Towards a Spatial Multidimensional Model

S. Bimonte
Laboratoire d'InfoRmatique
en Images et Systèmes d'information
UMR CNRS 5205
INSA, 7 avenue Capelle
69621 Villeurbanne Cedex, France
Sandro.Bimonte@insa-lyon.fr

A. Tchounikine
Laboratoire d'InfoRmatique
en Images et Systèmes d'information
UMR CNRS 5205
INSA, 7 avenue Capelle
69621 Villeurbanne Cedex, France
Anne.Tchounikine@insa-lyon.fr

M. Miquel
Laboratoire d'InfoRmatique
en Images et Systèmes d'information
UMR CNRS 5205
INSA, 7 avenue Capelle
69621 Villeurbanne Cedex, France
Maryvonne.Miquel@insa-lyon.fr

ABSTRACT
Data warehouses and OLAP systems help to interactively analyze huge volume of data. This data, extracted from transactional databases, frequently contains spatial information which is useful for decision-making process. Integration of spatial data in multidimensional models leads to the concept of SOLAP (Spatial OLAP). Using a spatial measure as a geographical object, i.e. with geometric and descriptive attributes, raises problems regarding the aggregation operation in its semantic and implementation aspects. This paper shows the requirements for a multidimensional spatial data model and presents a multidimensional data model which is able to support complex objects as measures, inter-dependent attributes for measures and aggregation functions, use of ad-hoc aggregation functions and n to n relations between fact and dimension, in order to handle geographical data, according to its particular nature in an OLAP context.

Categories and Subject Descriptors
H.2.1 [Logical Design]: Data models, H.2.8 [Database Application]: Spatial databases and GIS, Statistical databases

General Terms
Design

Keywords
Spatial OLAP, spatial data, multidimensional data, data modelling.

1. INTRODUCTION

It has been estimated that about 80% of the data stored in corporate databases integrates spatial information [8]. It is obvious that this meaningful data is worth being integrated in decision making process as a first class knowledge. Research works on integration of spatial information into multidimensional models lead to the definition of Spatial On Line Analytical Processing (SOLAP) concepts.

Geographic Information Systems (GIS) are powerful tools used to manipulate, query and visualize spatial databases. They also provide various functions to analyze spatial data (e.g. research or proximity functions) [11] [20]. Therefore GIS play an important role in spatial making-decision process, and they are sometimes considered as Supporting Decision Systems [5]. SOLAP solutions usually lie on coupling OLAP functionalities used to provide multidimensionality and GIS functionalities used to store and visualize the spatial information [10] [23]. However, SOLAP can not be reduced to a simple coupling architecture, but implies a real re-thinking of OLAP concepts. Spatial information can be integrated in multidimensional models as axis and/or subject of analysis. The numerous on going works (amongst them [3, 10, 14, 15, 17, 22]) confirm the importance and the innovating character of SOLAP. A SOLAP paradigm should define spatial measure with adapted aggregation function, spatial OLAP algebra extending OLAP operators to topological, spatial operators, dimension hierarchies based on spatial relations, suited GUI interface for navigation mixing tabular and cartographic features: Spatial facts, spatial measures and aggregation, and spatial dimensions characterize the SOLAP. Defining a formal, logical and physical framework, tracing the perimeter of analysis processes associated to SOLAP, are challenging research themes.

This work aims to provide a proposal for a formal model for spatial multidimensional databases. Spatial OLAP applications present some peculiar characteristics which can not be taken into account in classical multidimensional models, in particular the concept of spatial measure as geographic object taking in account its various attributes. This paper is organized as follows. Section 2 shows definitions of spatial information, main SOLAP concepts and related works. In Section 3, a use case is given and several requirements for a spatial multidimensional model, focusing on spatial measure, are discussed. In Section 4, we present our Spatial Multidimensional Model. We conclude in Section 5 with discussion and current and future works.

