MANAGING THE RISK RELATED TO THE SIMULTANEOUS USE OF SPATIAL DATACUBES

Sboui Tarek12, Bédard Yvan12, Brodeur Jean 31, Badard Thierry1

1 Department of Geomatic Sciences and Centre for Research in Geomatics, Université Laval, Quebec, Qc, G1K 7P4, Canada.

2 NSERC Industrial Research Chair in Geospatial Databases for Decision-Support

3 Natural Resources Canada, CIT, 2144-010 King West St, Sherbrooke, Qc, J1J 2E8, Canada.

Tarek.Sboui.1@ulaval.ca, Yvan.Bedard@scg.ulaval.ca, brodeur@nrcan.gc.ca, Thierry.Badard@scg.ulaval.ca

ABTRACT

Data warehouses are being considered as efficient components of decision support systems. They are usually structured as datacubes, i.e., according to the multidimensional paradigm. Spatial datacubes contain spatial components that allow spatial visualization and aggregation. One may need to use several spatial datacubes that may be heterogeneous in design or content. This heterogeneity may cause concerns about the risk of data to be misinterpreted, and hence about the repercussion of that risk on decision analysis. Supporting data interpretation by overcoming the heterogeneity problems has been the principal aim of semantic interoperability. In spite of successful initiatives and widespread use of international standards for representing spatial information, today's solutions, on one hand, still have a considerable vulnerability to the risk of misinterpretation, and on the other hand, do not address yet spatial datacubes. This paper describes the risks of misinterpretation related to the simultaneous use of spatial datacubes and proposes a risk management approach based on context-based semantic interoperability.

Keywords: Decision Analysis, Spatial Datacubes, Semantic Interoperability, Risk Management, Context.

1. INTRODUCTION

Over the last decades, there has been an exponential increase in the amount of data available from multiple sources. Furthermore, there have been significant innovations in information technology, especially in database technologies, Decision Support Systems (DSS), knowledge discovery and communication between information systems. DSS are information systems that support decision analysis. They provide techniques, data, models and knowledge to identify and solve problems and to improve making strategic decisions [41]. Basically, DSS consist of components that have (1) modelling capabilities, (2) databases management capabilities, and (3) friendly user interface designs [37]. Databases management components have to be designed to manage data, support ad hoc exploration of data at different levels of aggregation and at different epochs, and support discovering data patterns in a way that agrees with user’s mental model of the data [43], [34]. These goals appear attainable if we consider data warehousing [4].

Data warehouses are databases which are designed to supply DSS with data at different levels of aggregation. Data warehouses are often structured according the multidimensional paradigm that may use either a relational, object-oriented, or hybrid implementation. The multidimensional paradigm facilitates the navigation within the different levels of granularity of data. It allows to go to finer or coarser granularity and change dimension of analysis with one or two mouse clicks. Thus, using the multidimensional paradigm, data warehouses provide the basis for knowledge discovery and decision analysis (users can rapidly get a global picture of a phenomenon, get more insight into that phenomenon information, compare with other phenomena or analyse its evolution over time) [4]. Multidimensional databases (hereafter called datacubes) support user-driven interactive analysis and guide the organization's strategic decision analysis, especially when different epochs and levels of information granularity are involved. However, one may need to use several scattered spatial datacubes at the same time. More specifically, users may need to simultaneously and rapidly navigate through different spatial datacubes, to quickly create (virtually or physically) a new spatial datacube from existing scattered ones, or to expeditiously insert information in one spatial datacube using another one.

Simultaneously using different spatial datacubes presents some risks such as the risk of non-accessibility to data stored in spatial datacubes, the risk of non-technical commonality, the risk of hidden spatial referencing differences leading to conflicting locations and shapes, and the risk of misinterpreting data. These risks lie in the fact that spatial datacubes are typically heterogeneous in design, in content and in spatial referencing. For example, a health organization willing to analyze the risk of the West Nile virus to a human population, may use two datacubes; one containing data related to water bodies and the other containing data related to the location of dead birds reported by the population. Both datacubes are modeled differently (e.g., different conceptualizations, purposes) and developed using different techniques. This heterogeneity may lead to the risk of data misinterpretation and hence to the risk of making wrong strategic decisions (e.g., from using different definitions and categories of marshes, swamps, streams, etc. or from using different map coordinates to locate dead birds and water bodies).

Interoperability has been widely recognized as an efficient paradigm for reducing the risk that results from the problem of the heterogeneity when dealing interactively with different systems [14], [5], [17], [6]. It aims at resolving technical, organizational, and semantic heterogeneities between various systems in many fields (information management, engineering technologies, etc.). Although the interoperability of information systems and more especially in the geographic information domain has attracted the attention of many researchers ([8], [5], [17], [6]) and standardisation bodies ([30], [19]), semantic heterogeneity still remains a challenge for enabling interoperability among databases and the risk of simultaneously using different databases still relatively high. In this paper, we discuss the risk related to the simultaneous use of heterogeneous spatial datacubes and we propose an approach that combines both automation and human intervention in order to identify and control such risk.

In the next section, we present the relation between risk management and the use of different datacubes at the same time. In Section 3, we propose an approach to manage the risks that are related to such a simultaneous use of different datacubes. We conclude and present further works in Section 4.

