providing decision-makers the possibility for them to roll-up from vineyard parcels to a more general wine region level.

Notice that a spatial level, such as county, may have more than one way to be generalized to high-level concepts, and the generalized concepts can be spatial, such as map representing larger regions, or non-spatial, such as area or general description of the region [22]. As a consequence a spatial dimension may aggregate different spatial levels. We define a spatial level as a level for which we need to keep its spatial characteristics.

Spatial levels also relate to each other with topological relationships between spatial components, such as contain, equals, intersects, overlaps, etc. [14]. Therefore it is possible to overlap spatial dimensions rows based on the spatial level they describe.

Our spatial dimension modeling approach approximates Kimball's type II dimensions [10] definition because each update is stored as a new record with topological relationships (i.e. tuple versioning). Nevertheless, we need to be careful to avoid overcounting because we may have multiple rows in the spatial dimension for the same spatial feature. We use a most recent row indicator to do counts based on the most up-to-date descriptive values for a spatial object (e.g., vineyard parcel). We also employ effective and expiration dates to deal with spatial features counts at a given historical point in time and make use of two techniques to deal with rapidly changing dimensions. The first one is for browsing and tracking changes of key attributes in changing dimensions. This is accomplished by breaking off one or more minidimensions from the dimension table, each consisting of small clumps of attributes that have been administered to have a limited number of values. This is the case of DVINEYARDS dimension in Figure 3. The second technique uses variable depth hierarchies. The representation of an arbitrary spatial hierarchy is an inherently difficult task [7, 11, 14] in a relational environment. For example, decision-makers may want to monitor changes to vineyard parcels derived from buy/sell operations. Thus, in Figure 3, the DPARCELS spatial dimension rows can play the role of parent as well as child based on a downward or upward date operation.

To handle with unpredictable hierarchies we inserted a bridge table between the DPARCELS dimension and the fact table, as depicted in Figure 3. The bridge table contains one row for each spatial object change. Each row contains the parcel key of the parent roll-up entity, the parcel key of the derived entity, the number of levels between the parent and the derived entity, a bottom-most flag that identifies a derived entity with no further nodes beneath it, and a top-most flag to indicate that there are no further nodes above the parent.

Table 3. Spatial Dimension Auxiliary Attributes.

Attribute Name	Description	Example
Cartographic Parameters		
ScaleFraction	Scale value, i.e., representative fraction within which the spatial object is valid.	[1:1000, 1:100]
ProjectionDec	Projection description.	--
ProjectionType	Type of projection.	WGS 84
Coordinate System	Specifies the coordinate system (geographical or projected) that applies to the positions of the objects on the earth's spherical surface	--
MapUnits	Indicates the map measure unit	meters
Generic Attributes		
Feature Type	Specifies the dimensionality of the spatial data, including multipoint, multiline, multipolygon or 3D objects	0D, 1D, 2D or 3D
Feature Category	Defines business data nomenclature to help in the ordering/classification of spatial data	
Feature Status	An optional flag to simplify query performance	
Edition Date	The edition date is a timestamp field	
Version	Useful for multiversion proposes	
Flag	A most recent row indicator, helpful to do counts based on the most up-to-date descriptive values	

