TOWARDS SPECIALIZED INTEGRITY CONSTRAINTS FOR SPATIAL DATACUBES

M. Salehi [a,b,]*, Y. Bédard [a,b], M.A. Mostafavi [a], J. Brodeur [c]

[a] Center for Research in Geomatics and Department of Geomatics Sciences
[b] Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support
Laval University, Quebec City, Canada
mehrdad.salehi.1@ulaval.ca, (yvan.bedard, mir-abolfazl.mostafavi)@scg.ulaval.ca
[c] Center for Topographic Information, Natural Resources Canada, Sherbrooke, Canada - brodeur@nrcan.gc.ca

KEY WORDS: Integrity Constraint, Spatial Datacube, Spatio-temporal Data Quality, SOLAP, Multidimensional Database

ABSTRACT:

Spatial datacubes (also called "spatial multidimensional databases") are the cornerstone of the emerging Spatial On-Line Analytical Processing (SOLAP) technology. They are aimed at supporting Geographic Knowledge Discovery (GKD) as well as certain types of spatial decision-making. Although these technologies seem promising at first glance, they may provide unreliable results if one does not consider the quality of spatio-temporal data stored in them. In traditional spatial databases, spatial integrity constraints have been employed to improve internal quality of spatial data. However, for defining integrity constraints of spatial datacubes, these traditional spatial integrity constraints should be revisited. To this end, this paper presents the characteristics of spatial datacubes that differentiate them from transactional spatial databases from an integrity constraint perspective. These characteristics concern the datacubes model structure, the presence of thematic, spatial, and temporal data, and the various levels of data reliability for different decisions. Based on these characteristics, we propose spatial multidimensional integrity constraints and identify their fundamental characteristics. These characteristics include (1) considering the building elements of multidimensional data structures, (2) restricting thematic, spatial, and temporal data cross-tabulation, and (3) including a range of tolerance within the definition of integrity constraints. The analysis of today's solutions shows that existing spatial integrity constraint specification languages cannot express efficiently spatial multidimensional integrity constraints. Finally, future research directions for a formal model of spatial multidimensional integrity constraints is discussed as well as integrity constraints specification languages.

1. INTRODUCTION

Decision Support Systems (DSS) help strategic managers to make decisions efficiently. Decision makers need fast answers made up of aggregated and summarized large units of data. To this respect, DSS are often based on datacubes (or, multidimensional databases, as defined in the field of Business Intelligence, see Section 3). In datacubes, *dimensions* are the axis of analysis and *measures* are the numeric data analyzed against the different granularity levels of dimensions (Rafanelli 2003, Gray 1997). Both dimension and measure can refer to location data. Even tough location data has been integrated in datacube applications, it is usually represented in an alphanumeric manner (Malinowski, and Zimányi 2005). Taking into account the geometric representation of location data integrates the power of spatial data with the efficiency of datacubes in decision making and leads to an efficient DSS tool known as Spatial OLAP (SOLAP) (Bédard, *et al* 2006).

Although SOLAP seems a promising decision support tool, without considering data quality in its spatial datacubes, it may provide unreliable results. In transactional spatial databases, which are considered as the data sources for spatial datacubes, spatial integrity constraints are defined along the database conceptual models to preserve spatial data quality (Normand 1999; Mostafavi *et al.* 2004; Vallières *et al.* 2005). However, for maintaining data quality within spatial datacubes, additional integrity constraints must be considered. In the database community, some research works study integrity constraints for non-spatial datacubes (Carpani 2001; Hurtado and Mendelzon 2002; Ghozzi *et al.* 2004). However, in order to study spatial datacubes' integrity constraints, the specific characteristics of datacubes as well as spatial and temporal data features should be considered together. It appears that no study insofar has attempted to address these issues simultaneously.

Consequently, the objective of this paper is to propose fundamental considerations about spatial datacubes integrity constraints. More particularly, we focus on the conceptual level which is independent from technology choices and consider the algorithms used for integrity checking beyond the goal of this paper.

The remainder of this paper is structured as follows. Section 2 reviews integrity constraints in transactional spatial databases and their role in spatial data quality, and proposes a classification. Section 3 explains the need for spatial datacubes in decision making and their structure. Section 4 discusses the characteristics of spatial datacubes from a spatial integrity constraint point of view. Section 5 introduces multidimensional integrity constraints. Finally, Section 6 concludes and draws more research directions for the definition and implementation of the integrity constraints for spatial datacubes.