2. RISK RELATED TO THE USE OF DIFFERENT SPATIAL DATACUBES

2.1 Spatial Datacubes

A datacube is based on the multidimensional approach (i.e., a set of measures aggregated according to a set of dimensions [40]). A dimension contains a set of members that are organized hierarchically into levels. The aggregations of measures are typically pre-calculated on all the possible combinations of members are optimized for rapid ad hoc information retrieval and knowledge discovery. Datacubes support the user's mental model of the data and, hence, it is very appropriate for data exploration and strategic decision analysis. Moreover, based on a multidimensional paradigm, tools exploiting datacubes, such as On-Line Analytical Processing (OLAP) tools [32], provide intuitive ways to query data (mostly by simply clicking on the desired data). OLAP tools allows to query data using different methods and specific operators such as drill-down (show details), roll-up (show a more general level of information), drill-across (show another theme at the same level of detail) and swap (change from a dimension to another) [9], [4]. A datacube can be implemented using relational database management systems (RDBMS) alone, a DBMS with a multidimensional server, directly with a multidimensional server, or with hybrid models. In the context of OLAP, these different architectures are known as Relational OLAP (ROLAP), Multidimensional OLAP (MOLAP) and Hybrid OLAP (HOLAP) [32], [43].

In spatial datacubes, both dimensions and measures may contain spatial components (e.g., street address, map coordinates) that allow the spatial visualization of phenomena and spatial data aggregation. These characteristics facilitate the extraction of insight from real-world phenomena and offer a better understanding of the relationships contained within the dataset [4], [34], [26].

2.2 Risk and Risk Management

There are various understandings of the term risk depending on the circumstances in which it is used. However, this term usually refers to the possibility that an undesirable outcome may occur as a result of an event. ISO/IEC considers the notion of risk as a "combination of the probability of occurrence of harm and the severity of that harm" [20].

But how can we qualify the result of an event as undesirable (or harm)? Technical approaches consider "harm" as physical dysfunction or error. Psychological approaches consider subjective judgment and preferences to determine if the outcome of an event as "harm." For example, downhill skiing may be viewed by some people as undesirable while perceived as desirable by others [25]. Also, risk, in economy, may include gains and losses [33].

Risk management refers to the process of reducing the risk to a level considered acceptable by an individual or an organization [28], [33]. Managing risks consists basically of four phases: 1) identifying the risks, 2) assessing the risks (i.e., determining their probability of occurrence and their degree of harm), 3) taking proper actions to reduce the risk to an acceptable level, and 4) documenting the previous phases [25], [20].

But what risk level is considered acceptable? (i.e., How safe is safe enough?). The answer to this question depends on the circumstances that surround each case. For decision analysis, the answer for this question depends on the organization's strategy, organization's business goals, user's skills, available alternatives, possible consequences, organization's legal compliance responsibilities, etc. Thus, there is no single "How safe is safe enough?" problem [10]. However, the acceptable risk can be calculated (cf. technical approaches) or estimated (cf. psychological approaches) with regards to what is prepared to be lost compared with what is to be gained.

For technical approaches, the risk can be objectively identified and assessed by scientific methods (i.e., probabilistic assessment methods). On the other hand, psychological approaches consider human values and preferences as well as the circumstances where undesirable events may occur in order to identify and assess risks [33], [38].

2.3 Risk of Data Misinterpretation Related to the Simultaneous Use of Several Spatial Datacubes

Simultaneously using different spatial datacubes favourites reuse of existing spatial data and improve making strategic decisions that require information scattered in disparate spatial datacubes. However, this reuse is associated with some risk of misusing data. Such a risk may be critical to the strategic success of an organization. This risk may be due to technical aspects, or to the content of spatial datacubes. A disturbing concern related to content of a datacube is the risk of misinterpreting data. Such a risk may have significant drawbacks on the organization. In fact, strategic decisions made on the basis of inappropriate interpretations of data may lead users (i.e., decision analysts) to have inappropriate judgment and to make unwarranted inferences about some aspects of the problem to be solved, and thus to make faulty decisions. The risk of data misinterpretation is due to the fact that:

- in spatial datacubes, data have undergone complex ETL (Extract, Transform and Load) procedures that can impact the meaning of its content. This adds to the fact that data are generally collected from other heterogeneous sources having themselves undergone complex procedures (see [2] for a detailed explanation). During the transformation phase, some interpretations may be formed and several rules, functions and decisions may be applied to the collected data in order to fit business needs (e.g., modifying the terms used to be understood by business people). For example, modifying the legend related to spatial data in order to put the emphasis on certain layers;
- in spatial datacubes, data are also aggregated or summarized using different methods. This aggregation adds another level of complexity of interpretation. In fact, in order to interpret data, spatial datacube users may need to understand first the method or the pattern used for aggregating data since the same source data may be summarized with different methods. For example, the buildings enclosed by four streets can be aggregated to form what we call a building block if the density is higher than a given threshold. In order to truly understand the meaning of such a polygon, users need to know first the criteria used for the aggregation;
- spatial datacube structure promotes a rapid and easy use of data. The rapidity and ease of data use may lead users to 1) misunderstand the inherent characteristics of data, and to 2) to mistakenly feel that data are made-to-order for their decision analysis needs, and hence to deter them from adopting an informed behaviour towards data [24].

If the relevant information about the meaning of spatial datacubes contents (e.g., information about data transformation, method of aggregation) is not explicitly represented, users will have significant concerns related to data interpretation. These concerns are even more serious when we try to use, at the same time, more than one spatial datacube for a specific need. In fact, spatial datacubes are usually modeled differently (e.g., different conceptualizations, languages, intended uses). Consequently, many differences may occur between the content of spatial datacubes (i.e., semantic heterogeneities between spatial datacubes). Semantic heterogeneities may mislead spatial datacubes users to make wrong interpretations. For example, simultaneously using two different spatial datacubes, a user may think that the term "forest" has the same intended meaning in both datacubes, while actually it has not. The term "forest" may refer, in one spatial datacube, to "a wooded area set aside for hunting" and to "an area with a high density of trees" in the other [42]. Another example of a semantic heterogeneity may occur when two concepts have the same nomination but different meanings. For instance, the term "risk" may have more than one meaning: 1) possibility that an unwanted state of reality may occur or 2) economic gains and loses (in the economic field) [33]. Also, a typical example of a semantic conflict may arise when two different terms have the same meaning. For instance, the terms "Brook" and "Burn" may refer to a small stream [42].