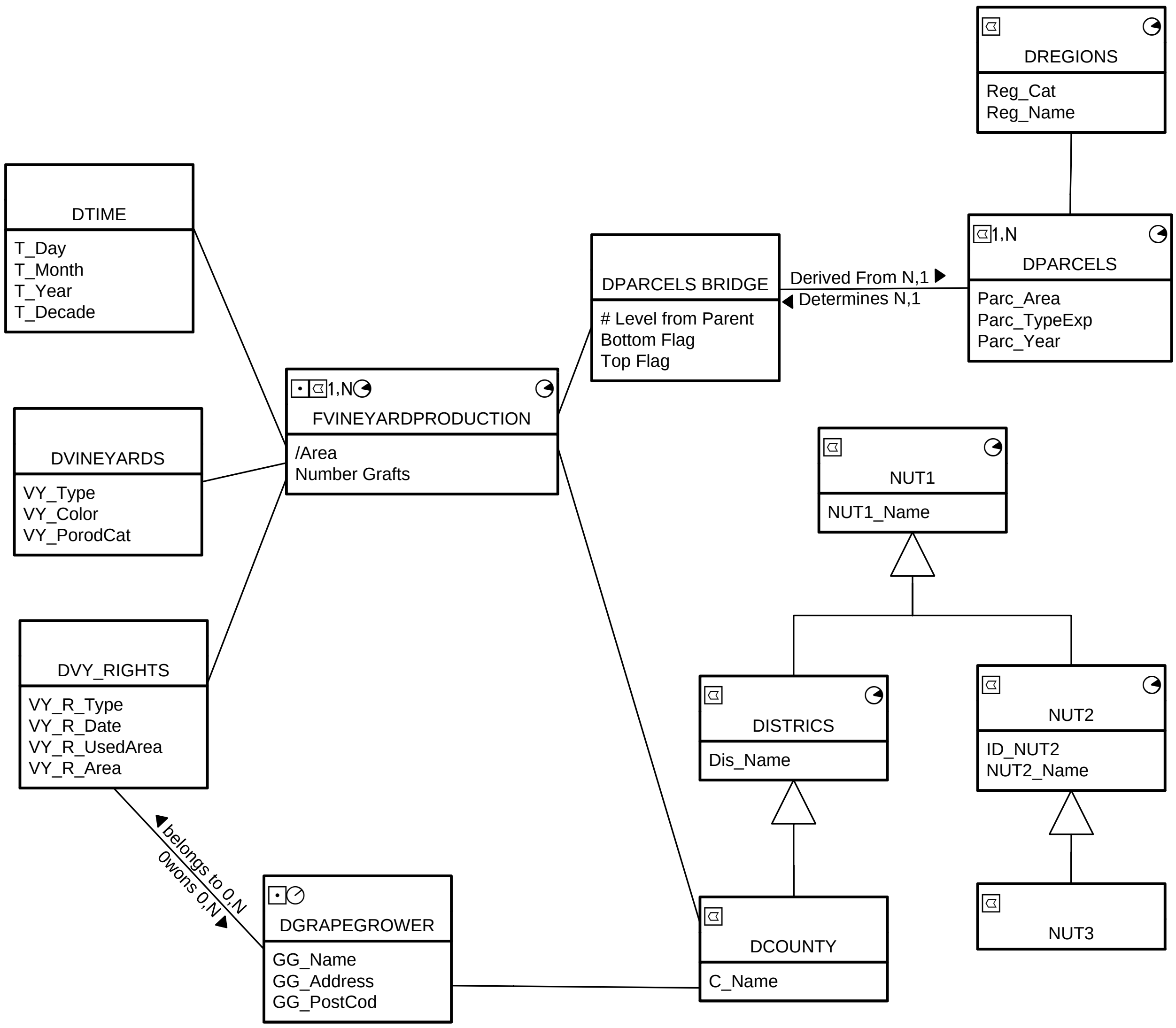

Fig. 3. A Simplified Version of the IVV Star Schema.

Spatial Measures

We distinguish two types of measures in the spatial DW:

Numerical measure is a measure containing numerical values. For example, a typical DW measure is specified by monthly revenue of a region, and a roll-up operation may get the total revenue by year, by county, etc. These numerical measures can be further classified into distributive, algebraic, and holistic [11]. A measure is distributive if it can be computed by partitioning the DW, such as count, sum, max; it is algebraic if it can be computed by algebraic manipulation of distributed measures, such as average, standard deviation; otherwise, it is holistic, such as median, most_frequent, rank.

Spatial measure is a measure that either (1) is represented by geometry with a spatial function used for aggregation along the hierarchies, or (2) represents a numerical value that is calculated using spatial or topological operators. For example, during a generalization (roll-up) operation, the regions of the same area might be grouped forming a spatial measure compose of a multi-polygon spatial data type.

Spatial measures can be associated to a fact relationship, independently if the relationship is spatial or not. We must define how spatial measures should be determined during roll-up and drill-down operations [14] because the output of these operations can be simple or complex geometries.

VI. CASE STUDY

The Wine and Vineyard Institute (IVV, in Portuguese) has collaborated in our research work as an end-user. IVV is mainly focused on wine-policy making and seeing to the application of the instruments required to increase the competitiveness of Portuguese wines.

The queries that we present in this section cover three scenarios for spatial OLAP queries: when there are no predefined hierarchies, when there is a need to support multiversioning and when there is a need to improve spatial OLAP queries performance.