2. INTEGRITY CONSTRAINTS FOR TRANSACTIONAL SPATIAL DATABASES

One of the fundamental aims of defining integrity constraints is to improve data quality for databases. Integrity constraints are the rules defined along the conceptual models to prevent entering incorrect data into a database (Godfrey *et al.* 1997). In

* Corresponding author.

the database community, different types of integrity constraints have been proposed (Elmasri and Navathe 2000; Tansel 2004). They are also implemented in diverse places such as within SQL Data Definition commands, within the forms to fill, with cleaning processes, etc.

In the geomatics community, spatial data quality involves internal data quality and external data quality (Devillers *et al*, 2007). While the later deals with the fitness for use of the data, the ISO/TC211 (2002-a) suggests the following elements for the internal quality: completeness, positional accuracy, temporal accuracy, thematic accuracy, and logical consistency. Spatial integrity constraints have been defined to preserve logical consistency in spatial databases (Normand 1999; Servigne *et al.* 2000; Vallières *et al.* 2005). Spatial integrity constraints express the correct spatial property or relationship between spatial objects. They differ from non-spatial ICs in the sense that they refer to geometries (e.g. closure of polygons) and convey geometric properties and relationships (e.g., overlap, distance).

We classify integrity constraints in transactional spatial databases into three categories. First, *geometric integrity constraints* are based uniquely on the geometric properties and relations of the spatial objects, such as "a polygon must be closed". Second, *thematic integrity constraints* are defined by relying only on the thematic properties of spatial or non-spatial objects and are like the business rules defined in non-spatial databases, for instance "the number of floors of a house must be greater than zero". Third, *geo-spatial integrity constraints* are defined according to spatial properties and relationships of spatial objects in addition to their semantics, such as "a railway cannot intersect with an airplane landing strip". In a former project, more than 600 such constraints have been defined for the Quebec Topographic Database at the end of the 1990s (Normand, 1999). A similar exercise was also done by the second author for Canada's National Topographic Database. We are currently revisiting these projects and similar research done elsewhere to define a comprehensive classification and model for the integrity constraints used in spatio-temporal databases (Salehi *et al.*, 2007).

The required spatial integrity constraints can be specified by considering a conceptual data model and by using an integrity constraint specification language, such as a controlled natural language (Normand 1999), a spatial extension of first-order logic (Hadzilacos and Tryfona 1992), or a spatial extension to a formal language such as the Object Constraint Language (OCL) (Duboisset *et al.* 2005).

3. FROM TRANSACTIONAL SPATIAL DATABASES TO SPATIAL DATACUBES

Transactional spatial databases are designed to store, protect, update, and disseminate detailed up-to-date data while ensuring minimum redundancy and maximum integrity. However, decision-makers need fast answers made up of aggregated data summarizing large units of works. They need to analyze many aspects that may interact at different levels of granularity, including varying spatial and temporal granularities. To facilitate and accelerate these complex analysis and visualization operations, the databases for spatial decision support systems are typically modelled using the spatial datacube paradigm. Spatial datacubes add spatial components to multidimensional database structure (Bédard *et al.* 2001) (details on multidimensional structure is given in Section 4.1).

Since spatial datacubes are being used as backends of spatial decision support tools (e.g., SOLAP), preserving data quality within them is necessary. As it is the case for transactional spatial databases, data quality in spatial datacubes can be maintained with integrity constraints. However, the integrity constraints that are used for transactional spatial databases do not capture all the characteristics of spatial datacubes. In the next section, we will discuss these characteristics.

4. CHARACTRISTICS OF SPATIAL DATACUBES

Several characteristics such as multidimensional data model structure, the simultaneous existence of temporal, spatial, and thematic data, and the varying levels of data reliability required for decisions differentiate spatial datacubes from transactional spatial databases. These differences impact on the required integrity constraints that need to be implemented for spatial datacubes. This section highlights these characteristics.

4.1 Multidimensional data model structure

The transactional databases conceptual schemas show an abstraction of a part of reality from users' data needs point of view. However, as datacube is organized for a specific decisional need, its conceptual model reflects how data is analyzed in the decision making process. This difference in modelling requirements results in different structures and constructs for datacubes models versus transactional models.