We believe that geometric and graphic aspects belong to spatial data semantics. That is, geometric and graphic aspects may convey meanings about spatial data. For example, a sinusoidal line on a map may indicate a road with several left and right curves. Also, on a map, polygons may indicate houses, whereas points may indicate light poles (when they are blue) or fountains (when they are black). The fact that these two types of object were represented with different geometries and graphic visual variables would help us to deduce when visualising the map that an object represented by a polygon is a house, an object represented by a blue point is a pole and an object represented by a black point is a fountain. Moreover, geometric and graphic aspects are not inherent to objects but defined according to the needs of a given application (for example, a polygon representing a building may correspond to the roof and may be measured using photogrammetry for a given application, while it may correspond to the foundations and measured using land surveying for another application).

Consequently, any difference in the geometry or the graphic aspect (e.g., colour, pattern, weight) is indeed considered as a semantic heterogeneity of spatial data. Differences of the geometry include differences in geometric primitives used for the same or related spatial data (i.e., point, line, and polygon), differences in reference systems (for example, North American Datum of 1927 (NAD 27) and North American Datum of 1983 (NAD 83)), differences in topological relations (for example, the related relations cross and intersect are described by different words), different scales, vague vs. clear concepts (e.g., close vs. next to), etc.

In order to cope with the risk of misinterpreting data when simultaneously using different spatial datacubes, we need to identify the causes that may lead to data misinterpretation and propose ways to overcome them. In the following section, we discuss first some approaches and techniques that may be useful to control the risk related of the simultaneous use of different spatial datacubes.

2.4 Overview of Approaches for Managing Risks of Data Misinterpretation

Some approaches and techniques have been proposed to identify the risk of data misinterpretation related to the simultaneous use of different information systems, and suggest ways to reduce such risk. These approaches vary from taking precautions that must be applied to avoid possible risks of data misinterpretation (i.e., prevention approaches) to more complex methods that propose to assist data interpretation (i.e., creative approaches). Both types of approaches may be based on different techniques such as standards (cf. subsection 2.4.3).

2.4.1 Prevention Approaches

Some prevention approaches (also called a priori approaches) define measures to prevent certain predicted risks of data misinterpretation and misuse (that may be caused by a human error or a system failure). These measures vary from simple restriction of data access for certain individuals or groups, to more complex procedures such as training users in order to improve their ability to interpret data, enhancing data selection tools, defining relevant information to help users to assess data quality (e.g., providing metadata or context-sensitive warnings), and developing context-aware systems that help users to better adapt data to their specific use. An example of context-aware systems is NAMA [23]. This system is based on a context-aware agent that predicts user behaviour and helps him/her to interpret the content of commercial web sites in order to make specific purchases.

While they assist users to interpret data by predicting human behaviours and system failures, prevention approaches are not sufficient to reduce the concerns of data misinterpretation to an appropriate level because, first, it is not easy (nor convenient) to predict all possible situations where the risk of misinterpretation can occur. Second, it is extremely difficult to unify the way of defining data related to different domains. Moreover, prevention approaches are usually manually-defined and are not adequate for a large scale use.

2.4.2 Reactive Approaches

Reactive approaches (also called *a posteriori* approaches) propose to take certain actions while using data in order to reduce the risk of data misinterpretation. Some approaches communicate relevant information about data use or provide warnings to end-users when a risky operation is performed (e.g., measuring a distance without having the map units defined) [1], [18], [24] or when a pattern is not considered. Other approaches propose to automatically interpret data (i.e., semantic interoperability between information systems). In this article, we focus on the latter approach.

Semantic interoperability between information systems

Semantic interoperability has been defined as the ability of heterogeneous systems to exchange information and applications in an accurate and effective manner [5], [8], [20]. With the emergence of software agents, the semantic interoperability has been viewed as the technical analogue to human communication [6], [7], [13], [21], [31]. According to this view, each agent interprets the exchanged data as it has been originally intended by another one.

Semantic interoperability aims to resolve the semantic differences (i.e., semantic heterogeneities) that may exist between the content of different information systems [5], [17]. Semantic heterogeneities occur when there are differences about the meaning, the interpretation, or the intended use of the same or related data [36]. These differences may lead to ambiguities in data interpretation. For example, the English term "river" may be interpreted by a French speaker as 1) a stream of water that flows into a sea or as 2) a stream of water that flows into other water bodies. Considering the interoperability as a communication between systems, suppose that a System A sends some data (i.e., a map and a sentence: "The river 2 intersects the river 1", as shown in Fig. 1) to System B. We also suppose that System A uses the English language to model data, while System B uses the French language. Receiving the data from System A (1), System B may wonder about the synonym of the term "river" in French (i.e., "fleuve": a stream of water that flows into a sea, or "rivière : a stream of water that flows into other water bodies ?). If System B is not aware of relevant information that would allow it to appropriately interpret the term "river", it could ask for more information to clarify the meaning of the term "river" (System B may ask, for example, "Which water body do the rivers flow into?" (2), and System A may respond: "The river 1 flows into the sea, the river 2 flows into the river 1" (3)). System B could then appropriately interpret the term "river" (i.e., river 2 is a "rivière" while river 1 is a "fleuve"). Thus, interoperating together, systems could resolve such ambiguities and reduce the risk of data misinterpretation. Hence, semantic interoperability may be seen as an expeditious way to facilitate data interpretation.

