FROM TRANSACTIONAL SPATIAL DATABASES INTEGRITY CONSTRAINTS TO SPATIAL DATACUBES INTEGRITY CONSTRAINTS

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 multidimensional databases (also called "spatial datacubes") are the cornerstone of the emerging Spatial On-Line Analytical Processing technology (SOLAP). 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. In traditional spatial databases, spatial integrity constraints have been employed to improve internal quality of spatial data. However, spatial datacubes require additional integrity constraints in comparison to the traditional databases found into transactional GIS systems. These extra constraints concern the supplementary information included in these datacubes, such as spatial dimensions and hierarchies, aggregated data, multidimensional cross-tabulation of data, and the existence of a temporal dimension with several levels of granularity. This paper presents the characteristics of spatial datacubes that differentiate them from transactional spatial databases from a spatial integrity constraint perspective. Based on these characteristics, we propose fundamental considerations for the classification of these integrity constraints and for the use of integrity constraint specification languages tailored for geospatial datacubes. Finally, the paper concludes and addresses some questions that contribute to a research agenda for the definition of spatial integrity constraints in spatial datacubes.

1. INTRODUCTION

Decision Support Systems (DSS) help strategic managers to make decisions efficiently. Decision makers need fast answers made up of aggregated data summarizing large units of data. To this respect, DSS are often based on multidimensional databases (datacubes) as their database backend. In multidimensional databases, dimensions are the axis of analysis and measures are the numeric data analyzed against the different granularity levels of dimensions (Rafanelli 2003). Both dimension and measure can refer to location data. Although location data has been integrated in multidimensional database 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 multidimensional database in decision making and leads to an efficient DSS tool known as Spatial OLAP (SOLAP) (Bédard, Rivest *et al.* 2006).

Although SOLAP seems a promising DSS tool, without considering the quality of data stored in its spatial multidimensional database, it may provide unreliable results. In traditional spatial databases, spatial integrity constraints are defined along the database conceptual models to preserve the quality of spatial data (Normand 1999; Mostafavi, Edwards *et al.* 2004; Vallières, Brodeur *et al.* 2005). Additionally, in non-spatial multidimensional databases, some research works study the integrity constraints in multidimensional models (Carpani 2001; Hurtado and Mendelzon 2002; Ghozzi, Ravat *et al.* 2004). However, it appears that no study insofar has attempted to address simultaneously these two concerns. For maintaining spatial datacube quality by specifying integrity constraints, the specific characteristics of multidimensional databases as well as spatial data features should be considered together.

This paper reviews the characteristics of spatial multidimensional databases and explains how they necessitate the definition of novel spatial integrity constraint. In other words, this paper contributes to the improvement of spatial datacubes quality by proposing different types of integrity constraints that should be addressed along the conceptual models of spatial multidimensional databases.

The remainder of this paper is structured as follows. Section 2 reviews spatial integrity constraints in spatial databases and propose a new classification for them. Section 3 explains the need for spatial multidimensional databases in decision making and its structure. Section 4 discusses the characteristics of spatial multidimensional databases from a spatial integrity constraint point of view. Section 5 introduces how the characteristics of spatial multidimensional databases necessitate new types of spatial integrity constraints. Finally, section 6 concludes and draws more research directions for the definition and implementation of the integrity constraints for spatial multidimensional databases.

2. SPATIAL INTEGRITY CONSTRAINTS

In the database community, different types of integrity constraints have been proposed (Elmasri and Navathe 2000; Tansel 2004). Integrity constraints are the rules that are defined

* Corresponding author.

along the conceptual models of databases to prevent entering of incorrect data into a database and to preserve database integrity (Godfrey, Grant *et al.* 1997). In spatial databases, spatial integrity constraints been defined to preserve spatial data quality (Normand 1999; Servigne, Ubeda *et al.* 2000; Vallières, Brodeur *et al.* 2005). Spatial integrity constraints are the rules that express the correct spatial property or relationship between spatial objects. In spatial transactional databases, they are defined along with the conceptual models of applications (within and aside) and then, the data entered in the database are compared with these rules in order to improve their internal spatial data quality.