At the conceptual level, transactional database models consist of concepts like object classes or entities, attributes, identifiers, relationships between objects, multiplicities, constraints between relationships, etc. However, the metamodel of datacubes is quite different. Data organization in conceptual datacube models rely on concepts such as dimension, hierarchy, granularity, member, measure, properties and fact.

Dimensions reflect axis of analysis for a user and are structured into one or several hierarchies. Each *hierarchy* is a directed graph whose vertex stands for a level, and edges connect and show the relationship between these levels (see Figure 1). A hierarchy organizes the granularity *levels* from lower-level (e.g., day) to higher-level (e.g., month) in a dimension (e.g., temporal dimension). Hierarchical structure in the dimensions allows database users to view and analysis dimensions at different levels of detail. *Members* are instances of levels within a dimension (e.g., Monday is a member of the day level in the temporal dimension). In a hierarchy, the members of the lower-level roll-up to the members of the higher-level. Roll-up can be considered as a partial function from the member set of a lower-level to the member set of a higher-level. *Measures* are numerical attributes such as "number of accidents" and are analyzed against all granularity levels of all hierarchies. Each unique combination of the members of all dimension hierarchies' levels and of the resulting measure value represents a *fact*. A *datacube* or *hypercube* is made of all the possible combinations of dimensions' granularity levels and their corresponding computed measure values.

4.2 Simultaneous existence of thematic, spatial, and temporal data types

In transactional spatial databases, thematic and spatial data exist together. These databases typically replace past data by the new data updates and do not maintain the history of objects. Even today, spatio-temporal databases still remain rare exceptions and are difficult to maintain and query. However, temporal data are important components for decision making in order to understand the evolution of phenomena, detect trends or predict what may happen in the future. To overcome this limitation of spatial transactional databases, spatial datacubes typically keep temporal data in addition to thematic and spatial data. In other words, datacubes typically include spatio-temporal facts and are built in a way that facilitates their management.

To better understand the multidimensional data model structure (Section 4.1) and their built presence of thematic, spatial, and temporal data (Section 4.2), in the remainder of this section we give a number of examples. For the situations which will arise from these examples, in Section 5.2 we further address some required integrity constraints.

Example 1: Temporal dimension

As opposed to spatio-temporal transactional databases, integrating temporal dimensions in spatial datacubes is usual practice. Figure 1 shows a model of a temporal dimension consisting of one hierarchy with the lower-level Day, middle-level Month, and higher-level Year.

Figure 1. Example of a temporal dimension model with three levels.

Example 2: Spatial dimension with spatial hierarchies

A spatial dimension is a specific type of dimension where its members have a cartographic representation or spatial analysis results (Bédard *et al.* 2006). Each spatial dimension is structured by one or several spatial hierarchies.

Spatial dimensions are of three types (Bédard *et al.* 2001). 1- *Non-geometric spatial dimension* is a dimension containing only non-geometric data. For example, if the instances of "Administrative Region" dimension are all nominal data referring to the name of cities, provinces, etc, then this spatial dimension is non-geometric. 2- In *fully geometric spatial dimension*, the instances of all dimension's levels have geometric representation. 3- *Hybrid spatial dimension* is a dimension whose some levels of data are geometric and others are not.

For instance, consider a spatial dimension called "Administrative Region". The purpose of this dimension is to conceptually reflect countries' territorial divisions and the user intends to analyse data for the US and Canada together. Canada consists of provinces and territories; however, the US is made up of states. Hence, we need two spatial hierarchies within this dimension to model the different territorial divisions for these two countries. Figure 2 shows this dimension and its two hierarchies. In this figure, City, Province, Territory, and Country are the levels of the first hierarchy representing Canadian territorial division, and City, State, and Country are the levels of the second hierarchy standing for the US territorial division. In the first hierarchy, the cities in Canada roll-up to either provinces or territories and both provinces and territories roll-up to country Canada.

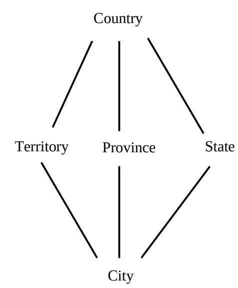

Figure 2. Administrative Region spatial dimension model with two hierarchies for Canada and the USA.

