Modern Education
and Computer Science
PRESS

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business Intelligence Knowledge

Belko Abdoul Aziz Diallo
West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL),
Competence Centre, Ouagadougou, Burkina Faso
Email: diallo.b@wascal.org

Thierry Badard
Centre de Recherche en Géomatique, Université Laval, Québec, Canada G1V 0A6
Email: thierry.badard@ulaval.ca

Frédéric Hubert
Centre de Recherche en Géomatique, Université Laval, Québec, Canada G1V 0A6
Email: frederic.hubert@ulaval.ca

Sylvie Daniel
Centre de Recherche en Géomatique, Université Laval, Québec, Canada G1V 0A6
Email: sylvie.daniel@ulaval.ca

Abstract—To contribute in filling up the semantic gap in data warehouses and OLAP data cubes, and enable semantic exploration and reasoning on them, this paper highlights the need for semantically augmenting Geo/BI data with convenient semantic relations, and provides OWL-based ontologies (ODW and OOLAP) which are capable of replicating data warehouses (respectively OLAP data cubes) in the form semantic data with respect of Geo/BI data structures, and which enable the possibility of augmenting these semantic BI data with semantic relations. Moreover, the paper demonstrates how ODW and OOLAP ontologies can be combined to current Geo/BI data structures to deliver either pure semantic Geo/BI data or mixed semantically interrelated Geo/BI data to business professionals.

Index Terms—Business Intelligence, Geospatial Business Intelligence, semantic Geo/BI, OLAP, Data warehouse, metadata, semantic gap, semantic relations, data semantics, OWL Ontology, semantic layer, data analysis, knowledge discovery, Decision support

I. INTRODUCTION

Business Intelligence (BI) technologies are among decision support systems (DSS) which are widely and increasingly adopted by companies [1] and global revenue in the business intelligence (BI) and analytics software market was forecast to reach $18.3 billion in 2017, and $22.8 billion by the end of 2020, according to the latest forecast from Gartner, Inc. [2]. Thanks to their multidimensional and multilevel data structures, data warehouses and OLAP data cubes provide (i) an effective way of quickly crossing data, (ii) a straightforward means of data aggregation, and (iii) a quick calculation of data, allowing then an intuitive analysis and exploration of data.

However, despite all these capabilities, BI and derived Geospatial BI (Geo/BI) data structures do not provide answers to all concerns regarding business analysis. A good illustration of this, is their lack of semantic information.

2   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

The semantic gap within BI data (data warehouses, OLAP cubes) as stated by [3] is well known among BI practitioners and researchers and several solutions have been proposed to overcome or minimize that gap. All these have generally been based on the principle of providing more semantics to metadata to get enough description of BI concepts and properties.

But semantics is also about data itself and relations that may exist between the data occurrences. For instance, the fact of knowing that the "company A" that purchased from us our company $ 500,000 of products last year is competing with "company B" that provides us with these products, and is in partnership with the "company C" which ensures our deliveries to customers, provides more valuable insights to a decision maker. Such a knowledge is explicitly absent in Geo/BI data and is generally unearthed by the analyst/decision maker after an additional effort of information search from other data sources: e.g. browsing the web, calling the customer service, etc. Moreover, this knowledge, once established, is very poorly exploited since it is not saved in a formal and structured manner. It usually remains buried in the memory of the decision maker having deducted it or is mentioned in unstructured documents (oral exchanges, notes, reports, etc.).

Taking into account such semantic relations between Geo/BI data can, not only enrich the data, but also provide decision makers with semantic-oriented analysis, exploration, and discovery of the data and knowledge. As an illustration, a salesperson might want to know the part of sales realized with client companies competing between them, and the part realized with partnering client companies, targeted to his current location (e.g. Beauport district in Quebec City).

Regrettably, nowadays, Geo/BI data structures do not provide such a semantic support. And to the best of our knowledge, there is not yet a work regarding semantic enrichment of Geo/BI data with semantic relations between data occurrences.

The present paper addresses this problem as follows. After reviewing major proposals on semantic enrichment of BI (section 2), the paper through a realistic case study justifies why there is still a need to semantically enrich BI data, this time, with semantic relations (section 3). Then the paper exposes its approach towards enabling full semantic Geo/BI solutions: to overcome the lack of semantics in data warehouses (respectively OLAP data cubes) and enable semantic exploration and reasoning on them, the authors have designed OWL-based ontologies (O.DataW and OOLAP) which are capable of replicating data warehouses (respectively OLAP data cubes) in the form semantic data with respect of the data structure, and which provides the possibility of augmenting these BI semantic data with semantic relations (Section 4). Finally, section 5 demonstrates how O.DataW and OOLAP ontologies can be combined to current Geo/BI data structures to deliver either pure semantic Geo/BI data or mixed semantically interrelated Geo/BI data.

II. Related Work on Semantic Enrichment of BI Data

BI data is usually loaded into a logical multidimensional data model (e.g. Fig. 5) and physically stored in a huge database called data warehouse. The logical data model is supposed to hold and organize data in the same way as expressed in the conceptual data model it derives from.

Unfortunately, as reported by [3] and recalled by [4], there still is a semantic gap between advanced conceptual data models and relational or multidimensional implementations of data cube [5]. Additionally, it appears to be an open problem how to represent dimension constraints [6] or even less expressive context dependencies [7], both of which explain the existence of null values in dimensions in logical implementations and allow to reason about summarizability with respect to sets of attributes.

To overcome or minimize the semantic gap within BI data, several authors have proposed different solutions ranging from creating semantic bridges and enriching business/semantic metadata, to annotating BI data cubes with ontological models of OLAP cubes. Here are some major works regarding these various proposals.

Semantic bridges: to fill the gap between conceptual and logical models, [8] proposes the construction of a semantic bridge between the two models by using a model-driven architecture (MDA) to translate semantics from the conceptual level into OLAP logical system. An OLAP algebra is built by using OCL to express needs and semantics at the conceptual level. This algebra is then transformed into a logical schema (e.g. SQL) by using QVT language. [9] also used MDA [10] method, OCL [11] and QVT [12] languages to build a semantic derivation from conceptual geospatial data warehouse specifications into their suitable logical models.

Fig.1. Enriched business metadata connected to fact data from [13]

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL 3
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

Enriched semantic/business metadata: To help OLAP users in establishing a link between OLAP metrics values and business goals they have to reach, [13] proposed to enrich business metadata with a UML-based meta-model which defines details regarding enterprise goals (e.g. Goal name, Goal perspective, Metric name, Metric target value + unit, etc.). That model of goals is then linked to the data warehouse containing the BI data, by using the technique of model weaving, which consists of establishing links describing the relationships between the goals model and the data warehouse model. This linkage is then used to display business metadata (e.g. business goals) related to OLAP fact data (e.g metrics values) such as illustrated in Fig. 1 provided by the authors. The same technique is used by [14] to *"integrate Goals with Process Warehouse for Business Process Analysis"*.