2. RELATED WORKS

2.1 Spatial information

Many authors have defined spatial information. [20] defines a "geographical" object as a real world entity that can be described using two components: the *descriptive* component is a set of alphanumerical attributes (e.g. name, population of a region) and the *spatial* component corresponds to its geometry (e.g. a point, a polygon…) and its topology relationships with other geographical objects (e.g. adjacency). Many other authors also define the *metric* part of the data which can be calculated (e.g. area). [26] calls "spatially referenced objects" or "geo-object" an object with spatial, temporal, graphical and alphanumerical attributes. In all these definitions, authors agree to define spatial information as an object composed of a set of attributes that contains descriptive and geometric data. We define a complex object as a real world entity represented by a set of attributes (age, name…). We call spatial information or geographical object a complex object that contains spatial attributes (i.e. geometry).

2.2 SOLAP

A first way to take benefits of spatial information in a decisional application is to use it as an analysis axis i.e. as a dimension in a multidimensional model. Spatial dimensions, as they are defined in literature, can be *spatial non geometric* (text only), *spatial geometric* (all levels have a cartographic representation) or *mixed spatial* (combine cartographic levels and textual levels) [3] [17]. In [15] a conceptual model for spatial OLAP is presented, based on integration of MultiDimER and MADS models. This work analyzes various scenarios of SOLAP applications using spatial dimensions, spatial measures and spatial facts. Representation of spatial hierarchies and of their topological relationships within this model is provided in [16]. [9] proposes a multidimensional formal model which is not based on any specific spatial data model, but is conceived to support the particular property of spatial hierarchies. The model introduces the partial containment relation in order to handle the case where several spatial objects representing members at different levels of the same hierarchy overlap (for example Roadway and District). [19] provides a unique formal framework to integrate a spatial database and a multidimensional database, exploiting the full-contains relation between hierarchy levels.

A different way to introduce spatial information in decisional process is using it as analysis subject, i.e. as a fact. Spatial measure then can be analyzed through non spatial and/or spatial dimensions. Among works related to SOLAP concepts, different authors have focused their attention on the concept of spatial measure. Different definitions can be found in literature: measure is sometimes represented as a collection of references to spatial objects [22] [21], as objects resulting from topological operations (e.g. union or intersection), or metric operators (e.g. distance) [21] [15], or as measure associated to a spatial dimension [14]. In [22], the spatial measure is reduced to its geometric part, and to metric attributes which are derived directly from the geometry. The proposal focuses on strategies to optimize aggregation and fusion of geometric values. A solution in order to associate to a spatial measure its descriptive characteristics is to report it, or to replicate it into a dimension [7]. Then the problem is to formalize and to manage this redundant or distributed representation throughout navigating across the cube. Indeed, in traditional cases, a measure is simple typed, and its semantic is limited to a quantitative description (a quantity, an amount…). On the contrary, in these solutions, the spatial measure is linked to an object which is member of a dimension, strongly typed, and constrained. Does the aggregation should be extended to the non geometric part of the measure, how hierarchy and/or semantic relationships are handled between the measure and its dimension representation are pending questions.

3. MOTIVATION

We now illustrate with a use case the requirements that are needed for a multidimensional model handling geometrical objects.

3.1 Use case

The application used to exemplify this work concerns the supervision of French departments in relation to mortal diseases. The mortality data represents numbers of deaths registered in France during the study period for a selected list of causes from the ICD (International Classification of Diseases), location of death, patients' gender and 5 year age-groups.

A first possible design model for this application is given figure 1. It uses an alphanumeric measure (`#deaths`) and 4 dimensions:

– `time` organized following the hierarchy (`day, month, year`),
– `location` with hierarchy (`department`, `region`),
– `sexage`, that represents the patients' gender and 5 year age-groups,
– and `causes` with hierarchy (`disease, classes`).

`location` is a spatial dimension, and `department` contains various descriptive attributes and the `geometry` attribute which is a reference to a geometric object. The aggregation function applied against the measure is the SQL SUM operator.

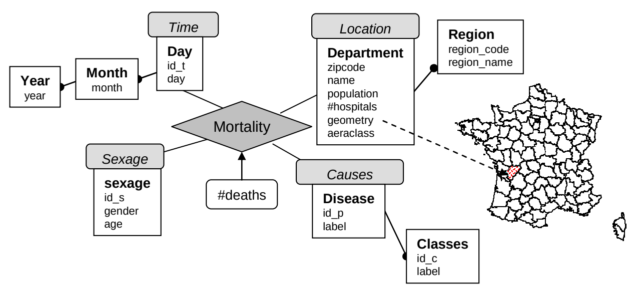

Figure 1. Mortality DW with Spatial Dimension Location.

