A Conceptual Framework to Support Semantic Interoperability of Geospatial Datacubes

Tarek Sboui[1,2], Yvan Bédard[1,2], Jean Brodeur[3,1], and Thierry Badard[1]

[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

Abstract. Today, we observe a wide use of geospatial databases that are implemented in many forms (e.g. transactional centralized systems, distributed databases, multidimensional datacubes). Among those possibilities, the multidimensional datacube is more appropriate to support interactive analysis and to guide the organization's strategic decisions, especially when different epochs and levels of information granularity are involved. However, one may need to use several geospatial multidimensional datacubes which may be heterogeneous in design or content. Overcoming the heterogeneity problems in a manner that is transparent to users has been the principal aim of interoperability for the last fifteen years. In spite of successful initiatives and widespread use of standards, today's solutions do not address yet geospatial datacubes. This paper aims at describing the interoperability of geospatial datacubes, defining the semantic heterogeneity problems that may occur when using different geospatial datacubes, and proposing a conceptual framework to support semantic interoperability of these datacubes.

Keywords: Geospatial datacubes, interoperability, semantic heterogeneity, ontology, context.

1 Introduction

In the last decades, we have witnessed an exponential increase in the amount of data available from multiple sources. Additionally, there have been important innovations in geospatial[1] information technology, especially in decision support systems, geographic knowledge discovery and interoperability. 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 defined in the field of Business Intelligence (BI). This paradigm enables making strategic decisions by supporting the user's mental model of data. It allows users to navigate

[1] The term "Geospatial" is used in this paper to refer to spatial information for objects located on the Earth (e.g. spatial reference systems, relationships between objects, constraints, etc.).

J.-L. Hainaut et al. (Eds.): ER Workshops 2007, LNCS 4802, pp. 378–387, 2007.

aggregated and summarized data according to a set of dimensions with different hierarchy [9], [4], [16], [15]. Multidimensional databases (hereafter called datacubes) are becoming widely used in the geographic field [4], [16], [15]. Geospatial datacubes contain spatial, temporal and thematic data which may differ in format and content (e.g. geospatial characteristics such as location and geometry, levels of abstraction and meanings) resulting in heterogeneous geospatial datacubes. The heterogeneity of geospatial datacubes presents a major obstacle when people need to make strategic decisions or discover spatio-temporal trends using information located in different geospatial datacubes. For example, a health organization, willing to analyze the risk of the West Nil virus on population, may use two geospatial datacubes; one containing data related to forest stand and the other containing data related to the population density. Both geospatial datacubes are modeled differently (e.g. different conceptualizations, purposes, etc.) and developed using different techniques.

Interoperability has been widely recognized as an efficient paradigm for joining heterogeneous systems to facilitate an efficient exchange of information [5], [6], [10] [13]. It aims at resolving technical, organizational, and semantic heterogeneities between various systems in many fields (information management, engineering technologies, etc.). However, resolution of semantic heterogeneity still remains a challenge for enabling interoperability among databases [13], [6].

Although the interoperability of information systems and more especially in the geographic information realm has attracted the attention of many researchers [8], [5], [13], [6], there have not been many works on the semantic interoperability among datacubes. Also, no work on semantic interoperability of geospatial datacubes has been found in the literature. The purpose of this paper is to discuss the need for interoperating geospatial datacubes and to propose a conceptual framework to support the semantic interoperability among geospatial datacubes.

In the next section, we present the need for interoperating geospatial datacubes, we define the interoperability between geospatial datacubes, and we classify the semantic heterogeneity problems that may occur in these datacubes. In section 3, we present a conceptual framework to overcome the semantic heterogeneity in geospatial datacubes. We conclude and present further works in section 4.

2 Interoperability and Datacubes

2.1 Interoperability Between Information Systems

Interoperability has been described in various ways. It has been generally defined as the ability of heterogeneous systems to communicate and exchange information and applications in an accurate and effective manner [5], [6], [14].