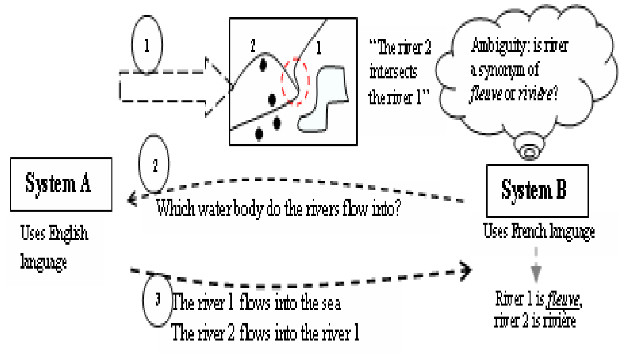

Fig. 1: Interoperating together, systems could resolve ambiguities in interpretations

The semantic interoperability is based on agreements between a set of agents to use a shared vocabulary (i.e. ontology) and on specific circumstances under which the vocabulary can be used (i.e. context). An ontology is a set of related concepts and a set of assumptions about the intended meaning of these concepts in a given domain or application [15], [16]. Ontologies contain some elements of context which are usually defined in the assumptions (for example, fountains should be represented with a point). However, other context elements may vary from a specific use to another (for example, the security measures vary from a province to another) and are normally not included in ontologies.

2.4.3 Standards and the Risk of Misinterpretation

Standards have been widely used in a variety of domains and applications in order to facilitate data interpretation when simultaneously using different information systems. In the context of spatial data, there is a diversity of standards that unify the way in which data can be discovered, described and represented, and hence facilitate data interpretation and information exchange [12]. Examples of standards are the Geography Markup Language (GML[1]) which based on the Extensible Markup Language (XML[2]), and the ISO 19115:2003 Geographic information - Metadata[3] standard. Some authors based their work on standards to overcome the problems of heterogeneity between non-spatial datacubes. Mangisengi et al. [27] and Frank and Chen [11] used XML standard to describe datacubes schema and to define a global schema to integrate heterogeneous datacubes.

1 http://www.opengeospatial.org/standards/gml
2 http://www.w3.org/XML/
3 http://www.iso.org/iso/iso_catalogue/

While standards are useful to ease data interpretation, they are not sufficient to reduce the risk of data misinterpretation to a tolerable level mainly because 1) in some cases, the more you want to get people to agree, the more differences will be found since there is no single geographic reality that overrides all others [29], [17]; 2) there still a lack of widely-accepted standards for data definition and transformation; 3) standards restrict the way humans or systems conceive the reality.

3. AN APPROACH FOR MANAGING THE RISK RELATED TO THE SIMULTANEOUS USE OF SPATIAL DATACUBES

We propose an approach to manage the risk of data misinterpretation when simultaneously using different spatial datacubes. This approach consists of 1) proposing a categorisation of heterogeneity conflicts in order to guide risk identification, 2) a framework of assessing risks, and 3) a model that facilitates the reasoning about context information in order properly interpret spatial data in datacubes.

3.1 A Categorisation of Heterogeneity Conflicts

Semantic heterogeneities may mislead users to reach wrong interpretations. In order to manage such risk, we need to overcome the semantic heterogeneities between spatial datacubes. The heterogeneities may be categorised as following:

1. Cube-to-Cube heterogeneity
   - *Cube-to-Cube context conflicts* arise when two cubes are created or used in different contexts (e.g., different purposes, languages, reference systems, etc.).
2. Dimension-to-Dimension heterogeneity
   - *Dimension-to-Dimension meaning conflicts* arise when related dimensions of different spatial datacubes have mismatched meanings (e.g., different names, definitions, spatial - i.e., difference in geometry or in temporality - or thematic properties). For example, a bridge may be represented with a line or a polygon. These problems may also occur when unrelated dimensions are named equivalently.
   - *Dimension-to-Dimension context conflicts* arise when related dimensions of spatial datacubes have been defined and used in different contexts.
   - *Dimension-to-Dimension hierarchy conflicts* arise when related dimensions of different spatial datacubes have different hierarchies (i.e., different schemas or semantically unrelated levels).
   - *Dimension-to-Dimension level conflicts* arise when semantically related levels of related dimensions have mismatched meanings or different contexts.
3. Fact-to-Fact heterogeneity
   - *Fact-to-Fact meaning conflicts* arise when related facts of spatial datacubes have different meanings (e.g., different names or definitions).
   - *Fact-to-Fact context conflicts* arise when related facts of spatial datacubes have been defined and used regarding different contexts.
4. Measure-to- Measure heterogeneity
   - *Measure-to-Measure meaning conflicts* occur when related measures have different meanings (e.g., different names or scales).
   - *Measure-to-Measure context conflicts* occur when related measures have different contexts.

Frank and Chen have defined a similar categorisation of heterogeneity conflicts in non-spatial datacubes [11].