The fact table is composed of dimensions' keys at their lower level that form the symbolic coordinates for the value of the measure. Thus the following functional dependency is enforced:

`zipcode, id_t, id_p, id_s → #deaths`

Dimension `location` can be represented as shown in table 1.

Table 1. Dimension `Location.`

zip co-de	name	po-pula tion	geom etry	#hosp itals	areaclass	reg-ion code
01	Ain	4	m01	1	Coastal and Country-side	RA
32	Gers	3	m32	2	Coastal and Country-side	A

38	Isère	6	m38	3	Mining and Manufac-turing	RA
69	Rhône	10	m69	10	Cities and services	RA

Another example of a spatial OLAP application is shown in figure 2. In this example, the number of deaths is transformed into dimension `incidence` which represents ranges of number of deaths. As in [22] we now use a spatial measure `department` which allows us to analyze locations in function of diseases, time, gender and age, and incidence. Aggregation will be applied against departments. These 2 application models correspond to different types of decision processes and consequently to diverse types of queries. In the first model the goal of the multidimensional analysis is the number of deaths, while the second model permits to answer to queries like: "*Where did cancer kill more than ten 50-years-old male persons in 1980?*" In this model, departments are the subject of the multidimensional analysis, so the user can deduce information about the influence of geographical location of departments in mortality problem. Moreover some other descriptive and metric attributes of the department can be useful to the decision process. For example the population and the number of hospitals can reveal an inadequate medical system or the causes of some diseases can have some relations with the areaclass of the department.

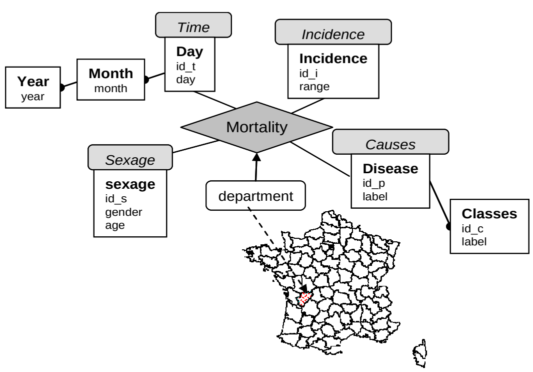

Figure 2. Mortality DW with Spatial Measure Department.

3.2 Requirements

Examining the two previously described application models, it is clear that the second one raises more questioning then the first one.

Primary, a same date, incidence range, disease and sex-age can characterize different departments. We do not use a list typed measure (e.g. a list of references) in order to individuate, at least at the lowest level, each department with its own description. Despite this, department still remains a measure and not a dimension, because it is the one on which we want to apply aggregation operations. This means that, in this particular example, there is no functional dependency between dimensions and measure as in the first application model.

`Department`, when defined in the dimension `location` of figure 1, is characterized by a set of descriptive attributes (`name`, `zipcode`, `population`, `nbhospitals`, `areaclass`) and geometric attributes (`geometry`) and by its hierarchy. A major problem is to determine how to take into consideration these attributes when a department is considered as a measure. Most of spatial models proposed in literature reduces a spatial measure to its geometric part and its aggregation function to a topological operator (as for example topological union), or a collection or list constructor. In our example the roll-up operation builds a new aggregated object from a set of objects at the most detailed level (for example the departments). A problem that arises from this scenario lies in establishing if the aggregated object has the same attributes of the departments and how to calculate them.

In this particular example, for the new aggregated object:

- The aggregated attribute name is not meaningful and should not be calculated.
- The aggregated attribute number of hospitals and the aggregated attribute geometry are calculated by means of the same detailed attributes and classical aggregation functions (sum and topological union).
- The aggregated attribute areaclass is calculated from the areaclass and the population of departments that have to be aggregated (ratio).