The hierarchies within spatial dimensions can have different structures (Malinowski and Zimányi 2005). For instance, a simple spatial hierarchy is a hierarchy that the relationship between the levels' members is represented by a tree. In multiple alternative spatial hierarchies there are several non-exclusive simple spatial hierarchies with sharing levels. Parallel spatial hierarchy is made up of several dependent or independent spatial hierarchies each one used for different analysis. In Figure 2, the Canadian territory division hierarchy is a simple generalized hierarchy in which each City member rolls-up to either a Territory or a Province. However, US Territorial Division hierarchy is a simple hierarchy where all the members of the lower-level roll-up to the same higher-level.

Data for the levels of a spatial hierarchy can be populated from maps at different scales. For example, in the US Territorial Hierarchy, City and State data can be populated from two different sources at different map scales. Alternatively, the map scale of a unique source necessitates using cartographic generalization operators (simplification, elimination, etc) to make the map appropriate for the scale of the higher-level of the hierarchy. Using cartographic generalization, potentially leads to the aggregation-generalisation mismatch problem (Bédard *et al*, 2006).

4.3 Varying Levels of Reliability for Decisions

Beside their structure and the types of data they store, the third characteristic of spatial datacubes refers to the goal for which they are used. As these datacubes are employed in the process of decision making, the definition of their integrity constraints should be oriented toward improving the reliability of decisions rather than solely focusing on individual data integrity (cf. measures in facts). This differs from spatial transactional databases where spatial integrity constraints must be defined to support detailed information for operational activities. In datacubes, it is quite frequent not to keep the detailed data used to calculate aggregated data and to see the decision-maker asking for "indicators", for "orders of magnitude" rather than overly-precise data. Consequently, the ultimate goal of spatial integrity constraints in transactional spatial databases and spatial datacubes is different, the latter accepting more easily "loosely constrained" data in favor of timely and pertinent aggregated data.

5. SPATIAL MULTIDIMENSIONAL INTEGRITY CONSTRAINTS

5.1 Existing research works

A number of research works investigate the integrity constraints within non-spatial datacubes. Carpani *et al.* (2001) propose a structure and a many-sorted logic language for supporting integrity constraints in multidimensional databases. Hurtado and Mendelzon (2002) and Hurtado *et al.* (2005) suggest dimension constraints for addressing correct aggregation paths in a hierarchical domain and to reason about summarizability. Ghozzi *et al.* (2004) study the integrity constraints between dimensions. These integrity constraints address the possible combinations of dimensions' hierarchies for a fact. For example, consider a multidimensional schema with two dimensions "Administrative Region" and "Product". An integrity constraint for this schema must express that "US Territorial Division" hierarchy in "Administrative Region" dimension cannot cross "Canadian Product" hierarchy in "Product" dimension. Instead, "US Territorial Division" hierarchy should cross "US Product" hierarchy.

To the best of our knowledge, there is no research work studying spatial integrity constraints within spatial datacubes. We call such integrity constraints spatial multidimensional integrity constraints.

5.2 Characteristics of spatial multidimensional integrity constraints

This section presents the specification of the spatial multidimensional integrity constraints following the characteristics of spatial datacubes discussed in section 4. Based upon these characteristics, we reveal the shortcomings of the existing languages used for specifying spatial multidimensional integrity constraints. Moreover, we discover different types of multidimensional integrity constraints by presenting a number of examples.

5.2.1 Refering to multidimensional data model structure

Transactional spatial databases integrity constraints mostly refer to the object-oriented modelling elements in a conceptual schema. Therefore, their specification language supports specific concepts such as classes, attributes, domains, relationships, multiplicities, and object instances. However, a multidimensional model consists of concepts like dimension, hierarchy, level, member, fact, and measure. Consequently, an integrity constraint specification language for spatial datacubes should hold specific syntax and semantics supporting these concepts. The proposed spatial multidimensional conceptual models (e.g., Malinowski and Zimányi (2005)) do not include such a formal language.

In order to better explain, let us take an example. Consider two levels "State" and "Country" of "US Territorial Division" hierarchy of "Administrative Region" dimension in Figure 2. The spatial integrity constraint between these two levels in natural language is: "for two levels State and Country in US Territorial Division hierarchy of Administrative Region dimension, the geometric union of States' members should be equal to Country's member geometry". The existing formal languages for the specification of spatial integrity constraints (e.g., Brodeur *et al.* 2005) do not support vocabulary expressing "dimension", "hierarchy", and "member" in this spatial integrity constraint. This is in a way similar to the SQL query language that does not support multidimensional concepts. In fact, since a de facto standard multidimensional language has emerged especially for datacubes (MDX), it indicates a need for an integrity specification language especially built for datacubes. However, from our point of view, it is preferable that such language builds on existing languages as much as possible.