Ontology-based semantic annotation: semantic annotation is another method proposed by authors to fill in the semantic gap within BI data. [15] for example proposed to enrich OLAP data cubes by annotating them with ontological descriptions. These annotations are then exploited to display the semantics attached to a dimension or a measure like for instance, how it is aggregated or calculated. Fig. 2 shows an example of semantic annotation regarding the calculation formula of the measure ROI (Return On Investment).
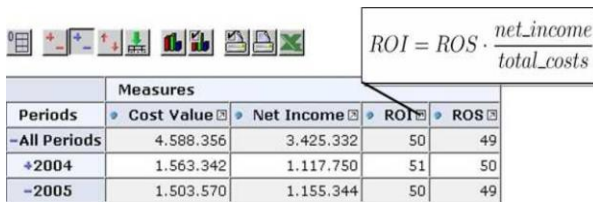

Fig.2. Example of semantic annotation from [15]

[16] also adopted the semantic annotation approach to *"facilitate the exchange of business calculation definitions"* between users and organizations and to *"allow their automatic linking to specific data warehouses through semantic reasoning"*.

Ontology-based ETL and OLAP: [17] and [18] propose the use of ontologies to conduct data extraction from their sources, and data integration into data warehouses and OLAP cubes. For this purpose, the authors define an OLAP ontology (Fig. 3) which describes the formal OLAP cube structure (e.g. dimensions, measures, etc.). Then, data sources are located and converted in an RDF *"format that makes the semantics of the data explicit"*. For each data source, a mapping ontology is used to convert the data in a way that matches the OLAP cube ontology. Thereafter, the OLAP ontology and the RDF data are used to construct the OLAP cube. Fig. 3 shows a graphical version of an ontological OLAP cube model proposed by the authors in [19]. [20] also provided an OLAP ontology (Fig. 4) to help integrate distributed energy sensor data and compose new data cubes from existing ones by alleviating schemata inconsistency such as "attribute differences, missing data, or semantic and functional gaps".
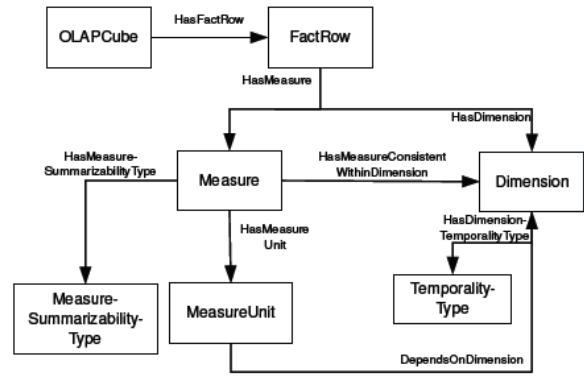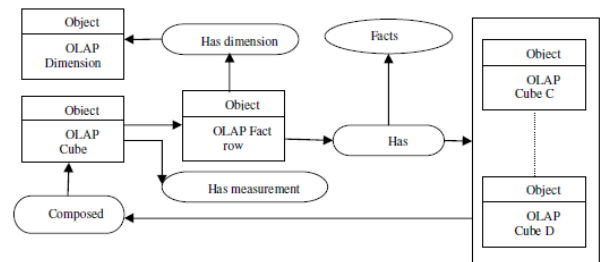

Fig.3. OLAP cube ontology proposed by [19]

Fig.4. OLAP ontology proposed by [20]

As it can be noticed, existing solutions mainly focus on semantically enriching BI metadata (e.g. concepts/classes, attributes/properties) rather than BI data itself (i.e. occurrences of concepts/classes or values of attributes/properties).

But semantics is also about relations existing between data occurrences. And to the best of our knowledge, there is not yet a work regarding semantic enrichment of BI data by considering semantic relations which may exist between the data.

III. Why Semantically Augment Geo/BI data

To highlight us how semantic-augmented Geo/BI data could enhance business analysis, let consider a realistic case study of a business professional named IdoBI Reason.

A. Case study: BioWYNX sales activities

M. Reason is a sales analyst and strategist moved from Washington DC to Quebec City to reorganize and expand the local branch of BioWYNX.

BioWYNX is a multinational firm specialized in selling biological food products. To minimize delivery fees, BioWYNX disposes of at least one storehouse per district from which salespersons can supply customers with desired products.

BioWYNX has its own salespersons but also deals with other mobile salespersons working for partnering companies and in accordance with these companies. So, these shared salesmen, when selling their own company products, can also sell BioWYNX products, and are rewarded by BioWYNX (the companies as well) according to sales they realized. A salesperson can

4   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

supervise other salespersons (e.g. a team) or a given company.

To monitor efficiently its business performance, BioWYNX has deployed a BI platform. Fig. 5 represents the snowflake schema of the data warehouse from which OLAP cubes and mini-cubes (Fig. 6) are built. The dimensions are Products, Seller, Time, Location, and Customer which has two hierarchies. The measures are number of sold products (NbProdUnits), sales amount of a given product (SalesAmount), average of offered discounts (Discount), and the unit price (avgUnitPrice).
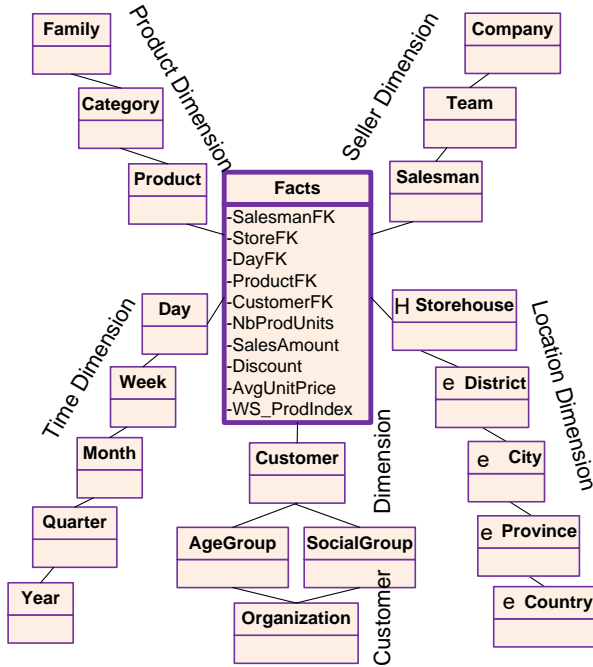

Fig.5. Snowflake-schema model for warehousing sales data

That general context of BioWYNX business activities will be used throughout this paper to highlight the lack of semantics in current BI data, and the need and relevance for semantic relations to provide semantic-based analysis, exploration, and discovery of BI data and correlations within data.

B. Lack of semantics between BI data

Let us consider say that IdoBI, after replacing the former salesperson in chief, is meeting on the field, some key salesmen (e.g. Jack, Jim, John). While discussing with Jack, M. Reason is also exploring (e.g. drill down, roll up, etc.) from his smartphone, a SOLAP mini-cube related to sales performed by salesmen (including Jack) over the last five years. And from time to time, to argue what he is saying or proposing to Jack, M. Reason shows him the analysis data related to his sales.

After exploring and discussing Jack's sales over the last five years thanks to SOLAP mini cubes, M. Reason would like now to explore sales related to salesmen whose Jack is the supervisor (e.g. Jim), and the sales performed by the supervisors of Jack (e.g. John) in a way like:

In my current location (e.g. Beauport district in Quebec City), who are the supervisors and supervisees of Jack that cumulated more than $100.000 sales of chocolate family products last month, and that have their offices near to my hotel or near my current position?