It is obvious that, as attributes are semantically dependant, changing an aggregation function for an attribute implies different aggregations for the other dependant attributes (for example geometric union implies sum of number of hospitals otherwise intersection implies difference). Consequently as attributes are dependant, aggregation functions are dependant too. Moreover, an aggregated attribute could be defined as very different from the detailed one, using for instance different data types [6], or different classification.

Thus, amongst the requirements for a spatial multidimensional model dealing with complex spatial measure are:

1. Support n to n relationships between facts and dimensions
2. Model measures as complex entities:
    2.1 Usage of every kind of spatial and alphanumeric attributes as one single measure
    2.3 Support of inter-dependent attributes and aggregation functions
    2.4 Use of ad hoc aggregation functions

Different models have been proposed in literature to model multidimensional databases. A complete survey of the properties of these models can be found in different works [4], [18]. Amongst all the proposed models we report models presented in [18], [24], [1] because they are representative of some important and advanced proprieties. [18] provides a formalism and an algebra for multidimensionality that focuses on the support of complex types of dimension hierarchies (supporting non-onto, non-covering hierarchies, etc…). Other important properties are provided as for example symmetrical treatment of measure and dimension and n to n relations between fact and dimension. Some models have been developed to represent multidimensional databases exploiting Object Oriented (OO) concepts. [24], [12], [13] present an OO multidimensional model based on UML that, among all its others features, consents to support aggregation semantics (specify what aggregation functions are allowed) using the Object Constraint Language (OCL) and measures sets. Moreover it emphasis modular conception allowing the design of complex multidimensional models using the package mechanism of UML. [1] provides an OO conceptual model based on UML too. In this work, the OO paradigm is used to model OLAP concepts (fact, dimension, etc…) and to establish relations between them (specialization of facts, etc…). The model permits to represent user defined aggregation functions and derived measures too. In [25], the proposed model utilizes the concept of basic cube (a set of cells associated at bottom levels of dimension hierarchies) and cube (a possible aggregated view of the basic cube) to support the drill-down operator.

4. A SPATIAL MULTIDIMENSIONAL MODEL

We provide a multidimensional model which permits to represent dimensions and spatial facts supporting the requirements individuated in section 3.2.

4.1 Overview

We introduce the concepts of entity schema and entity instances which allow modeling indifferently all objects of the analysis universe: members of dimensions and facts. Entity schemas and their instances are organized into hierarchies in a classical way. A base cube is assimilated to a multidimensional space. Instances of entities used to represent the members at the most detailed dimension levels and facts are projected on the axes of the multidimensional space. We permit n to n relation between fact and dimensions. The concept of aggregation mode is provided in order to define a way to calculate aggregation of entity schema's attributes, and so to support measure as complex object, and in the case of spatial measure as geographical object. In a similar way to the concept of base cube, a cube is a multidimensional space where at least one dimension is not at its most detailed level and the fact schema corresponds to an aggregated schema.

4.2 Data Model

Supposing a set of domain dom_1,… dom_n, we call an *attribute* an alphanumeric identifier a_i associated to a domain represented by dom(a_i). An example of attribute is `population` whose domain is dom(`population`) = $\mathbb{N}$. The *value* of an attribute a_i is a value belonging to the domain dom(a_i) and is denoted val(a_i). For example val(`population`)=254000.

Definition 1. (*Entity Schema*)

An entity schema S_e is a tuple of attributes denoted $S_e = \langle a_1, \dots a_n\rangle$ where a_i is an attribute defined on dom(a_i).
We note $S_e.a_i$ the i^{th} attribute a_i of S_e and $\mathcal{A}(S_e)$ the set of the a_i attributes of the entity schema S_e: $\mathcal{A}(S_e) = \{a_1, \dots a_n\}$

Example 1. S_{dept} is an entity schema modeling French departments.
S_{dept} = ⟨`zipcode, name, population, areaclass, nbHospitals, geometry`⟩
dom(`areaclass`) = {`'Cities and Services', 'Coastal and Countryside','Mining and Manufacturing'`}
$S_{dept}.a_1$ = `zipcode` $S_{dept}.a_3$ = `population`
`geometry` is a reference on a spatial object. `population` is given in thousands of inhabitants.