Geospatial interoperability is considered here as the ability of information systems to a) communicate all kinds of spatial information about the Earth and about the objects and phenomena on, above, and below its surface, and b) cooperatively run applications capable of manipulating such information [17]. Brodeur and Bédard defined the interoperability as a communication between different agents who use their common background to establish a common understanding of reality [7]. Brodie proposed an elegant definition of geographical databases interoperability: “Two geographical databases X and Y can interoperate if X can send requests for services R to Y on a mutual understanding of R by X and Y, and Y can return responses S to X based on a mutual understanding of S as responses to R by X and Y" [8], [5].

Typically, there are three levels of interoperability among geographic information systems: technical, semantic, and institutional [10]. The technical level aims to resolve the differences in format, languages and user interfaces. The semantic interoperability aims to provide a mutual understanding of different meanings. For example, the term forest may have more than one meaning: 1) "a defined area of land formerly set aside in England as a royal hunting ground" or 2) "a land that is covered with trees and shrubs" [1]. Finally, at the institutional level, the interoperability aims at unifying the organization process. For spatial information systems, we include the consideration about object's geometry in the semantic level since geometry is 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).

2.2 Geospatial Datacubes

Data warehouses are being considered efficient components of decision support systems. They may be structured as datacubes, i.e. according to the multidimensional paradigm, and implemented using database management systems (DBMS) with a multidimensional server, directly with a multidimensional server, or with hybrid models. From an On-Line Analytical Processing point of view (OLAP), these different architectures are known as Relational OLAP (ROLAP), Multidimensional OLAP (MOLAP) and Hybrid OLAP (HOLAP) [4], [9], [15], [16].

A datacube is composed of a set of measures aggregated according to a set of dimensions with different levels of granularity. Both dimensions and measures of a geospatial datacube may contain geospatial components [4]. Geospatial datacubes support the user's mental model of the data and enables him to make strategic decisions [2], [4], [16]. In fact, they enable users to interactively navigate through different levels of granularity so; he can get a global picture of a phenomenon, and can get more insight into that phenomenon detailed information. Moreover, geospatial datacubes contains spatial component (e.g. street address, geographic coordinates, map coordinates) which allow the visualization of phenomena and, hence, help users to extract insights that can be helpful to understand these phenomena [2].

2.3 Need for Interoperability Between Geospatial Datacubes

In many situations, we may need to use several heterogeneous geospatial datacubes. We group these situations into three categories:

1. *A simultaneous and rapid navigation through different datacubes*: Users from different disciplines may need to access and navigate simultaneously through heterogeneous geospatial datacubes. Navigating separately through each datacube would be an arduous work for users, since they likely need to make extra efforts to manually resolve the problems of heterogeneity between datacubes (e.g. comparing the meaning of concepts, establishing a mapping between them, etc.). The principal aim of interoperability is to automatically overcome such differences and, hence,

can considerably facilitate the navigation task. Interoperability of geospatial datacubes would enable, for example, to create a temporary multidimensional structure (i.e. structure without data records) that is connected to the different dimensions of geospatial datacubes. Such structure is useful especially in emergency situations when users need to rapidly navigate through data stored in the geospatial datacubes without preoccupying themselves with the problems of heterogeneity. An example of a situation is a natural disaster that affects adjacent countries. In such situation, we may need to navigate different geospatial datacubes, developed in these countries, in order to get the right information and respond quickly at different levels (local, provincial, federal, etc.) or at different domains (geographic, political, etc.). In such cases, interoperability of geospatial datacubes is crucial to prevent catastrophic losses.

2. *A rapid insertion of data in a datacube*: While data in datacubes are usually collected from legacy systems, they can be imported from other heterogeneous datacubes [4]. We may need to rapidly insert new data in a geospatial datacube from other datacubes. For example, we may need to rapidly insert new data in a geospatial datacube which contains data about the construction of winter bridges, from two other geospatial datacubes; one of them contains data about traffics. The other cube contains data about lakes.
3. *An interactive and rapid analysis of phenomena changes*: In order to analyze phenomena changes (such as the forest stand evolution), we need to compare data describing these phenomena at different epochs. We may need to compare data stored in geospatial datacubes build also at different epochs. Interoperating geospatial datacubes would enable to interactively compare data and analyze phenomena changes. For example, we may need to rapidly compare forest stand changes following an environmental disaster. Interoperability of geospatial datacubes would enable to detect changes in the wood volume.