The first example concerns queries involving a fixed spatial dimension, i.e., there is a static set of vector features such as parcels polygons. On the second example, the spatial dimensions may be volatile, i.e., the regions at the finest granularity may evolve over time. In the third example, changes to the geometric shape are seen as spatial measures.

These queries bring out the difference between traditional and spatial DW. In the former, OLAP results are often shown as summary tables of text and numbers, whereas in spatial DW the data are organized as collection of maps or geometric shapes.

DATA MODEL

The geometric shapes might be stored as a descriptive attribute in a spatial dimension or as a spatial measure in the fact table. When stored as a dimension attribute, it corresponds to a vector feature which might be used to filter, sort, and group the query results. Therefore, used by queries involving descriptive attributes of spatial dimensions. On the other hand, when the geometric shape is the subject of the spatial analyses it is computed as a spatial measure and hence stored in the fact table.

It is also expected that decision-makers may need to drill-around (zoom-in and zoom-out using a GIS nomenclature) dimensions hierarchy, requiring the overlay of multiple thematic maps and causing spatial data to be expressed at different granularities. Decision-makers may want to backtrack parcels changes, for instance, to monitor/validate parcels derived from split or merge operations.

Figure 3 shows a simplified version of the IVV multidimensional spatial data model, in which the spatial dimensions may change according to buy/sell operations, or Grape-Grower personal data updates originating the addition of a new row to the dimension. The DGrapeGrower is modeled as a spatial dimension because its data are presented as points on the map.

The data model contains a Time dimension, a Vineyard dimension (describing the types of vineyards), a Parcel dimension, a Grape-Grower dimension and a fact table that stores the area and number of grafts for each vineyard parcel. The fact table also reports changes to the geometric shape of spatial phenomena over time.

The DRegion is a spatial dimension that stores administrative descriptive attributes. Although it could be modeled as a non-spatial dimension, we explicitly present it as a spatial dimension to allow us to spatially roll-up history based on the current assignments.

In our approach we include in each spatial dimension additional descriptive attributes corresponding to the cartographic parameters; these attributes are resumed in Table 3.

EVALUATION

The IVV decision-makers usually need to monitor vineyard parcels' boundaries evolution. This type of information requires a multiversioning search, usually expressed on the following form:

> *What was the evolution in vineyards parcels derived from new or replanting rights being applied over the last decade?*

This query operates directly over DParcels, requiring only a multi-tree index structure to access the whole history (last decade) of geometry shapes which overlap the search region. This is particularly true for long periods covering a large geographic area.

On the other hand if decision-makers intend to monitor incidents caused by fire they may use a map-based layout to visually backtrack for patterns. A typical query would be:

> *What was the evolution of burned areas on vineyard plantation from 1990 to 2000 at Alentejo and Terras do Sado? Where did it start and through where did it spread?*

Although this query could be answered by using SQL-92 statements, it would be very expensive to check and enforce. Furthermore, topological relationship among spatial objects would be neglected making it difficult to cross-examine spatial and non-spatial data over maps and use spatial functionalities to seamlessly discover geographical patterns.

IMPLEMENTATION

The multidimensional spatial data model was implemented on Oracle Spatial. We use the spatial data types provided by the Oracle SDO table structure.

Each change to spatial data in vector format (e.g., geometry shape changes) is stored as an SDO_GEOMETRY [4] type in a single row. In this version raster data (e.g., satellite imagery, scanned topographical maps, or DTM grid data) are stored in the DBMS as descriptive attributes.

Once the database is implemented, the access to the stored data is not restricted to a certain GIS-package (like ArcGIS in our case) but is obtainable via any Oracle compatible system (e.g. GeoMedia, MapInfo, ArcView). We use ArcGIS 9 .Net as the development platform because it offers a wide and flexible range of capabilities for a systematic analysis of the available data.

We opted for COTS applications because they present some advantages over prototypes that are developed from scratch, namely reliability and features. In addition, COTS applications constitute an excellent starting point to expand software capabilities and improve system performance for the commercial marketplace. Consequently, prototypes based on commercial applications can usually be utilized to solve real-world problems, such as the IVV case study. At the time of publishing this paper, the first commercial Spatial OLAP running on top of COTS should be out and further save months of development (cf. 3rd author).