Such a semantic-oriented Geo/BI request ("is a supervisor of" defines a semantic relation between salespersons in seller dimension) brings an interesting new way of analyzing BI data and may ease and speed up the discovery of correlations between data. Indeed, through that request, M. Reason is trying to discover a correlation between salespersons performance and their supervisors/supervisees performance.
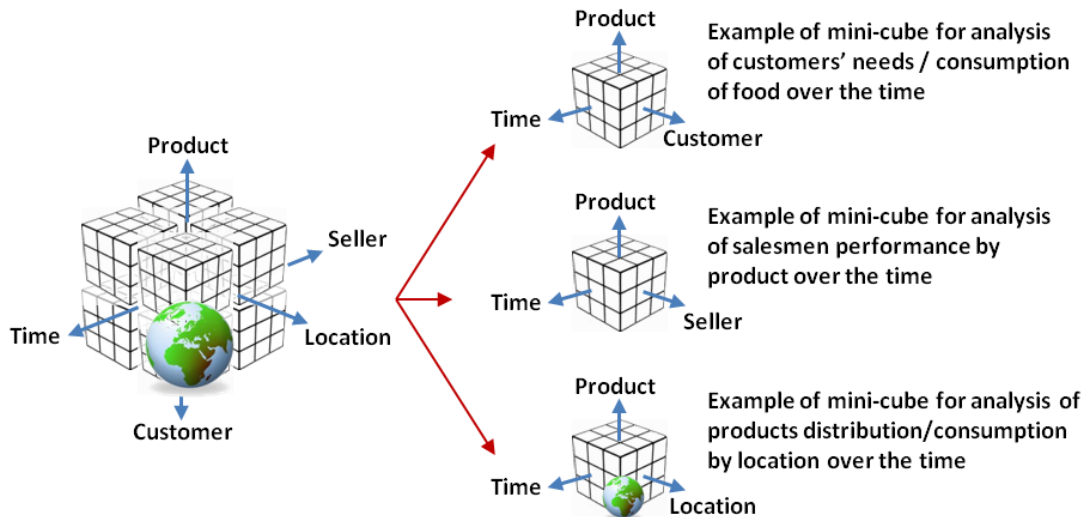

Fig.6. Example of OLAP hypercube and mini-cubes which can be generated from the previous data warehouse model (Fig. 5) thanks to server-side OLAP tools

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL   **5**
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

Given the current capabilities of OLAP/SOLAP cubes, the only way for M. Reason to explore and analyze data regarding sales performed by the supervisors and supervisees of Jack is to proceed as follows:

(i) Firstly, identify the names of salespersons Jack is supervising or that are supervisors of Jack. Such information is not available in OLAP cubes nor in data warehouses. Thereby, given that M. IdoBI does not yet know by heart all his employees and their organizational hierarchy, he might have to interrupt his discussion with Jack to look for that information by calling his secretary, or remotely accessing the employees' file, etc.

(ii) **A.** Thereafter, manually browse all members of the dimension level "Salesman" in the dimension "Seller" (cf. Fig. 5) of the SOLAP mini-cube until he finds a name among the names of Jack's supervisors (John, etc.) or supervisees (Jim, etc.).

B. Or write a BI data request using the dedicated MDX (Multidimensional Expressions) language to select sales regarding the list of salespersons previously identified as being supervisors or supervisees of Jack.

Such additional search and request tasks might be time consuming and inappropriate for competitive decision making and will not even allow M. Reason to navigate directly from Jack's sales figures to John's (supervisor) or Jim's (supervisee) sales figures and vice-versa.

The today's difficulty to make decision makers benefit of such semantic-oriented BI requests is due to the lack of semantics (especially semantic relations) in OLAP cubes as well as in data warehouses from which cubes are built. Indeed, there is no information attached to the S/OLAP cube that indicates for example that Jack is supervised by John and supervises Jim (In the data warehouse model in Fig. 5, there is no relation indicating that a salesperson may have a link with another one).

C. Need for semantic exploration, analysis, and discovery of BI data and correlations within data

Fig. 7 visualizes the situation aforementioned, points out the lack of semantics in S/OLAP cubes, and highlights an example of how semantic relations, if they were present, would have been taken advantage to provide business professionals with a semantic support which can offer an advanced and meaningful exploration, analysis and discovery of BI data. Different examples of semantic relations are illustrated (e.g. "is trainer of", "is supervisor of", "is friend of") to bring a wider view of semantic relations richness. Red crosses express the absence of these relations in today's S/OLAP technologies.

The lack of semantic relations between BI data does not concern only data within the same level of dimension as depicted in Fig. 7. This also concerns the members of different levels in the same dimension (e.g. to which organization the customer "M. Ido Buy" belongs to? in order to access directly sales made with that organization from sales made with "M. Ido Buy"), as well as data in different dimensions (e.g. Are there some client organizations which are in competition with selling companies?).

Indeed, let consider that M. IdoBI Reason is now performing a drill down operation from Team level (in Seller dimension) to Salesman level in order to explore in detail, sales performed by each salesman of a given team (e.g. BioTeamX1). The list of salespersons (e.g. John, Jack, Jim) is clearly known as "belonging" to the selected team (i.e. BioTeamX1). Now, if he wants to get back to the team's sales figures (i.e. BioTeamX1) from one of its salespersons (e.g. Jack), by applying the inverse operation (i.e. roll-up), M. IdoBI will get the list of all teams (e.g. BioTeamX2…BioTeamX10) rather than the desired team. He will not then be able to identify and focus only on the team of Jack since he is not meant to memorize all the employees' names and their related
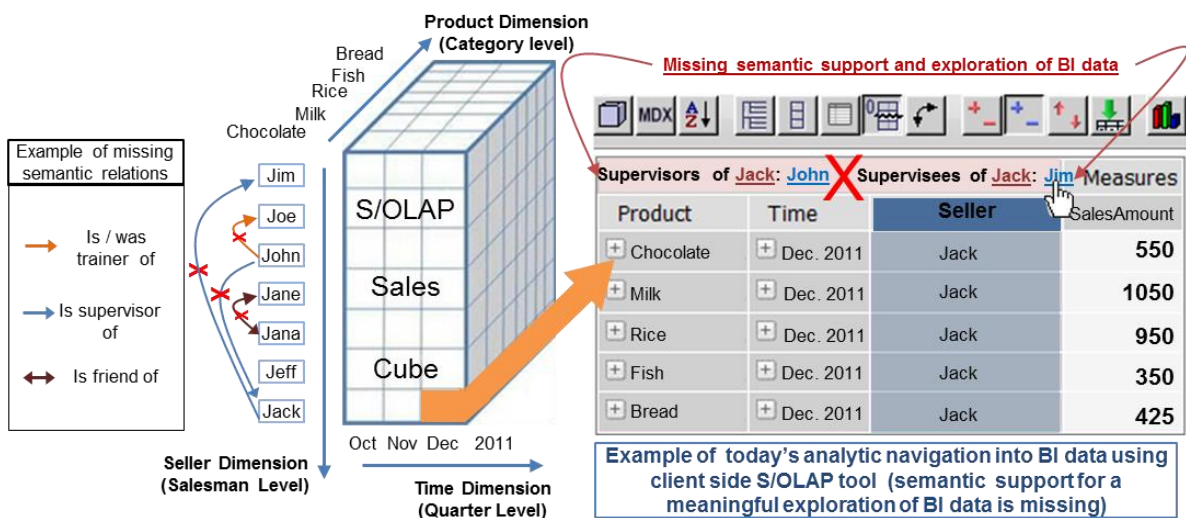