However, data in datacubes are usually collected from different source systems. One may ask "why shouldn't we just interoperate the source systems rather than datacubes?" There are three reasons for interoperating geospatial datacubes:

1. We possibly have no more access to data source systems from which we created the datacubes due to multiple reasons including administration policies (such as security reason).
2. We need to use long-period data (i.e. historic data) that usually exist only in datacubes. In fact, in source systems, long-period data are usually modified or replaced by new data and then destroyed or archived, whereas datacubes keep historic data for strategic decision-making purposes [4]. So if we need to reuse such data, we have no choice but to consider datacubes.
3. In the context of decision-making, interoperating datacubes is more efficient than interoperating source systems. In fact, within a geospatial datacube, contrary to source systems, possible aggregations of measures for all possible combinations of members are pre-calculated using different operators (e.g. mathematical, metric such as distance and area, and topological relations such as disjoint, interior intersection, etc.) [16]. These aggregations usually require an arduous work for geospatial datacubes developers (e.g. defining procedures for aggregation, defining a new spatial layer as an aggregation of others). Reusing datacubes means that we don't need to define such aggregations from scratch, thus saving time and money.

2.4 Defining the Interoperability Among Geospatial Datacubes

Inspired from the definition provided by Brodie [8], we define *the interoperability among two geospatial datacubes C1 and C2 as the ability of C1 to request a service in a manner that can be understood by C2, and the ability of C2 to respond to that request in a manner that can be understood by C1. The request and response are conducted automatically*. Services could include:

- importing/exporting data contained in dimensions or facts;
- comparing a dimension/fact against another;
- getting information about the dimensions or the facts (e.g. language used);
- taking into account a concept evolution (e.g. meaning or format changes); or
- adapting the meaning of a concept when the context changes.

While at first sight, there may not seem to have differences between semantic interoperability of transactional databases and that of datacubes, the latter particularly stresses the importance of dealing with the semantics of aggregation and generalization relationships, the semantics of summarizing methods and algorithms, the semantics of summarizability conditions and the semantics of cross-tabulations for every level of details and every member of the datacube dimensions. Although one could do it using traditional transactional solutions, the efficiency can be improved with an enriched framework that explicitly supports datacubes.

We are especially interested in the semantic interoperability which still remains the main issue for the geographic information systems communities [5], [13]. Accordingly, the remaining part of this paper focuses specifically on semantic interoperability of geospatial datacubes. Such interoperability aims at overcoming semantic heterogeneity problems in geospatial datacubes.

2.5 Semantic Heterogeneity Conflicts in Geospatial Datacubes

Semantic heterogeneity represents a major challenge for enabling interoperability among information systems [5], [6]. It occurs when there are differences in the meaning or the interpretation of the related concepts (i.e. concepts having a similarity, generalization, or specialization relationship). For geospatial databases, the semantic heterogeneity may come from the difference in the abstraction of concepts (for example, the concept forest may be represented, with different geometries, as vegetations, as trees, or as wooded areas) and databases schemas [6]. These problems appear more difficult to deal with when working with geospatial datacubes since concepts in these databases are typically more complex (dimensions with different levels, measures to analyze, etc.). In order to support geospatial datacubes interoperability, we first propose a classification of semantic heterogeneity types that may occur between different geospatial datacubes.

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, etc.).
2. Dimension-to-Dimension heterogeneity
   - *Dimension-to-Dimension meaning conflicts* arise when related dimensions of different datacubes have mismatched meanings (e.g. different names, definitions,

geospatial – 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 datacubes have been defined and used in different contexts.
- *Dimension-to-Dimension hierarchy conflicts* arise when related dimensions of different datacubes have different hierarchies.
- *Dimension-to-Dimension level differences* arise when dimensions of datacubes have different levels.