VII. CONCLUSION

In this paper we presented a practical example in the field of land parcels, to evaluate the implementation of a spatial multidimensional model. The extension to the spatial domain provides a concise and organized representation of spatial phenomena [15] that facilitates the delivery of data for spatial OLAP systems, spatial data mining, or spatial statistical analysis.

A conceptual multidimensional model establishes a communication bridge between users and designers. It reduces the difficulties of spatial data handling as decision-makers do not usually possess the expertise required by software currently used for managing spatial data.

Concurrently, we are directing our research towards aspects such as object roles, constrains, data quality, SOLAP functionalities and spatial ETL (extract, transform, load) for multi-granularity spatial DW. When modeling spatial data these are aspects that have to be considered for the spatial properties, to avoid inconsistency and ambiguity. This concern has been neglected insofar.

Moreover, when considering interoperability of spatial data, we need to have a modeling notation (in our case Perceptory) compatible with todays' standards [22].

VIII. REFERENCES

[1] M.L. Gonzales, "Seeking Spatial Intelligence," *Intelligent Enterprise Magazine*, 3(2), 2000.

[2] J. Han and M. Kamber, *Data Mining: concepts and Techniques*, Morgan Kaufmann Publisher, Inc, 2001.

[3] Y. Tao and D. Papadias, "Historical Spatiotemporal Aggregation," *In Proc. ICDE*, 2002.

[4] C. Murray, "Oracle Spatial User's Guide and Reference," Release 10.1 (B10826-01), Oracle, 2003.

[5] S. Shekhar and S. Chawla, *Spatial Databases: A tour*, Prentice Hall, 2003.

[6] P. Rigaux, M. School and A.Voisard, *Spatial Databases with Application to GIS*, Morgan Kaufman 2002

[7] P.A. Longley, M.F. Goodchild, D.J. Magurie, D.W. Rhind.,*Geographical Information Systems, Principals and Technical Issues*, 2nd ed. Wiley & Sons, 2002.

[8] M. Body et al, "Handling Evolutions in Multidimensional Structures" IEEE 19th Int. Conf. on Data Engineering (ICDE), March 5-8, Bangalore, India, 2003.

[9] Y. Bedard et al. "Modeling Geospatial Databases with Plug-Ins for Visual Languages" ER2004, LNCS 3289, 17–30, 2004.

[10] R. Kimball, R. Merz and M. Ross, *The Data Warehouse Toolkit*, 2nd ed., 2002.

[11] N. Stefanovic et al, "Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes," *IEEE Transactions on Knowleadge and Data Engineering, vol. 12*, nº. 6, p. 938-958, 2000.

[12] D. Papadias et al "Indexing Spatiotemporal Data Warehouse," 2002.

[13] M. Indulska, M. Orlowska, "On Aggregation Issues in Spatial Data Management," *In Proc. 13th Australasian Database Conference*, p. 75-84, 2002.

[14] E. Malinowski, Zimányi, "Representing Spatiality in a Conceptual Multidimensional Model," *GIS*, 2004.

[15] C.S. Jensen et al,"Multidimensional data modeling for location-based services" *VLDB Journal*, 13:1-21, 2004

[16] E. Camossi, M. Bertolotto, E. Bertino and G. Guerrini, "A Multigranular Spatiotemporal Data Model," *In Proc. ACM-GIS*, 2003.

[17] G. Pestana et al "A Prototype Implementation of a Spatial Data Warehouse for Integrating Business, Historical and Spatial Data," *5th Int. Conf. of Intelligent Data Engineering and Automated Learning*, 2004.

[18] Open GIS Consortium, OpenGIS Specifications, 2003. http://www.opengis.org/ogcSpecs.htm.

[19] ISO TC211, 2003. http://www.isotc211.org

[20] P. Marchand et al, "Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis", Int. Soc. of Photogrammetry and Remote Sensing Journal, 59:1, 6-20, 2004

[21] Y. Bedard et al, "Integrating GIS Components with Knowledge Discovery Technology for Environmental Health Decision Support", Int. J. of Medical Informatics, 70:1, 79-94, 2002,

[22] J. Brodeur, et al, "Modelling Geospatial Application Database using UML-based Repositories Aligned with International Standards in Geomatics", ACMGIS 2000, Nov. 10-11, 2000