We classify these integrity constraints 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, *semantic integrity constraints* are defined by relying only on 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 that zero”. Third, *geo-semantic 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 former projects, more than 600 such constraints have been defined for the Quebec Topographic Database and the Canada National Topographic Database at the end of the 1990s (Normand, 1999). We are currently revisiting these projects and similar research done elsewhere to define a comprehensive classification and model for the integrity constraints in spatio-temporal databases (Salehi, Bédard *et al.*, 2007).

The required spatial integrity constraints can be specified by considering the defined 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, Pinet *et al.* 2005).

3. FROM SPATIAL TRANSACTIONAL DATABASES TO SPATIAL MULTIDIMENSIONAL DATABASES

Transactional 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 decision support systems are typically modelled using the *multidimensional* (or datacube) paradigm. Some key concepts in multidimensional databases are *dimensions, members, measures, facts,* and *data cubes*. *Dimensions* reflect axis of analysis for a user and are structured by one or several hierarchies. Each hierarchy organizes the granularity levels from lower-level (e.g., day) to higher-level (e.g., month) in a dimension (e.g., temporal dimension). *Members* are instances within a dimension (e.g., Monday is a member of day level in the temporal dimension). *Measures* are numerical attributes such as “number of accidents” and analyzed against different granularity levels of dimensions. Each combination of the members of dimensions’ levels and its resulting measure value represents a *fact*. A *datacube* or *hypercube* is all the possible combinations of dimensions’ granularity levels and their corresponding computed measure values.

In the context of spatial decision making, the transition from spatial transactional databases to spatial multidimensional databases has significantly improved the capacity to answer complex queries in a timely manner. Spatial multidimensional databases add spatial components to multidimensional databases structures such as spatial dimensions which are dimensions with members having a cartographic representation or spatial analysis results, and spatial measures which may also contain spatial analysis results or collections of pointers to spatial objects. Spatial datacube is a datacube constructed by at least one spatial dimension or measure (Bédard, Merrett *et al.* 2001; Bédard, Rivest *et al.* 2006)

Spatial multidimensional databases are being used as backend of SOLAP tools. Consequently, the quality of spatial data stored in spatial multidimensional databases affects the quality of SOLAP analysis and of the overall process of decision-making. As for spatial transactional databases, data quality can be treated in spatial multidimensional databases with spatial integrity constraints. However, due to some characteristics of these databases, the spatial integrity constraints of spatial transactional databases are not rich enough for this task. In the next section we will discuss these characteristics.

4. CHARACTRISTICS OF SPATIAL MULTIDIMENSIONAL DATABASES

Several characteristics such as existence of temporal dimensions, spatial dimensions with hierarchies, and reliability of decisions differentiate spatial multidimensional databases from spatial transactional databases, impacting on the required spatial integrity constraints. This section highlights these characteristics.

4.1 Existence of temporal dimension

Temporal dimensions are important components for decision making in order to understand the evolution of phenomena or to predict what may happen in the future. On the one hand, spatial transactional databases typically replace past data by the new data updates and do not maintain the history of objects (spatio-temporal databases still remain rare exceptions). On the other hand, in spatial multidimensional databases the temporal dimension is omnipresent. Temporal dimensions similar to spatial dimensions may have several granularity levels. As opposed to spatio-temporal transactional databases, integrating temporal dimensions is usual practice. Figure 1 shows a model of a temporal dimension consisting of the lower-level Day, middle-level Month, and higher-level Year.

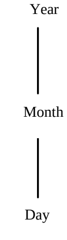

Figure 1. Example of a temporal dimension model with three levels.

4.2 Existence of spatial dimension with hierarchies

In spatial multidimensional databases, a spatial dimension reflects an axis of analysis where the data associated to this dimension is spatial. Each spatial dimension is structured by one or several spatial hierarchies which represent analysis perspectives. Hierarchical structure in the dimensions allows database users to view and analysis dimensions at different levels of detail.

For instance, let us consider a spatial dimension called "Administrative Region". This dimension represents country territorial divisions in the US and Canada together. Canada consists of provinces and territories; however, the US is made up of states. Since each of these countries consists of different territorial divisions, the dimension cannot be modelled with a same hierarchy. Consequently, "Administrative Region" dimension consists of two hierarchies, one for Canada and another for the US territorial divisions. Figure 2 shows this dimension and its two hierarchies. In this model of hierarchy, each vertex in the directed graph stands for a level and edges connect and show the relationship between the levels. For two connected levels, the one which is lower in the hierarchy (e.g., City) rolls-up to the higher one (e.g., State). 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 territory and both provinces and territories roll-up to country.