3. Fact-to-Fact heterogeneity

- *Fact-to-Fact meaning conflicts* arise when related facts of datacubes have different meanings (e.g. different names or definitions).
- *Fact-to-Fact context conflicts* arise when related facts of 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.

In order to overcome these conflicts, we defined a conceptual framework based on human communication, ontology, context and multidimensional structure.

3 A Conceptual Framework to Support Semantic Interoperability Between Geospatial Datacubes

3.1 Human Communication, Ontology, Context, and Semantic Interoperability of Geospatial Datacubes

The conceptual framework presented in this paper is based on human communication, ontology and context. The human communication process constitutes an improved form of interoperability [6]. It 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 [12], [11]. Ontologies play an important role in enabling semantic interoperability between agents by providing them a common understanding of the reality. Ontology of geospatial datacubes would include definitions, assumptions, and properties of the datacubes concepts. The properties include non-spatial and spatial aspects (i.e. geometries and graphical representations).

Context is any information that surrounds and facilitates the interpretation of concepts. Ontologies contain some elements of context which are usually defined in the assumptions (for example, fountains should be represented with a point). However, other elements of context 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. In order to support semantic interoperability of geospatial datacubes, we aim to explicitly identify all possible elements of context related to geospatial datacube. To guide context definition, we identify four context levels: *Goal Context* level which defines the purpose for which the geospatial datacubes will be used (such as population density), the *Domain Context* level which contains the context elements of the domain (such as forest stand), the *Dataset Context* level which consists of elements related to dataset of geospatial datacubes (such as the specification used to describe concepts), and *Concept Context* level which includes the characteristics of dimensions, or measures of datacubes (role, properties, etc).

3.2 A Theoretical Framework for Geospatial Datacube Interoperability

Some research works on interoperability were based on the communication process between people [7], [18]. Brodeur and Bédard defined a conceptual framework for spatial data interoperability based on human communication [7]. They considered the context of a given concept as a) the meaning of this concept and b) its dependencies with other concepts. We build our work on their point of view; however, we define a richer model that explicitly represents not only elements extracted from datacubes metadata, but infers new context information from these elements. We define an agent communication framework which is based on a *Context Agent* that helps agents, representing geospatial datacubes (called *Datacubes Agents*), to appropriately interpret information exchanged between them (see Fig.1). 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, and stores them in what will form the context knowledge base that will be available for *Datacubes Agents,* to reason about and interpret the concepts of their geospatial datacubes.

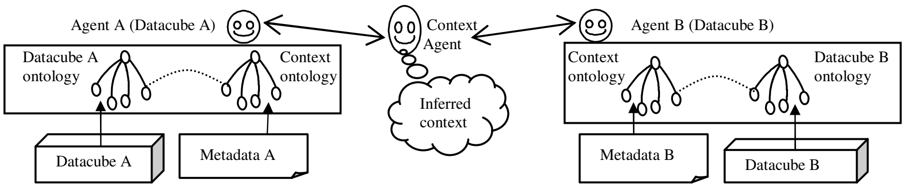

Fig. 1. A communication between agents representing geospatial datacubes

For example, *Agent A* (representing a geospatial cube developed in Ontario using English) communicates the following information to *Agent B* (representing a geospatial cube developed in Quebec using French): "The river *R*, represented with a line, intersects the forest *F*". The context elements of this model are:

- Goal Context: building bridges.
- Domain Context: geographic domain, civil engineering (large scale map).
- Dataset Context: The language used is *English*. The modeling language is UML.
- Concept Context: Time of the concept definition was 1995. The place is Ontario.

The *Context Agent* helps *Agent B* to interpret the data within its context (i.e. figure out what *Agent A* means by the concept *river*). In other words; will *river* be interpreted as *rivière* or *fleuve* (in French)? In order to make the appropriate interpretation, agents would reason about the semantic of the concept *river*, including its definition, its geometry and its context. We defined a model to reason about the semantics of the concepts in order to facilitate their interpretation and hence support the interoperability between geospatial datacubes.