Fig.7. Lack of semantic support for a meaningful exploration, analysis and discover of BI data in today's S/OLAP technologies

6   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
          Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
                                    Intelligence Knowledge

teams. This means that exploring, analyzing and discovering data from Parent-Level to Child-level is possible while getting back from a child (in a Child-level) to its exact parent (in a Parent-level) is not offered by

today's BI technologies. Putting a semantic relation between children and parents (e.g. belongs to) can overcome that issue.
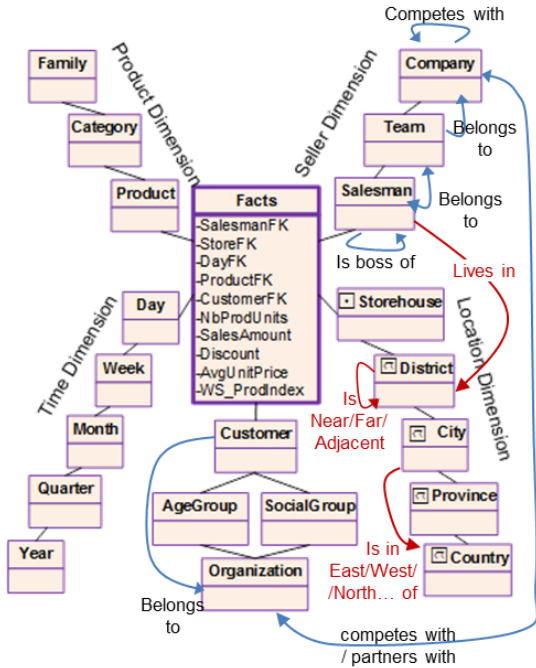
A- Overview of relevant semantic relations which might exist within a data warehouse
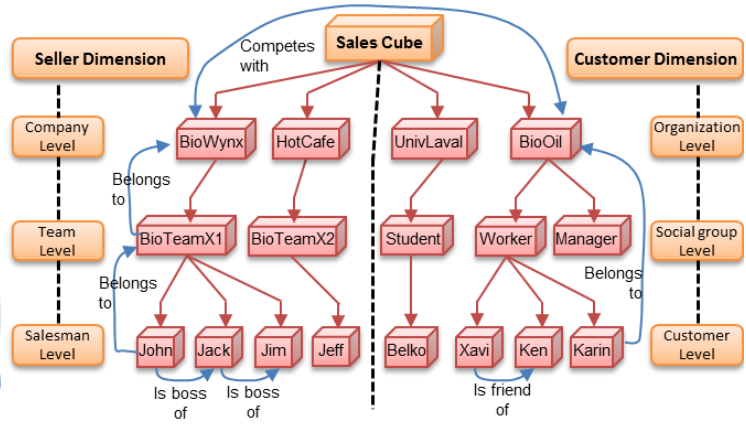
B- Excerpt of relevant semantic relations which might exist within a S/OLAP Cube

Fig.8. Example of semantic relations that might exist between data within the same level, between levels of the same dimension, and between dimensions

Fig. 8 provides various examples of semantic relations that might exist between data within the same level, between levels of the same dimension, and between dimensions. For instance, a selling company may be in competition with another one or be a partner of a customer organization. It also underlines the relevance of the "Belongs to" relation for providing a direct roll up to the expected member instead of getting all members of the upper level.

Semantic spatial relations might also exist between the location dimension and other dimensions members, or within the location dimension members. Some examples of these spatial relations are emphasized in dark red in Fig. 8-A. For instance, a district may be near to/far from/adjacent to another district; a given city may be situated in the east, west, north or the south of a given country, etc. If Geo/BI systems usually provide spatial analysis capabilities that can easily compute topological relations (ex. adjacency) between geospatial objects, non-geospatial BI systems do not, and both BI and Geo/BI systems do not take into account semantic relative positions such as:

(i)   Relative distances: near, far, in front of, behind, etc.
(ii) Relative levels: above, below, etc.

(iii) Relative orientations: left, right, east, west, north, south, north-east, south-west, etc.

Given this semantic deficiency of BI data highlighted throughout this case study, let us review the various solutions proposed in the literature to semantically enrich data warehouses and data cubes.

IV. DESIGN APPROACH FOR SEMANTICALLY ENRICHING GEO/BI DATA

Semantic relations already exist in OLTP databases, which are often used as data sources for data warehousing BI data. For instance, in Human Resources Management systems, one can know which employees supervise the others thanks to various joins connecting tables. By contrast, for the purpose of speeding up calculations and queries, and providing a quick, efficient and simplified access to analysis-oriented data, Geo/BI data structures are built to reduce as much as possible, these joins. This means that additional relations (joins) except those linking fact tables to dimensions tables are to avoid, or even not desired. Therefore, any solution aiming to add and establish semantic relations within BI data should be, somehow, external to Geo/BI data structures.

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL   **7**
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
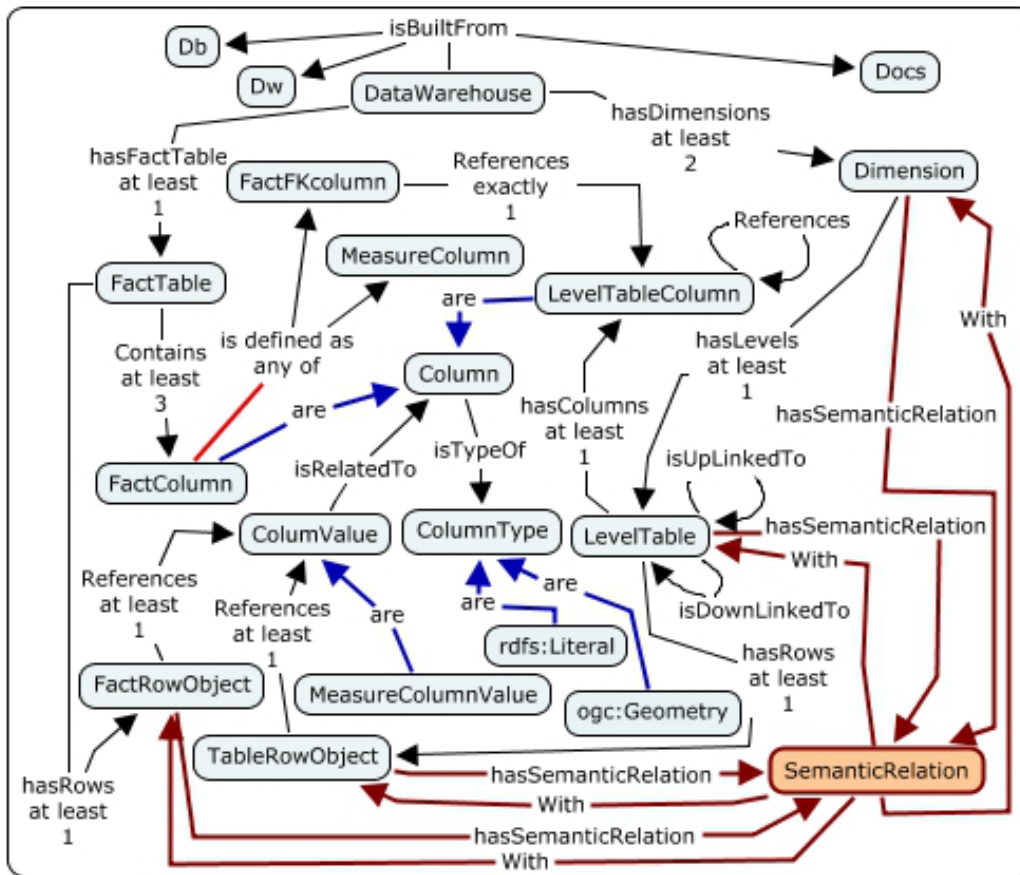Intelligence Knowledge

Fig.9. ODW ontology for replicating and semantically augmenting OLAP data warehouses

To enrich BI data with semantic information, we suggest the use of OWL-based semantic layers to describe not only the data stored in Geo/BI structures (data warehouses and cubes) but also the semantic relations that may connect them.