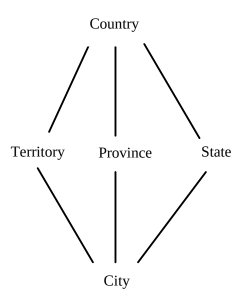

Figure 2. Administrative Region spatial dimension model with two hierarchies for Canada and the USA.

Spatial dimensions include three types (Bédard, Merrett *et al.* 2001). 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. Hybrid spatial dimension is a dimension whose some levels of data are geometric and others are not. For instance, in Figure 2, if the instances of City level are represented geometrically and instances of higher-levels such as Province are nominal, then the dimension is hybrid. In a fully geometric spatial dimension, the instances of all dimension's levels have geometric representation.

The spatial hierarchies can have different structures such as simple, multiple-alternative, and parallel spatial hierarchies (Malinowski and Zimányi 2005). 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 1, 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.

Additionally, the existence of spatial hierarchy may result in spatial data for 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 (e.g., simplification, elimination, etc) to make the map appropriate for the scale of the next level of the hierarchy, potentially leading to the aggregation-generalisation mismatch problem (Bédard et al, 2006). Therefore, populating several levels of spatial hierarchies can be considered as a multi-scale map population.

In section 5, we will discuss how the existence of spatial dimensions and hierarchies and their characteristics necessitate appropriate spatial integrity constraints for spatial datacubes.

4.3 Reliability of Decisions

For decision-making, the reliability of decisions has a higher priority than data integrity *per se*. Therefore, in spatial multidimensional databases, the definition of spatial integrity constraints should be oriented toward improving the process of decision making rather than solely focusing on data integrity. This differs from spatial transactional databases where spatial integrity constraints must be defined to support detailed information for operational activities. Consequently, the ultimate goal of spatial integrity constraints in transactional and multidimensional databases is different. This leads to different types of spatial integrity constraints.

5. SPATIAL INTEGRITY CONSTRAINTS FOR SPATIAL MULTIDIMENSIONAL DATABASES

5.1 The needs

In addition to research for spatial integrity constraints in spatial databases, some research works study the integrity constraints for non-spatial multidimensional databases. (Carpani 2001) proposes a structure and a many-sorted logic language for supporting integrity constraints in multidimensional databases. (Hurtado and Mendelzon 2002; Hurtado, Gutierrez *et al.* 2005) suggest dimension constraints for addressing correct aggregation paths in dimension models and to reason about summarizability.(Ghozzi, Ravat *et al.* 2004) study the integrity constraints between dimensions. These integrity constraints address the possible combinations of dimensions' hierarchies for a fact. An example is the integrity constraint in multidimensional model that expresses "US Territorial Division" hierarchy in "Administrative Region" dimension cannot cross "French 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 multidimensional database model. As the absence of these integrity constraints deteriorates the quality of decisions resulting from these databases, we must define them.

5.2 Aspects to Consider

This section presents the aspects that need to be considered when addressing the spatial integrity constraints in spatial multidimensional databases following the characteristics of these databases discussed in section 4.

5.2.1 Multidimensional data model structure

The existing spatial integrity constraint specification languages are mostly based on the object-oriented data modelling formalism for spatial transactional databases. Classes, attributes, domains, associations, cardinality and object instances play a central role in the definition of these spatial integrity constraints. However, concepts specific to multidimensional data models such as dimension hierarchical relationships, levels, member consistency, member's attribute domains, fact's referential integrity with dimension's members, and fact's measure domains, must be considered especially in the definition of spatial integrity constraints and specification languages to add the required syntax and semantics to support them.

5.2.2 Spatio-temporal integrity constraints

In spatial multidimensional databases, the granularity levels of spatial and temporal dimensions intersect with each other to make spatial datacubes. This necessitates defining the corresponding integrity constraints addressing both spatial and temporal restrictions. To better explain the types of integrity constraints resulting from the existence of spatial and temporal dimension with several granularity levels, we give three examples. 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 using 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).

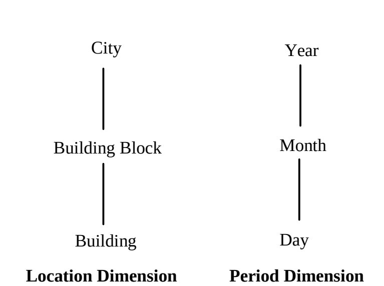

Figure 3. "Location" spatial dimension and "Period" temporal dimension for a spatial datacube.