It is worth noting that, the existing spatial integrity constraint specification languages can express this integrity constraint without referring to multidimensional vocabulary. However, spatial datacube conceptual model is constructed based on multidimensional concepts. Hence, it is more efficient to express spatial multidimensional integrity constraints referring to multidimensional elements such as the example above.

5.2.2 Restricting thematic, spatial, and temporal data

As stated previously, three types of data, i.e., thematic, spatial, and temporal data exist within spatial datacubes. Thus, the multidimensional integrity constraints restrict the allowable spatial, temporal, and thematic data. It turns out that any language developed for defining spatial multidimensional integrity constraints must support the syntax and semantics of thematic, spatial, and temporal restrictions. However, existing spatial integrity constraint specification languages do not support temporal restrictions.

The rest of this section introduces the possible types of multidimensional integrity constraints by giving a number of examples.

Example 3: **Spatio-temporal integrity constraints**

This section gives a number of examples for spatial and temporal data within multidimensional structure and resulting integrity constraints. The first example studies the effect of temporal granularity on the definition of spatial integrity constraints. The second example illustrates the combinations of quantitative temporal constraints with spatial integrity constraints. And, we explain the qualitative temporal constraints within spatial integrity constraints in the third example.

As a first example, let us consider a spatial dimension called "Location" and a temporal one called "Period" both consisting of a single hierarchy. The levels of the "Location" dimension are "Building", "Building Block", and "City" while the levels of the "Period" dimension are "Day", "Month", and "Year" (see Figure 3).

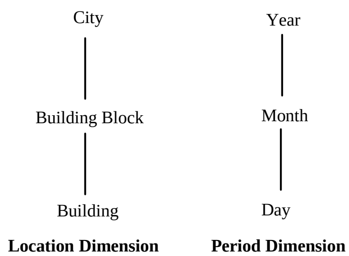

Figure 3. “Location” spatial dimension and “Period” temporal dimension for a spatial datacube.

We must intersect all the members of all of these granularity levels to construct the spatial datacube (e.g., City with Year, Building Block with Day, etc.). However, the spatial integrity constraint for each intersection may be different one from another. While intersecting “City” members with “Day” members, we need a spatial integrity constraint saying that “two cities do not overlap during the granularity level of one day”, while intersecting “City” members with “Year” members, a spatial integrity constraint could say “two cities may overlap within the granularity level of one year”. This may happen, as illustrated in Figure 4, when two different cities A and B that existed on August 15th, 2003 have merged together to create one unique and new city C on February 2nd, 2004. In such case, A and B do not overlap between them and do not overlap with C when temporal granularity is one day, e.g., on September 12th, 2003 or March 3rd, 2004. However, A and B may overlap with C when temporal granularity is one year, e.g., in 2004.

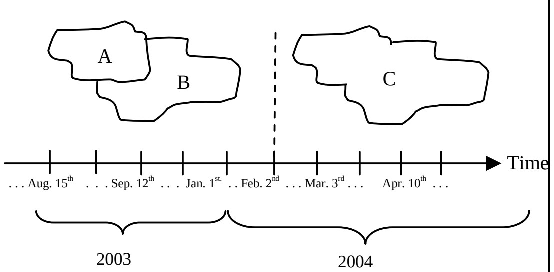

Figure 4. Cities A and B existed before February 2nd, 2003 and merged together on February 2nd, 2003 to create new city C

As a second example, let us consider two spatial dimensions, one called “Hazards” having a lower-level “Gas Station” and the other one called “Public Zones” having a low-level “School”, along with a temporal dimension. We can then define metric spatial integrity constraints combined with quantitative temporal constraints. Thus, a law stating that “the distance between gas station and school should be more than 300 meters” could have been valid from 1990 to 2000 and then revised during the year 2000 to “more than 500 meters” for the recent gas stations and schools.

Finally, a third example makes use of the 13 qualitative temporal constraints of Allen (1983) within spatial integrity constraints. This example rules that “the geometry of a province cannot evolve after its first creation”. Here, “after” is a qualitative temporal constraint combined with spatial integrity constraint.