3.3 A Model for Semantic Interoperability of Geospatial Datacubes

We believe that, in order to overcome the semantic heterogeneity of geospatial cubes, we should 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 geospatial datacubes concepts (e.g. language, techniques used to define spatial objects, etc.). Both ontology and context elements define the semantic characteristics of geospatial data cubes concepts. In order to guide the reasoning about the concepts semantics, and 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 and context represent the facts (see Fig.2). This model enables to explicitly represent the meaning and to define a relevant interpretation of a concept regarding the ontology and the context in which it has been defined and used. The ontology model has five dimensions (i.e. definitions, assumptions, geometries, time and graphical representations) and a fact table that has the ontology description of datacubes concepts (*Ont_Desc*) as its unique measure. *Ont_Desc* will contain textual definition, geometry, graphical and temporal properties, as well as axioms. The context model is defined according to four dimensions (i.e. Goal Context, Domain Context, Dataset Context, and Concept Context) and a fact table that has the description of context (*Context_Desc*) as its unique measure.

Since it is based on multidimensional structure, *SemEL* enables to rapidly navigate from one level to another and from dimension to another and apply reasoning capabilities (e.g. inference) to draw conclusions based on relations between semantic elements (i.e. ontology and context elements). For example, if the term *Forest* was used in England's royal context, then by inference, this term can be interpreted as a "hunting ground". More specifically, the model would allow to:

- provide the appropriate meaning of a concept (i.e. the concept defined in the specific context, represented with a specific geometry, a specific graphic, in a specific date and according a predefined assumption). For instance, the meaning of the concept *river* can be determined by 1) its definition within a general ontology: "Natural stream of water that flows in a channel" [1], and a general assumption specifying that it flows into the sea, 2) its geometry: *line*, 3) its graphic: *blue* and 4) its context elements: *English* as the language used, agriculture as the domain in which the concept is used, etc. Consequently, the appropriate meaning of the concept *river* in *French* would be *fleuve.*
- facilitate the conversion of concepts semantics. That is, navigating through different levels of dimensions, we can change the semantic characteristics of each concept and define the impact of that change on the interpretation of this concept.
- analyze phenomena changes by facilitating the comparison of different semantic elements of the same phenomenon. In fact, *SemEL* helps to rapidly navigate through different dimensions and compare different measures of a given

phenomenon (i.e. *Ont_Desc* and *Context_Desc*) and infer what changes have affected that phenomenon. For example, if an assumption, specifying that "people can easily walk", was added to the semantic of the concept *forest*, we can conclude that the forest has been managed for hiking.

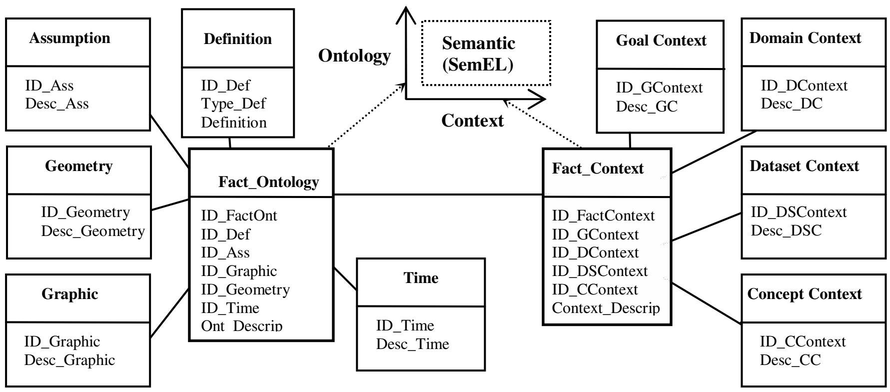

Fig. 2. A representation of the Semantic Element concept (*SemEL*)