Hence, the only way that remains to deliver semantic-augmented Geo/BI services to business professionals is to have, alongside Geo/BI data structures, other data structures that store semantic information regarding BI data and relations, and that reference the data stored in Geo/BI data structures. These external data structures will hence, act as semantic layers for common Geo/BI data structures.

Various data modelling languages like ER (Entity/relation), UML (Unified Modeling Language) or OWL (Web Ontology Language) could be used to design these layers. But Since reflexive (e.g. is boss of), symmetric (e.g. competes with), transitive (e.g. belongs to), opposite (e.g. competing vs. partnering), and other rule-based relations might be involved in describing semantic relations between data, the use of OWL-based ontology is more suitable (for semantic inference) than ER (Entity/relation) and UML [21] to design these semantic layers.

To conveniently reference BI data stored in Geo/BI data structures and establish semantic relations between these data, the semantic layers should have to reproduce, in one way or another, the multidimensional structure of Geo/BI data structures with explicit reference to dimensions, levels, measures, etc. In addition, they should also reference each data occurrence (data rows in data warehouses, members in S/OLAP cubes), since each data occurrence can potentially have a semantic relation with another one (e.g. Member "John" is the supervisor of member "Jim").

Referencing Geo/BI data stored in Geo/BI data structures can be made in various ways such as refereeing to members' positions in a dimension level, cells coordinates in an S/OLAP cube, etc.

But, to benefit of the richness and inferential potential of OWL-based ontologies, we propose, not only to reproduce Geo/BI structures, but also to replicate Geo/BI data into OWL-based semantic layers, and then augment these replicated data with semantic relations that can connect them when needed.

Simply put, the ontological semantic layers to be designed for Geo/BI data structures have to replicate the structure and the data of deployed data warehouses (semantic DW ontology) and OLAP cubes (semantic cube ontology), and provide the capacity to add semantic information that allows, not only to describe the data and metadata (semantic annotations) as do existing solutions, but also to interrelate data occurrences regarding relationships which may exist between them (semantic relations).

These requirements have led us to the design of ODW and OOLAP, two OWL-based ontologies proposed in sections 5 and 6 to semantically augment data

8   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

warehouses and OLAP data cubes, respectively. CmpaTools COE graphical editor and syntax [22] have been used to graphically design them.

V. ODW: An Ontology to Semantically Augment Data Warehouses

To enrich data warehouses with semantic relations establishing connections between data occurrences, ODW ontology, an OWL-based ontology for data warehouses is proposed. This semantic layer is dedicated to map and replicate the structure and the data of a data warehouse. Afterward, semantic relations can be established between these semantic data replicated from the data warehouse.

Fig. 9 presents a graphical upper-level model of a Geo/BI data warehouse semantic layer ontology with augmented semantic relations. The complete ODW ontology file is available for download[1].

The model without the SemanticRelation concept and its connections describes the multidimensional structure of common data warehouses. A data warehouse is considered as having at least one fact table and at least two dimensions. A fact table contains at least three columns (factColum), which is defined as a union of (at least one) column(s) regarding measures (measureColumn) and (at least two) columns regarding foreign key columns (FKColumns) that connect the fact table to involved dimensions. A dimension has at least one table designated by LevelTable. When there are several levels in a dimension, level tables are ordered using relations "isUpLinkedTo" and "isDownLinkedTo". The most granular level has no down-link while the most generic one does not have an up-link. In this way, hierarchies in the dimensions are also implicitly managed (These are only explicitly named in OLAP cubes not in data warehouses schemata). A level table has at least one column (LevelTableColumn) which can be referenced by a fact table's foreign key column, FactFKColumn (factTable↔ DimensionTable linkage) or by another LevelTableColumn (LevelTable↔ relationship via PrimaryKey↔ ForeignKey columns). FactColumns as well as LevelTableColumns are Columns (inheritance of the entity Column) which have ColumnTypes (rfds:Literal i.e. standard types, or ogc:Geometry for geometries). This describes the metadata of a common data warehouse structure.

When the model is populated with data, each instance of columnValue is related to the concerned Column; and each FactRowObject and TableRowObject reference at least one column value. The model is augmented with semantic relations by linking between them, dimensions, LevelTables, TableRowObjects and FactRowObjects thanks to the SemanticRelation class, which allows creating semantic relations objects that can be paired to others data instances through the links named "hasSemanticRelation" and "With". For instance, to specify that companies A and B are competitors, the

concerned rows (instances of TableRowObjects) describing these companies in the level table Company (instance of LevelTables), are to be linked by creating the relation "competes with". This is done by instantiating the class SemanticRelation with the value "competes with" as semantic relation object and by using the link "hasSemanticRelation" to relate Company A with the semantic relation object, and the link "With" to pair this latter object to Company B.

This upper-level ontology is suitable for generating and feeding a domain-specific data warehouse semantic layer from a deployed and operating data warehouse (e.g. sales data warehouse in Fig. 5) and without knowing in advance fact tables, dimension tables and data occurrences composing it. In order to conveniently generate relations linking data occurrences between them, the generation of that domain-specific semantic layer would be done during the ETL process at the same time the data warehouse is loaded. Indeed, since major semantic relations connecting data occurrences between them (e.g. John is boss of Jack) are usually present in OLTP databases (e.g. CRM, ERP databases, etc.), ETL tools might help extract these relations and generate them in the semantic layer (Fig. 9) when extracting data to warehouse.

Nevertheless, semantic Geo/BI designers can also choose to do not use the upper-level ontology model, but directly design a domain-specific data warehouse ontology as a semantic layer and accordingly populate it with data and semantic relations relating data occurrences. For instance, the structure of a domain-specific ontology for the data warehouse schema presented in Fig. 5 might look like the one depicted in Fig. 8-A. All depends on the degree of automation needed during the process of generating and populating the data warehouse semantic layer. The upper-level model ontology is generic, then more easily automatable without explicitly knowing the structure of underlying data warehouse schema. Whereas, a domain-specific semantic layer ontology is more explicit to the designer and easier to manipulate, but requires additional design effort. For instance, if there are ten different data warehouses to semantically enrich with semantic relations, ten domain-specific semantic layers will have to be manually designed, rather than generated them from a generic model.