We must intersect all the members of all of these granularity levels to construct the spatial datacube. However, the spatial integrity constraint for each intersection may be different from the other one. 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", whereas 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.

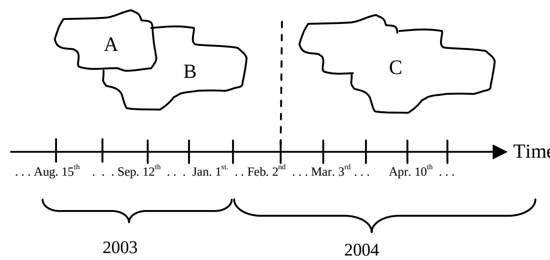

Figure 4. Cities A and B existed before February 2nd, 2003 and merged together on February 2nd, 2003 to create new city C

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) in 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.

Because of the existence of temporal dimensions, the specification of spatial integrity constraints in spatial multidimensional databases must consider the following three temporality aspects: temporal resolution (example 1), quantitative temporal constraints (example 2), and qualitative temporal constraints (example 3).

5.2.3 Spatial integrity constraints for spatial dimension and hierarchies

In dimension's hierarchies of multidimensional databases, lower-level members are classified and associated to a higher-level member in order to aggregate measure values at a more general level. For example, members of Day level are classified and associated to the corresponding Month members to aggregate measure values form a detail to a general level. In spatial multidimensional databases, for two consecutive levels (e.g., Building and Building Block, see Location Dimension of Figure 3) of a spatial hierarchy, this process of classification and association between lower-level (i.e., Building) and higher-level (i.e., Building Block) is done based on the topological relationships between them (e.g., inside). Indeed, all the Buildings which are "inside" a Building Block are associated to it. Accordingly, in addition to the visualization of spatial data, the process of aggregation from lower-level to higher-level in spatial dimensions is highly related to the quality of spatial data. Consequently, it is necessary to define the spatial integrity constraints for spatial dimensions to improve data quality of spatial datacubes, and hence, of the decisions.

Among the three types of spatial dimensions (see section 4.2), 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 non-spatial integrity constraint as well. However, when two levels are both geometric the integrity constraint between them is spatial. Hence, in this section when we refer to spatial dimensions' levels, we assume that these levels have geometric representations.

The spatial integrity constraints for a spatial dimension's hierarchy can be defined between the lower-level and the higher-level as well as geometric union of the lower-level members and higher-level, when lower-level members as a whole semantically represent a concept. These integrity constraints are part-whole spatial integrity constraints (Price, Tryfona *et al.* 2001). For example, in the US Territorial Division hierarchy in Figure 2 the spatial integrity constraint between a City and a State must express that each City is inside a State. Moreover, 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).

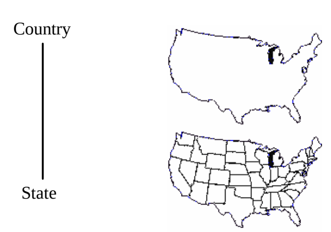

Figure 5. The integrity constraint between two levels State and Country in the US Territorial Division hierarchy should express that geometric union of states is equal to country.

In addition to semantics of the levels, the type of spatial hierarchy affects the definition of spatial integrity constraints. As discussed in section 4.2, a spatial dimension may consist of different types of hierarchies. For example, in a simple generalized spatial hierarchy when 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). In this case, the spatial integrity constraint between one lower and one higher level should be defined.

Another aspect that is necessary to be addressed by integrity constraints is the correct aggregate navigation paths in a dimension. Referring to the "Administrative Region" dimension in Figure 2, it can not be perceived by only dimension model if all the cities can roll-up to provinces, territories, and states, or those cities that roll-up to one of these higher-levels does not roll-up to the other ones which is the correct case. In fact, dimension's model itself is not semantically rich to express this information. (Hurtado, Gutierrez *et al.* 2005) study this problem in non-spatial multidimensional databases and suggest adding dimension constraints to dimension models to express this information. In spatial multidimensional databases, however, the definition of dimension constraints requires taking into account spatial relationship between the levels as well.