Example 4: Integrity constraints for spatial dimensions and hierarchies

Among the three types of spatial dimensions (see Section 4.2, Example 2), the integrity constraints of non-geometric spatial dimension are treated similarly to non-spatial dimensions. In hybrid spatial dimensions, when two levels are one geometric and one nominal, the integrity constraint between them is like a non-spatial integrity constraint. However, when two levels are both geometric, the integrity constraint between them is spatial. In this section, when we refer to spatial dimensions’ levels, we assume that these levels have geometric representations.

As a first example, let us consider two spatial levels City and State of the spatial hierarchy US Territorial Division (see Figure 5). Since the union of all states creates the country, the spatial integrity constraint between State and Country addresses that the geometric union of all the States is equal to the geometry of Country (see Figure 5). This integrity constraints is a part-whole spatial integrity constraints (Price *et al.* 2001)

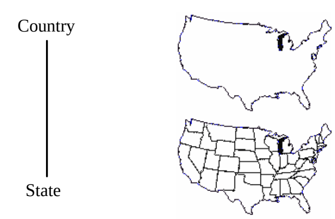

Figure 5. The integrity constraint between two levels State and Country in the US Territorial Division hierarchy should express that geometric union of states is equal to country.

As discussed in section 4.2, a spatial dimension may consist of different types of hierarchies. In addition to semantics of the levels, the structure of spatial hierarchy affects the definition of hierarchies’ spatial integrity constraints. For example, in a simple generalized spatial hierarchy when the members of two levels roll-up to one level, the spatial integrity constraint between the two lower-levels and the higher-level should be defined. An example is Canadian Territorial Division hierarchy in Figure 2 that geometrical union of two lower-levels Province and Territory must be equal to the geometry of higher-level Country. This is contrary to the simple hierarchy (e.g., US territorial division) where for each higher-level (e.g., Country) there is always one lower-level (e.g., State).

Another aspect that is necessary to be addressed by integrity constraints is the correct aggregate navigation paths in a dimension. In “Administrative Region” dimension in Figure 2, it can not be perceived by only dimension model if all city members can roll-up to provinces, territories, and states, or those members that roll-up to one of these higher-levels does not roll-up to the other ones. In fact, dimension's schema itself is not semantically rich enough to express this information. Hurtado, *et al.* (2005) resolved this problem in non-spatial datacubes by dimension constraints. In spatial datacubes, the definition of dimension constraints requires taking into account spatial relationships between the levels as well.

As previously stated, for populating a spatial hierarchy' levels it is likely to use spatial data at a different scale. Today, two methods exist to populate these levels. First method is to populate lower-level and use cartographic generalization procedures to produce higher-level data. Whereas the second method uses two different geospatial data sources at two different scales and they are linked using automatic matching. In both cases, higher-level data is in a different map scale than lower-level data. Traditional spatial integrity constraints do not consider the generalization operators (e.g., simplification) employed while changing the map scale. According to our experience in different practical projects, using these integrity constraints for verifying data quality of multi-scale multi-source data rejects too many spatial data to be entered into spatial datacube. In fact, these spatial integrity constraints which are "single-scale" are not appropriate for verifying spatial data quality of two levels at two scales. (N.B. a third method to populate spatial datacubes is being investigated in our Research Centre: multi-scale data acquisition using geometric patterns)

In order to explain this problem more clearly, consider, for example, US Territorial Division hierarchy with two levels State and Country. For populating this spatial hierarchy's levels, two different data sources at two different scales can be employed. The traditional spatial integrity constraint between State and Country indicates that geometric union of states must be equal to the geometry of country. The definition of this integrity constraint, which is derived from US country semantics, is for a single-scale spatial database. As US geometry is affected by cartographic generalization process, the geometry of Country is not exactly equal to geometric union of States. Therefore, spatial data for these two levels could not respect this integrity constraint and is rejected from entering into spatial datacube.

In the next section, we introduce the inclusion of a range within the definition of integrity constraints as another characteristic of spatial multidimensional integrity constraints. As we will discuss, integrity constraints with a range can be considered, among others, as a revision of single-scale spatial integrity constraints.