VI. OOLAP: An Ontology To Semantically Augment OLAP Data Cubes

As for data warehouses, an OWL-based ontology for OLAP cubes named OOLAP is proposed to replicate the structure and the data of the implanted data cube. Semantic relations can then be added to that semantic BI data. The ontology is depicted in a graphical format in Fig. 10. It presents a graphical upper-level ontology for an S/OLAP cube which we propose to replicate common cubes and then enrich the replicated data with semantic information regarding the data and semantic relations which may link them. The proposed model is partially inspired by the UML-based OLAP Cube metamodel

[1]:https://www.dropbox.com/s/njugakwaqhosdai/ODW.owl?dl=0

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL 9
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
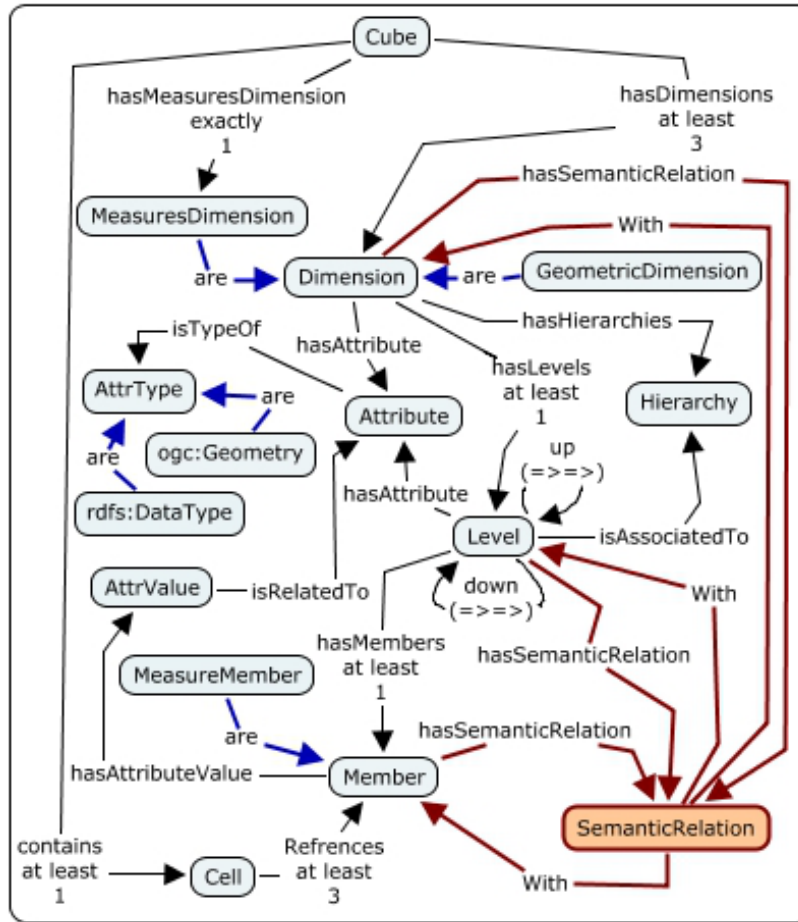Intelligence Knowledge

Fig.10. OOLAP ontology for replicating and semantically augmenting OLAP cubes

proposed by [23]. Additional considerations regarding major rules specified by the OMG group in *"Common Warehouse Metamodel (CWM) specification"* [24] have been integrated. The complete OOLAP ontology file is also available for download[2].

The S/OLAP cube ontology model, without the SemanticRelation concept and its connections, describes the multidimensional structure of any S/OLAP cube. We consider an S/OLAP cube as having at least dimensions among which exactly one dimension dedicated to measures (MeasuresDimension). MeasuresDimension is a special dimension containing all measures involved in the cube. Indeed as stated in CWM specification [24], *"the OLAP meta-model defines two special types of Dimension: Time and Measure"*. This is to provide a "complete symmetry between dimensions" and handle measures in the same way other dimensions are handled when querying data. In addition, several OLAP technologies such as Microsoft SQL Server Analysis Services, Mondrian, GeoMondrian and MDX language manage OLAP cubes in that way, i.e. measures are stored in a dedicated dimension [23]. Besides the measures dimension, the proposed semantic layer also distinguishes geometric dimensions apart in order to manage and process geometries conveniently.

Unlike models which consider that a dimension contains at least one hierarchy that contains in turn levels, the proposed model is built in respect of CWM specification that states: "*A Dimension has zero or more Hierarchies. A hierarchy is an organizational structure that describes a traversal pattern through a Dimension*" [24]. Therefore, we consider that a dimension has one or more levels which can be associated to dimension hierarchies (concept of Hierarchy) if any.

Levels are organized in a bottom-up manner so that the lowest level is the most granular and the upmost one is the most generic one. The transitivity rule (=>=>) characterizing relations "up" and "down" are necessary to determine in which order levels associated to a given hierarchy have to be ordered, mostly when a hierarchy associates levels which are not directly uplinked and downlinked together. For example, in a time dimension, we can have the following levels conveniently linked: Day, Week, Month, Quarter, and Year. All of these levels may be associated with a first hierarchy named "Time refined View" whereas only Day, Month and Year might be involved in another hierarchy called "Time Quick View". Since in the connection of levels, Day, Month and Year are not directly connected, nothing indicates the system which is up or down. But by making transitive the "up' and "down" semantic relations, the system can now infer and order the levels in the right way: Day-Month-Year. Fig. 11 graphically illustrates how these transitive

---

10   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

semantic relations designed in ODW ontology can be inferred to determine indirect links between levels within hierarchies of dimensions.
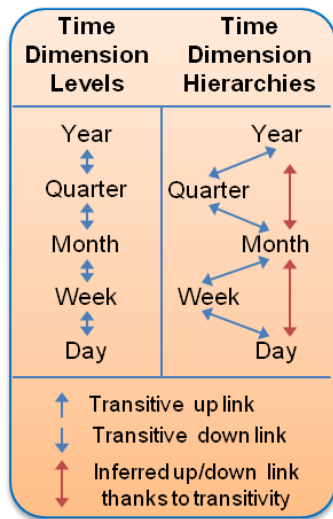

Fig.11. Up / Down linking levels in hierarchies using transitive semantic relations

Furthermore, the designed ontology also allows the annotation of entities like dimensions as well as levels with dynamic attributes/properties (ProductColor, SalesmanAge) thanks to the natural annotation power of OWL (Pottle, Lancaster, & Greenberg, 2008).

Therefore, the foregoing OOLAP semantic layer describes a kind of ontological metadata of S/ OLAP cube common structure.

When this semantic cube model is populated with data, each level is provided with members that can be members of common dimensions or members of measures dimension. These members can be crossed and referenced as belonging to a given cell of the cube. Members of levels having attributes are associated with their different values. The model is augmented with semantic relations by linking between them, dimensions, level, and members thanks to the SemanticRelation class and the links "hasSemanticRelation" and "With".

This upper-level ontology for an S/OLAP cube semantic layer is suitable for generating and populating a domain-specific semantic layer from a deployed and operating S/OLAP cube (e.g. sales SOLAP hypercube in Fig. 6) without knowing in advance, the dimensions, levels, hierarchies, and members composing it.

As it can be noticed, the ontological semantic layers can store both the BI data and the semantic relations linking them and empowering their semantics for a semantic-oriented analysis and exploration. Moreover, additional comments/annotations describing the BI data and metadata and their semantic relations can be added at any time (at runtime by users, or at design time by data administrators) through labels or other OWL data properties, thanks to the capacity of OWL ontologies to provide dynamic insertions and updates of semantic data.