3.2 Assessing the Risk

Risk assessment aims at determining the level of risk (i.e., the expected degree of error or dysfunction). Most assessment methods use a matrix with two axes that represent the probability that a risk will occur and the severity of the consequences if it occurs. These methods consider the severity and the probability as one-dimensional and provide a value of the overall severity and probability of the risk [22], [24]. But actually, the risk level (i.e., probability and severity of consequences of the risk of data misinterpretation) depends on the context in which data have been defined and used. For example, interpreting data modeled with a graphical language (such as UML[4]) is far from being obvious to people who are not used to work with this technique. In the field of decision analysis, the risk level depend on several aspects such as the impact of the risk on decision analysis, legal constraints, organization's goals, end user's skills, and other circumstances that surround the decision analysis process. Thus, we consider the severity

4 ww.omg.org/technology/documents/formal/uml.htm

and the probability of data misinterpretation risk as multidimensional concepts that depend on the context that surrounds the simultaneous use of different spatial datacubes. In order to guide the assessment of data misinterpretation risk and the definition of the acceptable level of this risk, we propose a multidimensional structure that allows to explicitly represent the context elements that may impact the level of risk. This structure includes a fact that represents the overall level of consequences related to the risk of data misinterpretation (i.e., Risk level fact). This fact has two measures that evaluate both the severity and the probability of risk. The fact is determined according to different dimensions that represent the contexts that may affect the risk of data misinterpretation (e.g., organization's strategic goal, domain, legal aspects), see Fig.2.

For example, in order to assess the risk of misinterpreting the term "forest management", an organization may consider the following aspects: the application, its goal, the legal aspect, the education degree of users, the region and period of interest. We suppose that the goal of the organization is to determine the evolution of the wood volume of Montmorency forest (Quebec City) in two different periods (1980 and 2000), that legal constraints are not important, and that users are novice to the geographic field (especially to the forestry domain). According to these aspects, the organization manager (or an expert) may consider that the overall risk of misinterpreting the term "forest management" is high (the probability of risk is moderate and the severity of such risk is high). So, for decision analysis, each organization, or even each manager, may have its own strategy and procedure to assess the risk of data misinterpretation (that is, its own aspects of assessing risk and its own evaluation of the risk according to these aspects). These aspects (context elements of assessing the risk) can be determined only if we take into account psychological profiles. To define these profiles, human intervention is required.

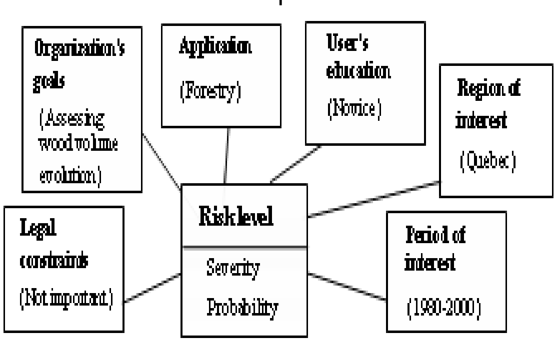

Fig. 2: A multidimensional structure for assessing the risk of data misinterpretation

3.3 Managing the Risk

3.3.1 A Framework for the Semantic Interoperability of Spatial Datacubes

In [35], we motivated the need for interoperating spatial datacubes, and we defined an agent communication framework which is based on a *Context Agent* that helps agents, representing spatial datacubes (called *Datacubes Agents*), to appropriately interpret data exchanged between them, see Fig. 3. We generate two ontologies for each *Datacube Agent*: one from the datacube model and another from metadata related to that model. The *Context Agent* defines and explicitly represents the context elements using ontologies. These elements allow *Datacubes Agents* to appropriately interpret the concepts of their spatial datacubes.

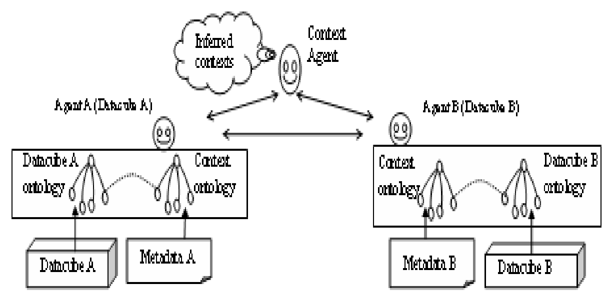

Fig. 3: A communication framework between agents representing spatial datacubes [35]

For example, *Agent A* (representing a spatial cube developed in the province of Ontario using the English language) communicates the message "The river R crosses the forest F" to *Agent B* (representing a spatial cube developed in the province of Quebec using the French language). The context elements of this model are:

- *Goal Context*: building bridges.
- *Domain Context*: Forest management domain.
- *Dataset Context*: The language used is English. The modelling language is UML.
- *Concept Context*: Time of the concept definition was 1995. The place is Ontario.

The *Context Agent* helps agents communicating together to interpret the data within its context. For example, it helps *Agent B* to figure out what *Agent A* means by the concept "river"; will river be interpreted as "rivière" or "fleuve" (in French)? *Context Agent* would explicitly represent the information related to the context in which concepts have been defined and used, and helps agents to reason about this information to appropriately interpret data. For example, *Context Agent* would notify *Agent B* that the language used by *Agent A* is English. Then, *Agent B* would ask for more information, for example "Which water body does the river flows into?", and *System A* may respond "The river flows into the sea". *System B* could then conclude that *Agent A* means a "fleuve" and not a "rivière (in French). Moreover, *Context Agent* would notify the possible differences in contexts to avoid data misinterpretation. For example, *Context Agent* would draw the attention of datacubes agents to the potential difference in the conceptualisation of the forest management domain (i.e., domain context) between the two provinces (i.e., Quebec and Ontario). For that, Context Agent may ask both agents "do you consider both public and privately-owned forests in the forest management?".