5.2.3 Including a range for integrity constraints

As explained in Section 4.3, for decision support systems, the reliability of a decision is important but it does not necessarily mean storing highly constrained data. Transactional spatial databases, such as some cadastral databases, aim at providing the correct extension of the land parcels and follow strict spatial integrity constraints, like: "a road does not overlap a building". However, spatial multidimensional integrity constraints can include a range accepting a tolerance for the integrity constraint. This range does not affect the process of decision making but accelerates spatial datacube populating. For example, a spatial integrity constraint such as "a road can overlap a building but the overlapping should not happen to more than *10%* of the buildings" is an integrity constraint with a range or threshold.

Another application of integrity constraint with a range is to revise single-scale spatial integrity constraints to multi-scale. Consider, for example, data populating for US Territorial Division hierarchy. For this hierarchy, as shown in Figure 6, State level data is the original source; however, Country level data is produced by generalization this source. Because of using generalization operators, US country border is simplified comparing to State's border. Therefore, spatial union of states is not equal to country, and integrity constraint "geometric union of states must be equal to country" does not accept this multi-scale data to be entered into spatial datacube. However, we can revise this constraint to "geometric union of states' geometries must be *between* country's geometry $\pm 5\%$". The range "$\pm 5\%$ of country's geometry" makes the integrity constraint suitable for data population from two different data sources at two different scales.

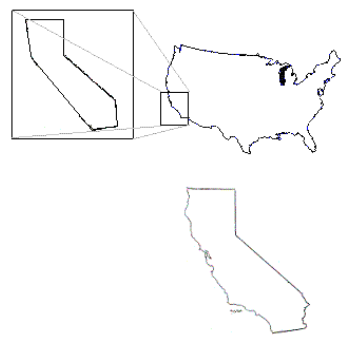

Figure 6. Populating two levels State and Country of US Territorial Division hierarchy from two sources at two different scales where border of country is simplified comparing state's border.

Noting that, the end-user, during conceptual modelling design, should be asked about the acceptable range for each integrity constraint. This range is added to the definition of integrity constraints by the data model designer. Consequently, we may have different sets of integrity constraints for different datacubes stemming from the same transactional sources. More investigation is underway to formally define the concept of range within integrity constraints.

6. CONCLUSIONS AND FUTURE RESEARCH

Spatio-temporal data quality in spatial datacubes has a significant effect on the decisions made using these databases. In spite of the research works for maintaining data quality by defining integrity constraint in spatial databases as well as non-spatial datacubes, no study has attempted to address integrity constraints in spatial datacubes.

This paper is the first work providing fundamental considerations for defining integrity constraints for spatial datacubes. Indeed, it is analysed as a spatial extension to the previously studied integrity constraints for non-spatial datacubes.

To this respect, we reviewed spatial data quality and its improvement by spatial integrity constraints. Additionally, we proposed a classification for traditional spatial integrity constraints. Next, we described the properties of spatial datacubes including multidimensional data model structure, existence of thematic, spatial, and temporal data, and reliability of decision. These properties characterize spatial datacubes from spatial transactional databases and influence on the required integrity constraints. We explained multidimensional integrity constraints as the integrity constraints to maintain data quality within spatial datacube. We discussed on the characteristics of these integrity constraints as referring to multidimensional data structure, restricting thematic, spatial, and temporal data, and including a range. These characteristics show that the existing integrity constraint specification languages for spatial transactional databases can not efficiently express the necessary integrity constraint of spatial datacubes. We are currently working on a formal model and language for spatial multidimensional integrity constraints.

This paper is a part of an ongoing research aiming at answering the following questions: 1- What should taxonomy of spatial multidimensional integrity constraints include considering the aspects mentioned in section 5.2? 2- What are the required syntax and semantic extensions needed for an integrity constraint specification language to express these integrity constraints? 3- What are the translation rules from the integrity constraint specification language to data definition languages included in SQL, MDX and programming languages such as Java? 4- What is the strategy to appropriately define a range on the integrity constraints that does not affect the process of decision making?

7. ACKNOWLEDGEMENT

The authors wish to thank for its support the NSERC Industrial Research Chair of Geospatial Database for Decision Support financed by the Natural Sciences and Engineering Research Council of Canada, Laval University, Hydro-Québec, Research and Development Defence Canada, Natural Resources Canada, Ministère des Transports du Québec, KHEOPS Technologies, Intélec Géomatique, Syntell, Holonics, and DVP-GS.

REFERENCES

Allen, J. F. (1983). Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11), pp. 832-843.