Definition 2. (*Entity Instances*)

An entity instance t_i of an entity schema S_e is a tuple that associates values from dom(a_i) to each attribute a_i of schema S_e:
$t_i = \langle val(a_1), \dots val(a_n)\rangle$ with $a_i \in \mathcal{A}(S_e)$ and $val(a_i) \in dom(a_i)$
We note $t_i.a_j$ the value of the j^{th} attribute of instance t_i and $\mathcal{I}(S_e)$ the set of the tuples t_i instances of S_e. We assume that we can build a total order (arbitrary or not) on the instances of a schema.

Example 2. The instances of departments are:
$\mathcal{I}(S_{dept})$ = {⟨`69,'Rhône',10,'Cities and Services',10,ptr69`⟩,⟨`38,'Isère',6,'Mining and Manufacturing',3,ptr38`⟩,⟨`01,'Ain',4,'Coastal and Countryside',1,ptr01`⟩,⟨`32,'Gers',3,'Coastal and Countryside', 2,ptr32`⟩, …}

t_1 is a tuple of $\mathcal{I}(S_{dept})$:
t_1 = ⟨`69,'Rhône',10,'Cities and Services',10, ptr69`⟩
$t_1.a_2$ = `'Rhône'`

4.3 Base Cube, Aggregation Mode and Cube Models

We use entity schemas and instances to model all the real world objects involved in a multidimensional application model. That is to say that, in our model, dimensions and facts are based on entities. We now define how entities participate to hierarchies and cubes.

Definition 3. (*Hierarchy Schema*)

A hierarchy schema is a tuple $SH_h = \langle \mathcal{L}, \perp_{SH_h}, \ddagger\rangle$ where $\mathcal{L}$ is a set of entity schemas, $\perp_{SH_h}$ is an entity schema, and $\ddagger$ is a partial order on $\mathcal{L} \cup \{\perp_{SH_h}\}$.

$\perp_{SH_h}$ is called the *bottom* schema of the hierarchy schema. The entity schemas belonging to $\mathcal{L} \cup \{\perp_{SH_h}\}$ are called *levels*, $\ddagger$ builds an oriented graph where entity schemas are nodes, $\perp_{SH_h}$ is the root and the arcs represent the partial order relationship.

Example 3. We can define the following hierarchy schema:
$SH_{time} = \langle \mathcal{L}_{time}, S_{day}, \ddagger\rangle$; $\mathcal{L}_{time} = \{S_{month}, S_{year}\}$ and $\perp_{SH_{time}} = S_{day}$
$\ddagger$: ($S_{day} \ddagger S_{month}$) and ($S_{month} \ddagger S_{year}$)

$SH_{location} = \langle \mathcal{L}_{location}, S_{dept}, \ddagger\rangle$; $\mathcal{L}_{location} = \{S_{region}\}$
and $\perp_{SH_{location}} = S_{dept}$
$\ddagger$: ($S_{dept} \ddagger S_{region}$)

$SH_{sexage} = \langle \mathcal{L}_{sexage}, S_{sexage}, \Uparrow \rangle$; $\mathcal{L}_{sexage} = \{\varnothing\}$ and $\perp_{SH_{sexage}} = S_{sexage}$

$\Uparrow : \varnothing$

$SH_{incidence} = \langle \mathcal{L}_{incidence}, S_{incidence}, \Uparrow \rangle$; $\mathcal{L}_{incidence} = \{\varnothing\}$ and $\perp_{SH_{incidencee}} = S_{incidence}$

$\Uparrow : \varnothing$

Definition 4. (*Hierarchy Instance*)

The instance of a hierarchy schema $SH_h = \langle \mathcal{L}, \perp_{SH_h}, \Uparrow \rangle$ is a set of tuples t_i so as if $t_i \in \mathcal{I}(S_i)$ and $S_i \Uparrow S_j$ then $\exists\, t_j \in \mathcal{I}(S_j)$ and an order relation $\uparrow$ so as $t_i \uparrow t_j$

In figure 3 is shown the instance of the hierarchy `time`. A link between two dates $date_1$ and $date_2$, means that $date_1 \uparrow date_2$

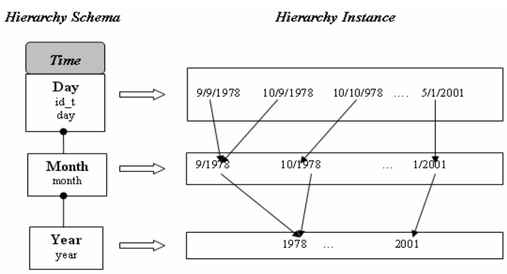

Figure 3. Hierarchy Time: schema and instance.