3.3.2 Combining Automation and Human Intervention

While semantic interoperability represents an effective way to facilitate data interpretation, it still has a considerable vulnerability to the risk of data misinterpretation. In fact, it may be extremely difficult (even impossible) for machines to take into account a certain context information that could be useful to appropriately interpret the content of different spatial datacubes. For example, consider the following message from *Agent A* to *Agent B*: "restaurants near Chateau Frontenac". In order to appropriately interpret the spatial concept "near", we may need to be aware of *Agent A*'s attitudes and preferences. *Agent A* may consider, for example, that restaurants within 5 kilometres to Chateau Frontenac to be "near". Such context information (i.e., agent's attitudes and preferences) may be relevant to appropriately interpret data. Taking into account such information requires an extensive knowledge and judgement which are unique to humans and cannot be completely automated [39]. Therefore, we propose a more global approach for managing the risk related to data misinterpretation when simultaneously using different spatial datacubes. This approach involves a human intervener that would support spatial datacubes interoperability (e.g., the intervener would extract explicit and implicit context elements, change the nomination of a given concept, add or delete concepts, adjust the context of concepts, infer other information from existing ones, make judgement about the interpretation of concepts and match them). In the ideal case, the intervener would be a group that consists of datacubes designers. Moreover, at least one of the members of the group would have experience in dealing with spatial data, in business intelligence (analysis decision process) and in other fields that may surround the interpretation and the use of spatial datacubes contents (e.g., legal aspects).

In order to guide the intervener, we defined a model that helps human intervener to reason about the semantic of the spatial datacubes contents and to appropriately interpret them.

3.3.3 A Model for Managing the Risk Related to the Simultaneous Use of Spatial Datacubes

In order to manage the risk of data misinterpretation when simultaneously using several spatial datacubes (i.e., to explicitly represent meanings and to overcome the semantic heterogeneity of spatial cubes), we should guide agents (i.e., machine or humans) to understand the concepts of these datacubes and reason about their semantic. We define semantic regarding the elements of ontologies (i.e., concepts, definitions, assumption, properties such as thematic, geometric, graphic and temporal aspects) and the elements of context of spatial datacubes concepts (e.g., language, techniques used to define spatial objects, etc.). Both ontology and context elements define the semantic characteristics of spatial data cubes concepts.

Inspired by the VUEL concept (View Element) [3], we introduce a model that is based on multidimensional structure called *SemEL* (i.e., Semantic Element) where ontology, context ans spatial characteristics represent the facts (as shown in Fig. 4). This model is in agreement with the human's mental model of the data and, hence, it facilitates analysis and reasoning about data represented in a hierarchical structure. It allows users to explore and navigate across different themes (i.e., dimensions) at different levels of detail and to rapidly visualize the information at the intersections of these dimensions. Moreover, based on the multidimensional paradigm, many tools, such as SOLAP tools (Spatial OLAP [34]), allow user to visual different views that represent different combinations of information. This is typically done in an intuitive ways. *SemEL* consist of three cubes (i.e., ontology, context and spatial cubes). Ontology cube has three dimensions (i.e., *Definition*, *Assumption*, and *Property*) and a fact table that has the ontology description of datacubes concepts ("Ont_Descrip") as its unique measure. "Ont_Descrip" will contain textual definition, geometry, graphical and temporal properties, as well as axioms. Context cube is defined according to four dimensions: *Goal Context, Domain Context, Dataset Context, and Concept Context* and a fact table that has the description of context

("Context_Descrip") as its unique measure. Spatial cube is defined according to three dimensions: *Geometry*, *Graphics* and *Position* and a fact table that has the description of spatial characteristics ("Spatial_Descrip") as its unique measure.

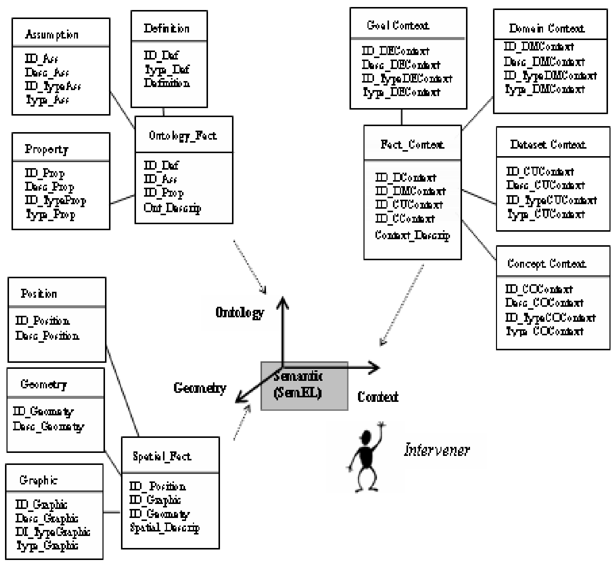

Fig. 4: A model to support data interpretation (*SemEL*)

Since it is based on multidimensional structure, *SemEL* allows interveners to have a global picture of context, and to get more insight into the details of that context (see Fig. 5). More specifically, the model allows to:

- Facilitate the identification and estimation of the risk of data misinterpretation. For example, the concept "contain" refers to a, inclusion relation between an object A and object B where 2) both interior and limit of B are completely within the interior of A (e.g., in ArcMap[5]), or where 2) the limits of both objects may touch each other (e.g., in Oracle9*i*[6]). Navigating within the dimension *Dataset Context* (the members ArcMap and Oracle9*i*), the intervener could figure out that there is a risk related to the context heterogeneity of the concept "contain".
- Facilitate the appropriate data interpretation (i.e., data defined in the specific context, represented with a specific geometry, a specific graphic, and according a predefined assumption). For instance, as it is shown if Fig.5, the meaning of the signal *contain* can be determined by 1) its definition within an ontology: "Topological relation that links two spatial objects", and its context elements: ArcMap as the tool used to specify the relation. Consequently, relation contain is interpreted as "a relation between two objects A and B where the interior of B are completely within the interior of A, the limits of both objects may touch each other".
- Facilitate the conversion of context and analyse the impact of that change on data interpretation. In fact, using *SemEL*, the intervener can rapidly navigate through different dimensions, modify one or more context aspects and reason about the effect of that change on data interpretation. For example, as it is shown if Fig.5, if we change the context information about the tool from ArcMap to Oracle9*i*, the interpretation of this relation will be affected (i.e., the interior and limit of B are completely within the interior of A.).