VII. Strategies for Delivering Semantically Interrelated Geo/BI Data

Since Geo/BI data structures only store the BI data, while the supplied semantic layers store both the BI data and the semantic relations augmenting the data, two strategies are possible to extract and deliver decision makers with BI data semantically interconnected:

- The first strategy consists of retrieving both the requested BI data and their semantic relations directly and exclusively from the data cube semantic layer (unless very detailed data are required from the data warehouse) by running semantic requests against this latter. Let us designate it as a purely semantic-oriented strategy. It is straightforward and exclusively based on semantic layers once these are created and populated. It consists of extracting directly from the semantic layers, both the BI data requested by the decision maker and the possible relations connecting them.

- The second strategy consists of retrieving the BI data as usual from the Geo/BI data structures (mainly the OLAP data cube, unless a drill-through operation requires an access to the data warehouse) by running MDX queries against the cube; and then, retrieving the semantic relations interconnecting these data from the data cube semantic layer, by issuing semantic queries. Let us name it as mixed semantic-augmented strategy.

- The purely semantic-oriented approach is straightforward but totally excludes the use of common Geo/BI technologies (e.g. OLAP cube, MDX) in the BI data retrieval. This involves that Geo/BI users wanting to write their own queries to extract data (as they usually do by using MDX) will have to learn a semantic query language such as SPARQL, or its geospatial and industry standardized extension, GeoSPARQL. To overcome this issue of query language from the user side, two variants of this strategy can be envisaged. The first one is providing the Geo/BI user with human-friendly primitives or functions they can easily write to retrieve desired data, such as in NoSQL systems. For instance, primitives such as selectMembers(dim, level, cond) or selectFacts (dimensionsList, measuresList, conditions) may be implemented to ease user-side query writing. The second variant consists of offering the possibility to end users to write MDX queries (as usual) that will be, afterward, handled and translated by the system into semantic Geo/SPARQL queries to be executed against the data cube semantic layer. This latter variant can be even enriched with the first one to provide a kind of a complete Not only MDX (NoMDX) query writing strategy.

The mixed semantic-augmented strategy is more complex to implement, but has the advantage of taking into the current practices regarding the BI data retrieval, both from user's side and the system side. Experienced users will keep using their MDX language to extract specific data they need, and organizations do not have to

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL   **11**
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

adopt another system to benefit semantic enrichment of BI data. But a kind of sophisticated plugin will have to be developed to integrate the semantic layers and their processing engines into current common GeoBI infrastructures. Indeed, S/OLAP technologies do not yet provide internal semantic engines for processing semantic data cubes. Therefore, semantic reasoning engines also have to be integrated.

Further work will propose a suitable architecture to integrate these semantic layers into current Geo/BI architectures and will implement a prototype of the proposed semantic Geo/BI solution.

VIII. CONCLUSION AND FUTURE WORK

After reviewing several research proposals (section 2) intended to overcome or minimize semantic gap within BI data, and after underlining that existing solutions mainly focus on semantically enriching BI metadata (e.g. concepts/classes, attributes/properties) rather than BI data itself (i.e. occurrences of concepts/classes or values of attributes/properties), this paper has shown, through a realistic case-study (section 3), the needs for enriching BI data with semantic relations to offer to the decision makers, the possibility to semantically analyze, explore and discover Geo/BI data and semantic correlations between them..

The paper has then explored (section 4) step by step, an approach for semantically enriching Geo/BI data with a suitable inferable semantic language (e.g. OWL) and by following a convenient method: semantically replicating both Geo/BI structures and data, and then providing the capacity to add semantic information that allows, not only to describe the data and metadata (semantic annotation) as do existing solutions, but also interrelate the data regarding relationships which may exist between them (semantic relations).

Afterward, thanks to the proposed method, two OWL semantic layers has been designed to enrich Geo/BI data and overcome the formerly identified semantic lack. These OWL-based ontologies for data warehouses (ODW) and OLAP cubes (OOLAP) allow to map and replicate the data structure and the data occurrences and add semantic relations to data occurrences. Moreover, additional comments/annotations describing the Geo/BI data and metadata and their semantic relations can be added at any time (at runtime by users, or at design time by data administrators) through labels or other OWL data properties, thanks to the capacity of OWL ontologies to provide dynamic insertions and updates of semantic data.

At last, the paper has exposed two strategies on how these semantic layers could be used in combination of common Geo/BI data structures, to extract and deliver semantically interrelated Geo/BI data to decision makers.

Further work is being conducted to provide an operational and technical solution for integrating the proposed semantic layers into current Geo/BI infrastructures. Future work will be conducted to implement a prototype of semantically-augmented Geo/BI application in order to test its concrete usability and performance compared to current Geo/BI systems.

REFERENCES

[1]   Gartner.com, "Gartner Forecasts Global Business Intelligence Market to Grow 9.7 Percent in 2011," Gartner, 18 02 2011. [Online]. Available: http://www.gartner.com/it/page.jsp?id=1553215. [Accessed 26 05 2012].

[2]   Gartner.com, «Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $18.3 Billion in 2017,» 17 02 2017. [En ligne]. Available: https://www.gartner.com/newsroom/id/3612617.

[3]   S. Rizzi, A. Abello, J. Lechtenborger et J. Trujillo, «Research in data warehouse modeling and design:dead or alive?,» DOLAP, pp. 3-10, 2006.

[4]   A. C. S. C. N. &. B. S. Sarkar, «Implementation of Graph Semantic Based Multidimensional Data Model: An Object Relational Approach,» International Journal of Computer Information System and Industrial Management Applications (IJCISIM), 3, pp. 127-136, 2011.

[5]   E. Malinowski and E. Zimányi, "Hierarchies in a multidimensional model: from conceptual modeling to logical representation," Data Knowl. Eng., 59 (2), p. 348–377, 2006.

[6]   C. A. Hurtado and A. O. Mendelzon, "OLAP dimension constraints," in ACM PODS, 2004.

[7]   J. Lechtenborger et G. Vossen, « Multidimensional normal forms for data warehouse design,» Information Systems, vol. 28, n°%15, p. 415–434, 2003.

[8]   J. Pardillo, J.-N. Mazón et J. Trujillo, «Bridging the semantic gap in OLAP models: platform-independent queries,» DOLAP, p. 89–96, 2008.

[9]   O. Glorio et J. Trujillo, «An MDA Approach for the Development of Spatial Data Warehouses,» chez DaWak, Turin, Italy, 2008.

[10]   Wikipedia.org, «Model-driven architecture,» en.wikipedia.org, 26 11 2012. [En ligne]. Available: http://en.wikipedia.org/wiki/Model-driven_architecture. [Accès le 26 11 2012].

[11]   Wikipedia, «Object Constraint Language,» wikipedia.org, 26 11 2012. [En ligne]. Available: http://en.wikipedia.org/wiki/Object_Constraint_Language. [Accès le 26 11 2012].

[12]   Wikipedia.org, «QVT,» en.wikipedia.org, 26 11 2012. [En ligne]. Available: http://en.wikipedia.org/wiki/QVT. [Accès le 26 11 2012].