As previously stated, the existence of spatial hierarchy may result in a situation where each level contains spatial data at a different scale. Practically, two methods exist to populate spatial hierarchies' levels when levels refer to geometric data. First method is to populate lower-level and use cartographic generalization procedure on it to produce higher-level. Whereas the second method uses different geospatial data sources at different scales for each level. In these cases, higher-level data is either the result of applying generalization operators on the lower-level data or can be considered as applying these operators on the external data similar to the lower-level data. However, traditional spatial integrity constraints convey semantic information independently of cartographic generalization process employed while changing the map scale. Using these integrity constraints rejects multi-scale spatial data to be entered into spatial datacube. In fact, these spatial integrity constraints which are "single-scale" cannot be applied directly for verifying spatial data quality of two levels that the spatial data associated to them is in 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, let us consider US Territorial Division hierarchy with two levels State and Country. As already said, while changing spatial hierarchy's granularity levels, scale changes could happen and cartographic generalization operators are applied. In this hierarchy, the spatial integrity constraint between State and Country levels indicates that geometric union of states must be equal to the geometry of country. The definition of this traditional spatial integrity constraint between State and Country, which is derived from US country semantics and which assume a single-scale spatial database, do not consider the cartographic generalization process. As levels' geometries are affected by cartographic generalization process, the geometry of Country is not exactly equal to geometric union of States. Therefore, spatial data for the two levels State and Country could not respect this integrity constraint and is rejected from entering into spatial datacube. Consequently, we need a revision of the traditional spatial integrity constraints to make them appropriate for multi-scale spatial data.

In the next section, we introduce the notion of “soft” and “hard” spatial integrity constraints. As we will discuss, soft constraints can be considered as a revision of single-scale spatial integrity constraints that can be applied for verifying spatial data quality of multi-scale spatial databases.

5.2.4 Hard and soft spatial 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. Spatial transactional databases such as a cadastral database aim at providing the correct extension of the land parcels and follow strict spatial integrity constraints, e.g., “a road does not overlap a building”. However, in spatial multidimensional databases, it is possible to define spatial integrity constraints with a looser nature by accepting a tolerance that does not affect the process of decision making. We call the first ones “hard” constraints and the latter ones “soft” constraints. For example, a spatial integrity constraint such as “a road can overlap a building but the overlapping area should be less than 10% of the buildings’ area” is a soft constraint.

Another application of soft constraints is to revise single-scale spatial integrity constraints and make them appropriate for multi-scale spatial data. For example, the spatial integrity constraint for State and Country levels of US Territorial Division hierarchy expressing that “geometric union of states is equal to country” is a hard spatial integrity constraint. We can revise this hard constraint to “geometric union of states’ geometries must cover at least 95% of country’s geometry” to make it suitable for multi-scale spatial datacube population. This revised version of the hard constraint is a soft spatial integrity constraint.

Such spatial integrity constraints allow the system to accept data which are not completely correct but which are considered acceptable in the process of decision making since their impact is null. Therefore, the number of errors during data entry will decrease and the process of populating spatial multidimensional databases will be faster. More investigation is underway to enrich these concepts of soft and hard spatio-temporal constraints.

6. CNCLUSIONS 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 existing research works for maintaining data quality by defining integrity constraint in spatial databases as well as non-spatial multidimensional databases, no study has attempted to address spatial integrity constraint in spatial multidimensional models. This paper provided the fundamental considerations for defining these integrity constraints by discussing the evolution from spatial transactional databases to spatial multidimensional databases. First, we reviewed spatial integrity constraints in spatial databases and proposed a classification for them. Next, we described the properties of spatial multidimensional databases characterizing them from spatial transactional databases which influence the required spatial integrity constraints. Then, we explained aspects to consider for addressing these integrity constraints adapted to the characteristics of spatial multidimensional databases. This led to recognizing the need of further expanding the typology of spatial integrity constraints in order to accommodate multidimensional, spatio-temporal, spatial dimensions and hierarchies, and hard versus soft integrity constraints. We believe that this typology of spatial integrity constraints helps to define appropriate integrity constraints to maintain data quality in spatial datacubes.

This paper is a part of an ongoing research aiming at answering the following questions: 1- How can spatial datacubes integrity constraints be classified to consider 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 select only the integrity constraints that affect the process of decision making or to make them “softer”?

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.

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.

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.