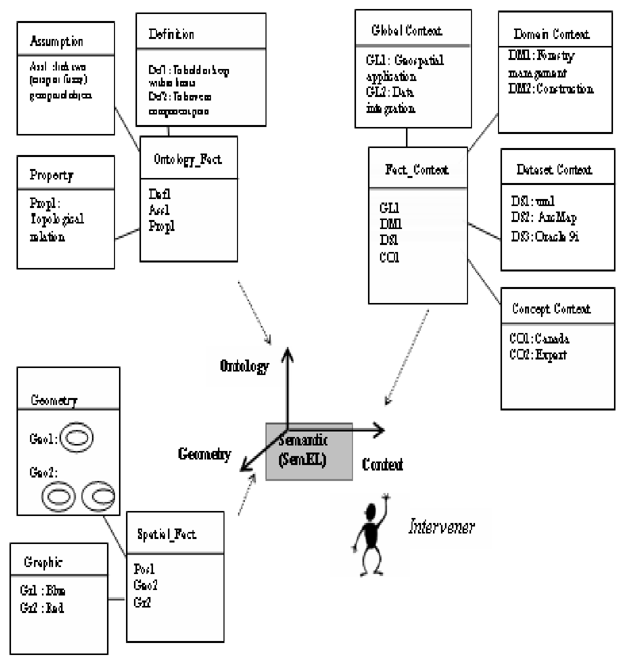

Fig. 5: *SemEL* of "Contain"

Fig. 6 shows an example of a *SemEl* implementation (context cube) using SOLAP technology. The user can easily explore data by navigating from one aspect to another, and compare context and spatial aspects related to data. The example illustrates the potential *SemEl* to support human intervener to rapidly navigate through different levels of context information and reason about this information. As such, the intervener can properly interpret data and thus reduce the risk of misinterpretation.

[5] www.rockware.com/
[6] www.oracle.com/technology/software/products/oracle9i

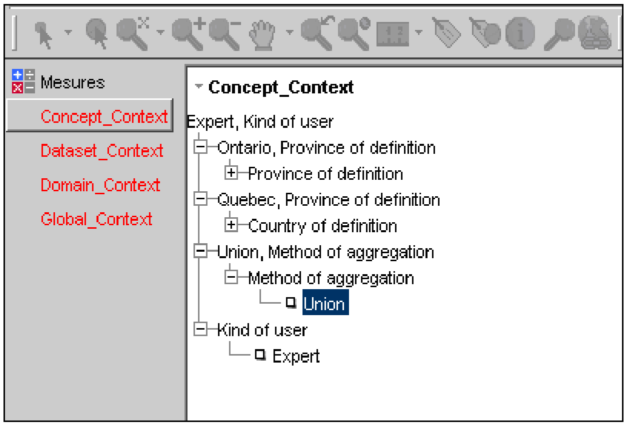

Fig. 6: An implementation of *SemEL* with SOLAP

4. CONCLUSIONS AND FUTURE WORK

In this paper, we explained the risk of data misinterpretation when simultaneously using different spatial datacubes, and we propose an approach to deal with such a risk. The approach is based on the combination of semantic interoperability and human intervention. We also proposed a model (i.e., *SemEL*) to guide human intervenes in their decision. Thanks to its multidimensional structure, *SemEL* forms a mental model of data semantics, thereby helping the intervener to gain insight into it, understand it and reason about it. Thus, human interveners are able to identify and reduce the risks of misinterpretation related to the simultaneous use of spatial datacubes.

Further work is required to refine *SemEL*, to further guide the reasoning about datacubes semantics (e.g., inference).

5. ACKNOWLEDGMENT

We wish to acknowledge the contribution of the NSERC Industrial Research Chair in Geospatial Databases for Decision Support.