[13]   V. Stefanov et B. List, «Business Metadata for the DataWarehouse - Weaving Enterprise Goals and Multidimensional Models,» chez 10th IEEE Int.Enterprise Distributed Object Computing Conference Workshops, 2006.

12   Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

[14] K. Shahzad et J. Zdravkovic, «Towards Goal-driven access to Process Warehouse: Integrating Goals with Process Warehouse for Business Process Analysis,» chez Fifth IEEE International Conference on Research Challenges in Information Sciences (RCIS), Guadeloupe, France, 2011.

[15] C. Diamantini et D. Potena, «Semantic Enrichment of Strategic Datacubes,» chez 11th International Workshop on Data Warehousing and OLAP (DOLAP 2008), New York, 2008.

[16] M. Kehlenbeck and M. Breitner, "Ontology-Based Exchange and Immediate Application of Business Calculation Definitions for Online Analytical Processing," in DaWaK 2009, 2009.

[17] M. Niinimaki et T. Niemi, «An ETL Process for OLAP Using RDF/OWL Ontologies,» Journal on Data Semantics XIII. LNCS, vol. 5530, Springer, Heidelberg, p. 97–119, 2009.

[18] T. Niemi, S. Toivonen, M. Niinimaki et J. Nummenmaa, « Ontologies with Semantic Web/grid in data integration for OLAP.,» International Journal on Semantic Web and Information Systems, Special Issue on Semantic Web and Data Warehousing 3(4), 2007.

[19] T. Niemi et M. Niinimaki, «Ontologies and summarizability in OLAP,» chez ACM Symposium on Applied Computing SAC 10, 2010.

[20] N. Shah, C. Tsai, M. Marinov, J. Cooper, P. Vitliemov et K. Chao, «Ontological On-line Analytical Processing for Integrating Energy Sensor Data,» Iete Technical Review 26, 375, 2009.

[21] T. Gu, X. Wang, H. Pung et D. Zhang, «An ontology-based context model in intelligent environments,» chez Communication Networks and Distributed Systems Modeling and Simulation Conference, San Diego, CA, USA, 2004.

[22] IHMC.US, "COE," 10 03 2012. [Online]. Available: http://www.ihmc.us/sandbox/groups/coe/wiki/welcome/att achments/d2a1b/COEmanual06.pdf. [Accessed 10 03 2012].

[23] É. Dubé, «Conception et développement d'un service web de constitution de mini cubes solap pour clients mobiles,» Msc Dep. Sciences geomatiques, Centre de Recherche en geomatique, Univ. Laval, Quebec, QC, Canada, 2008.

[24] OMG, «Common Warehouse Metamodel (CWM) specification,» Object Management Group, Inc., 2003.

Authors' Profiles

Belko Abdoul Aziz Diallo, Eng., PhD. (male) is a GIS and Big Data scientist. He holds the position of Senior Data Manager at the Competence center of WASCAL (West African Science Service Centre on Climate Change and Adapted Land Use). He completed his PhD research program on mobile Geospatial Business Intelligence at University of Laval (Quebec, Canada).

After the completion of his PhD, he has been serving as founder and CEO of a consultancy IT company, autonomous researcher, and visiting professor in several universities where he taught Geomatics, Data Mining, Internet Protocols, Web and Databases integration, Multi-layered Architectures, etc. Dr. Diallo has cumulated more than ten years of experience in conducting IT, teaching, and research and development activities in Burkina Faso and Canada. He joined WASCAL to head the data management and ICT unit. His fields of interest are related to Data Acquisition and Metadata Management, Data Transformation Storage and Quality Management, IT Architecture and Data Center Management, Data Warehousing and Decision support Management, Data/knowledge Visualization and Dissemination, Data Policy and security Management, etc.

Dr. Diallo is involved in the publications of several international peer-reviewed journals and conference talks. He is member of DAMA-I, the Data Management Association International, a Global Data Management Community.

Thierry Badard, PhD., Eng. is professor in geoinformatics at the Department of geomatics sciences of Université Laval in Quebec City (Canada).

He is the director of the Centre for Research in Geomatics (CRG) and is also on the steering committee of the Big Data Research Centre (BDRC) at Université Laval. He has more than 20 years of experience and he has been involved and has led national and international R&D projects of importance. His research interest deals with geospatial Big data, location analytics, data integration and fusion for better decision support, geospatial Business Intelligence, IoT and smart cities. He acts as a chair, editor and reviewer for numerous international journals and scientific conferences and has already an important record of scientific contributions. Dr. Thierry Badard is also actively involved in the geospatial free and open source community. He is developer, administrator and project coordinator of several open source projects : GeoKettle, GeoMondrian, SOLAPLayers and GeOxygene. He is an OSGeo charter member and has acted as a member of the OSGeo conference committee and a reviewer for the OSGeo Journal for several years. He is one of the founding co-chairs the OSGeo Quebec local chapter and a founding co-chair of the ICA (International Cartographic Association) commission on open source geospatial technologies.

He has also recently founded Ekumen, a company specialised in Location analytics & geomarketing where he acts as CTO. For further details, please visit http://www.ekumen.biz & http://www.crg.ulaval.ca.

Frédéric Hubert, Ph.D., received his PhD degree in computer sciences in 2003 from University of Caen (France).

He has more than 18 years of experience in the Geoinformatics field. Since 2007, he is a professor at the Department of Geomatics Sciences at Université Laval, Quebec, Canada. He is also member of the Centre de Recherche en Géomatique (CRG). His research interests are mainly concentrated on geovisualization, geospatial business intelligence, geospatial multimodal interactions, mobile spatial context, mobile augmented reality, and geospatial web services. Currently, he is more involved in research on cultural and noise mapping in urban and rural contexts. He has also been reviewer for various international scientific conferences and journals.

Dr Hubert is a member of different associations (ISPRS, ICA, CIG, OSGEO). He has published over 30 papers in peer-reviewed journals, conferences. He was also involved, as co-editor, in 3 books.

Towards Semantic Geo/BI: A Novel Approach for Semantically Enriching Geo/BI Data with OWL   **13**
Ontological Layers (OOLAP and ODW) to Enable Semantic Exploration, Analysis and Discovery of Geospatial Business
Intelligence Knowledge

Sylvie Daniel, Eng., M.Sc., Ph.D. graduated in electrical engineering in 1994 (France).

She received the PhD degree in signal processing and telecommunications in 1998 from University of Rennes I (France). She has been working with imagery and geospatial data for over 20 years, including in the industry where she led and managed international research and developments projects. Since 2004, she has been at Laval University (Quebec, Canada) where she is now a FULL PROFESSOR. She has held competitive research funding from several agencies including the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Innovation. Strongly committed to research, her interests include data acquisition (images and LiDAR data), image processing and artificial intelligence, 3D modeling, data fusion and augmented reality. Her research projects focus on 3D modeling of infrastructure and urban environment, on new solutions for hydrographic data collection and on new edutainment tools based on geomatics technologies and augmented reality. She has been actively involved in smart cities and communities and more specifically on 3D technologies contribution towards citizen participation and education. In such a context, in 2014, she was the acting director of the Institute for Information Technology and Society at Laval University, a key player in the field of smart cities.

Dr. Daniel is a senior member of IEEE and member of ACM. She has over 45 papers published in peer-reviewed journals and conferences. She was involved in the publication of 2 books and 5 book chapters.