4 Conclusion

In this paper we described the interoperability of geospatial datacubes, and we proposed a conceptual framework to overcome semantic heterogeneity problems when using geospatial datacubes. The framework is based on human communication, ontology, context, and the multidimensional structure. We defined a communication model which is based on agents representing geospatial datacubes (called *Datacubes Agents*) and a mediator agent (called *Context Agent*). The *Context Agent* helps the datacubes agents to appropriately interpret information exchanged between them. The interpretation is supported by a model which is based on the multidimensional paradigm (i.e. *SemEL*). The semantic of concepts will be discussed regarding the dimensions of this model (i.e. elements of both ontology and context).

Further work is required to refine *SemEL* and define a mapping between two different multidimensional models. Then we would implement our framework.

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

References

1. Answers, http://www.answers.com/ (Last visted September 20, 2006)
2. Bédard, Y., Rivest, S., Proulx, M.J.: Spatial on-line analytical processing (SOLAP): concepts, architectures and solutions from a geomatics engineering perspective. In: Koncillia, W.R (ed.) Data warehouses and OLAP: concepts, architectures and solutions (2005)

3. Bédard, Y., Bernier, E.: Supporting Multiple Representations with Spatial View Management and the Concept of "VUEL". In: Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, ICA Com. on Map Generalization (2002)
4. Bédard, Y., Merrett, T., Han, J.: Fundamentals of spatial data warehousing for geographic knowledge discovery. In: Miller, H.J., Han, J. (eds.) Geographic Data Mining and Knowledge Discovery (2001)
5. Bishr, Y.: Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geographical Information science 12, 299–314 (1998)
6. Brodeur, J.: Interopérabilité des données géospatiales: élaboration du concept de proximité géosémantique. Ph.D. Dissertation. Université Laval (2004)
7. Brodeur, J., Bédard, Y.: Geosemantic Proximity, a Component of geospatial Data Interoperability. In: Internat. Workshop, Semantics of Enterprise Integration, ACM Conference on OOPSLA, pp. 14–18 (2001)
8. Brodie, M.L.: The promise of distributed computing and the challenge of legacy lnformation systems. In: Proceedings of the IFIP WG2'6 Database Semantics Conference on Interoperable Database Systems (1992)
9. Codd, E.F., Codd, S.B., Salley, C.T.: Providing OLAP (On-Line Analytical Processing) to User- Analysts: An IT Mandate. Hyperion white papers, p. 20 (1993)
10. Goodchild, M.F., Egenhofer, M.J., Fegeas, R.: Interoperating GIS. Report of a specialist meeting held under the auspices of the Varenius project panel on computational implementations of geographic concepts. Santa Barbara, California (December 5-6, 1997)
11. Guarino, N.: Formal Ontologies and Information Systems. In: Guarino, N. (ed.) Proe. of FOIS 1998, pp. 3–15. IOS Press, Amsterdam (1998)
12. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Original paper presented at the International Workshop on Formal Ontology (1993)
13. Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C.: Semantic Interoperability: A Central Issue for Sharing Geographic Information Annals of Regional Science. Special Issue on Geo-spatial Data Sharing and Standardization, pp. 213–232 (1999)
14. ISO/IEC 2382. Information technology – Vocabulary – Part 1: Fundamental terms (1993)
15. Malinowski, E., Zimanyi, E.: Representing spatiality in a conceptual multidimensional model. In: Proceedings of ACM int. workshop on Geographic information systems, ACM Press, New York (2004)
16. Rivest, S., Bédard, Y., Proulx, M., Nadeau, M., Hubert, F., Pastor, J.: SOLAP Technology: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data. J. ISPRS Advances in spatio-temporal analysis and representation, 17–33 (2005)
17. Roehrig, J.: Information Interoperability for River Basin Management. J. Technology Resource Management and Development - Scientific Contributions for Sustainable Development 2 (2002)
18. Xhu, Z., Lee, Y.C.: Semantic Heterogeneity of Geodata. In: Proceedings of ISPRS Commission IV Symposium (2002)