Bédard, Y., T. Merrett, et al. (2001). Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery. *Geographic Data Mining and Knowledge Discovery*. H. Miller and J. Han, Taylor & Francis**,** pp. 53-73.

Bédard, Y., S. Rivest, et al. (2006). Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures, and Solutions from a Geomatics Engineering Perspective. *Data Warehouses and OLAP: Concepts, Architecture, and Solutions*. R. Wrembel and C. Koncilia, London, UK, Idea Group Publishing, pp. 298-319.

Carpani, F., and, Ruggia, R. (2001). An Integrity Constraints Language for a Conceptual Multidimensional Data Model. *Proceedings of XIII International Conference on Software Engineering & Knowledge Engineering*, Argentina.

Devillers, R., Y. Bédard, R. Jeansoulin & B. Moulin (2007) Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data, *International Journal of Geographical Information Sciences (IJGIS)*, Vol. 21, No. 3, pp. 261-282.

Duboisset, M., F. Pinet, et al. (2005). Integrating the Calculus-Based Method in OCL: Study of Expressiveness and Code Generation. *2nd IEEE International Workshop on Geographic Information Management (GIM 2005)*, Copenhagen, Denmark.

Elmasri, R. and S. B. Navathe (2000). *Fundamentals of database systems*, Addisom-Wesley.

Ghozzi, F., F. Ravat, et al. (2004). Constraintes pour modèle et langage multidimensionnels. *Ingénierie des Systèmes d'Information* 9(1), pp. 9-34.

Godfrey, P., J. Grant, et al. (1997). Integrity Constraints: Semantics and Applications. *Logics for Databases and Information Systems*. J. Chomicki and G. Saake, Kluwer, pp. 265-306.

Gray, J., Bosworth, A., Layman, A., Pirahesh, H., (1997), Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp. 29-53

Hadzilacos, T. and N. Tryfona (1992). A model for expressing topological integrity constraints in geographic databases. *International Conference GIS-From Space to Territory*, Italy.

Hurtado, C., A., C. Gutierrez, et al. (2005). Capturing summarizability with integrity constraints in OLAP. *ACM Transactions on Database Systems*, 30(3), pp. 854-886.

Hurtado, C. A. and A. O. Mendelzon (2002). OLAP Dimension Constraints. Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Madison, USA, pp.169-179.

Malinowski, E. and E. Zimányi (2005). Spatial Hierarchies and Topological Relationships in the Spatial MultiDimER model. *22nd British National Conference on Databases*, Sunderland, United Kingdom, Lecture Notes in Computer Science, 3567, pp. 17-28.

Mostafavi, M. A., G. Edwards, et al. (2004). An Ontology-Based Method for Quality Assessment of Spatial Data Bases. *International Symposium of Spatial Data Quality 2004*, GeoInfo Series, Austria.

Normand, P. (1999). Modélisation des contraintes d'intégrité spatiale, théorie et exemples d'applications. *Department of Geometics Sciences. M.Sc. Thesis*, Université Laval, Quebec City, Canada.

Price, R., N. Tryfona, et al. (2001). Modeling Topological Constraints in Spatial Part-Whole Relationships. *Proceedings of the 20th International Conference on Conceptual Modeling*. Lecture Notes in Computer Science, 2224, pp. 27-40.

Rafanelli, M. (2003). Multidimensional Databases: Problems and Solutions, Idea Group Publishing.

Salehi, M., Y. Bédard, M.A. Mostafavi, J. Brodeur (2007). Classifications of Integrity Constraints in Spatio-temporal Databases: Toward Building an Integrity Constraints Specification Language. Submitted to *Transactions in GIS.*

Servigne, S., T. Ubeda, et al. (2000). A Methodology for Spatial Consistency Improvement of Geographic Databases. *Geoinformatica*, 4(1), pp. 7-34.

Tansel, A. U. (2004). Temporal Data Modeling and Integrity Constraints in Relational Databases. *International Conference on Information Technology: Coding and Computing 2004*, Lecture Notes In Computer Science, 3280, pp. 459-469.

Vallières, S., J. Brodeur, et al. (2005). Les contraintes d'intégrité spatiale: un outil permettant d'améliorer la qualité interne des données géographiques. *Qualité de l'information géographique*, Information géographique et aménagement du territoire (IGAT). R. Devillers and R. Jeansoulin, Paris**,** pp. 1-19.