6. REFERENCE

1. Beard, K.: Use error: the neglected error component. In Proceedings of Autocarto 9, Baltimore, MD, USA, pp. 808-817 (1989)
2. Bédard, Y.: A Study of Data using a Communication based Conceptual Framework of Land Information Systems. Le Géomètre Canadien, Vol. 40, No. 4, pp. 449-460 (1986)
3. Bédard, Y., Bernier, E.: Supporting Multiple Representations with Spatial View Management and the Concept of "VUEL". Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, Ottawa, Canada, 7-8 July (2002)
4. Bédard, Y., Merrett, T., Han J.: Fundamentals of spatial data warehousing for geographic knowledge discovery. Geographic Data Mining and Knowledge Discovery, H. J. Miller and J. Han (eds.), London: Taylor and Francis, pp. 53-73 (2001)
5. Bishr, Y.: Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science, Vol. 12, No. 4, pp. 299-314 (1998)
6. Brodeur, J.: Interopérabilité des données géospatiales: élaboration du concept de proximité géosémantique. PhD dissertation, D. of Geomatics Sciences, U. Laval, Canada (2004)
7. Brodeur, J., Bédard, Y.: Geosemantic Proximity, a Component of spatial Data Interoperability. International Workshop, Semantics of Enterprise Integration, ACM Conference on OOPSLA (2001)
8. Brodie, M.L.: The promise of distributed computing and the challenge of legacy information systems. In Proceedings of the IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems. Lorne, Victoria, Australia (1992)
9. Codd, E.F., Codd, S.B., Salley C.T.: Providing OLAP (On-Line Analytical Processing) to User- Analysts: An IT Mandate. E. F. Codd and Associates, pp. 24 p (1993)
10. Derby, S.L., Keeney, R.L.: Risk analysis: understanding how safe is safe enough. Risk Analysis, 1(3), pp. 217–224 (1981)
11. Frank, S.C., Chen C.W.: Integrating Heterogeneous Data Warehouses Using XML Technologies. Journal of Information Science, Vol.31, No. 3, 2005, pp. 209-229 (2005)
12. Fonseca, F., Egenhofer, M., Davis, C., Borges, K.: Ontologies and Knowledge Sharing in Urban GIS. Computers, Environment, and Urban Systems, 24(3), pp. 25 l-272 (2000)
13. Gazzaniga, M.S.: Mind Matters, Houghton Miffin Co., ISBN 0-395-42159-4 (1988)
14. Goodchild, M.F., Egenhofer, M.J., Fegeas, R.: Interoperating GIS. Report Specialist Meeting under Auspices of the Varenius Project, U.S. NCGIA, St Barbara, USA (1997)
15. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, Vol. 43, Nos. 5-6, pp. 907–928 (1993)
16. Guarino, N.: Formal Ontologies and Information Systems. In Guarino N. (Ed.), Proceedings of FOIS98, IOS Press, Amsterdam, pp. 3-15 (1998)
17. Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C.: Semantic Interoperability: A Central Issue for Sharing Geographic Information. The Annals of Regional Science, Special Issue on Geo-spatial Data Sharing and Standardization, pp. 213-232 (1999)
18. Hunter, G.J., Reinke, K.J.: Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations. M Lemmens and G Heuvelink (eds.). Proceedings of 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Amsterdam, pp. 313-319 (2000)
19. ISO: http://www.iso.org/iso/iso_catalogue/ (Last visted January 5, 2008)
20. ISO/IEC Guide 51: Safety aspects – Guidelines for their inclusion in standards (1999)
21. Janowicz K.: Service Composition and Interoperation with DAML-S: A critical view, Diploma thesis, Institute for Geoinformatics, University of Münster (2003)
22. Kerzner, H.: Project Management: a systems approach to planning, scheduling, and controlling (9th Edition), John Wiley & Sons, New Jersey, 1004 p (2006)
23. Kwon, O.B., Choi, S.C., Park, G.R.: NAMA: a context-aware multi-agent based web service approach to proactive need identification for personalized reminder systems. Expert Systems with Applications. Vol. 29, pp. 17-32 (2005)
24. Levesque, M.-A., Bédard, Y., Gervais, M., Devillers, R.: Towards managing the risks of data misuse for spatial datacubes. ISSDQ 2007. Enchede, The Netherlands (2007)
25. Machlis, G.E., Rosa, E.A.: Desired risk: Broadening the social amplification of risk framework. Risk Analysis, Vol. 10, pp. 161-68 (1990)

26. Malinowski, E., Zimanyi, E.: Representing spatiality in a conceptual multidimensional model. In Proceedings of the 12th ACM Symposium on Advances in Geographic Information Systems, pp. 12–21 (2004)
27. Mangisengi, O., Huber, J., Hawel, C., Essmayr, W.: A framework for supporting interoperability of data warehouse islands using XML. In Proceedings of the 3rd Conference on Data Warehousing and Knowledge Discovery, LNCS, Vol. 2114, pp. 328-38 (2001)
28. Morgan, M.G.: Choosing and Managing Technology-Induced Risks. T.S. Glickman and M. Gough (eds), Readings in Risk, pp. 5-15 (1990)
29. Nyerges, T.: Representing geographical meaning. Buttenfield BP, McMaster RB (eds), Map Generalization: Making Rules for Knowledge Representation, pp. 59-85 (1991)
30. OGC: http://www.openspatial.org/ (Last visted January 5, 2008)
31. Ogden, C.K., Richards, I.A.: The Meaning of Meaning: a study of the influence of language upon thought and of the science of symbolism. Routledge & Kegan Paul (1923)
32. OLAP Council: OLAP and OLAP Server Definitions http:/www.olapcouncil.org/research/glossarjfy.htm (1995)
33. Renn, O.: Three decades of risk research: accomplishments and new challenges. Journal of Risk Research, Vol. 1, No.1, pp. 49-71 (1998)
34. Rivest S. Rivest, Bédard Y., Marchand P.: Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP). Geomatica, Vol. 55, No. 4, pp. 539-555 (2001)
35. Sboui T., Bédard Y., Brodeur J., Badard T.: A Conceptual Framework to Support Semantic Interoperability of Spatial Datacubes. J.-L. Hainaut et al. (Eds.), ER Workshops 2007, LNCS, Vol. 4802, pp. 378–387 (2007)
36. Sheth. A., Larson, J.: Federated Database Systems for Managing Distributed, Heterogeneous and Autonomous Databases. ACM Computing Surveys, Vol. 22, No. 3, pp. 183-236 (1990)
37. Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C.: Past, present and future of decision support technology. Decision Support Systems, Vol. 33, pp. 111-126 (2002)
38. Slovic, P.: Perception of risk, Science, Vol. 236, No. 4799, pp. 280-285 (1987)
39. Swanson, D.R., Smalheiser, N.R.: An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artificial Intelligence, Vol. 91, No.2 pp. 183 -203 (1997)
40. Thomsen, E., Spofford, G., Chase, D.: Microsoft OLAP Solutions, Wiley & Sons, 495 p. (1999)
41. Turban, E., Aronson J.E.: Decision Support Systems and Intelligent Systems. 6th edition, Prentice Hall. (2000)
42. Wikipedia: http://www.wikipedia.com/ (Last visted January 10, 2008)
43. Yougworth, P.: OLAP Spells Success for Users and Developers. Data Based Advisor, pp. 38-49 (1995)