Definition 5. (*Base Cube Schema*)

A Base Cube schema SBC_{bc} is a tuple $SBC_{bc} = \langle S_1, \dots S_m, S_f, \delta \rangle$ where :

- $\forall i, j \in [1,..m, f]$, S_i is an entity schema and $S_i \neq S_j$
- $\forall i \in [1,..m]$, a hierarchy schema exists $SH_i = \langle \mathcal{L}_i, \perp_{SH_i}, \Uparrow \rangle$

such as $S_i = \perp_{SH_i}$

- δ is a boolean function defined from $\mathcal{I}(S_1) \times .. \times \mathcal{I}(S_m) \times \mathcal{I}(S_f)$ to $\{0, 1\}$

Thus, in other words, the base cube schema represents a fact table with all dimensions at their bottom level except one which is taken at any level. This latter will play the role of the measure. The boolean function allows to model a n to n relationship between dimensions and measure. The base cube can be viewed as a multidimensional space where instances of dimensions and facts are projected on axis and points are 0 or 1 as shown in figure 4. The 0 value means that there is no data corresponding to this combination of instances.

Example 4. We report a base cube schema used to represent the application model of figure 2.

$SBC_{mortality\_base} = \langle S_{day}, S_{sexage}, S_{disease}, S_{incidence}, S_{dept}, \delta \rangle$

Definition 6. (Base Cube instance)

The base cube instance for a base cube schema $SBC_{bc} = \langle S_1, \dots S_n, S_f, \delta \rangle$ is a set bc_c of tuples such as:

$bc_c = \{\langle t_1^j, \dots, t_n^j, f_m^j \rangle, j=1,..p\}$ where $t_i^j \in \mathcal{I}(S_i)$ and $f_m^j \in \mathcal{I}(S_f)$ and $\delta(t_1^j, \dots, t_n^j, f_m^j) = 1$

In other words, bc_c is the set of tuples composed of entity instances for which δ is 1.

Example 5. The instance of the base cube $SBC_{mortality\_base}$ is:

$bc_{mortality}$ = { ⟨⟨05jan2001⟩,⟨'H40.45','Male','40-45'⟩, ⟨201,'cancer'⟩,⟨2,'20-30'⟩, ⟨69,'Rhône',10,'Cities and Services',10, ptr69⟩⟩, ⟨⟨05jan2001⟩,⟨'F40.45',Female,'40-45'⟩,⟨201,'cancer'⟩,⟨2,'20-30'⟩, ⟨69,'Rhône',10,'Cities and Services', 10,ptr69 ⟩⟩, …}

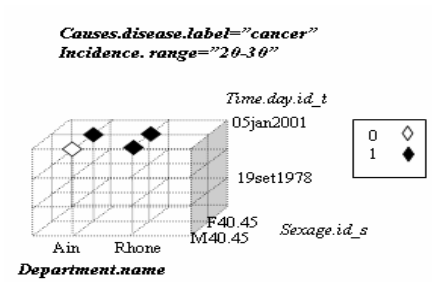

Figure 4. Multidimensional space view of Mortality application.

From figure 4, we can observe that in 5-01-2001, there were from 20 to 30 male persons and from 20 to 30 female persons who died from cancer in Rhône. No male, but from 20 to 30 female persons in 5-01-2001 are dead from cancer in Ain.

Definition 7. (*Aggregation Mode*)

An aggregation mode Θ_k is a tuple $\langle S_a, S_b, \Phi \rangle$ where S_a is an entity schema $\langle a_1, ..a_m \rangle$, S_b is an entity schema $\langle b_1, ..b_p \rangle$, and Φ is a set of p ad-hoc aggregation functions ϕ_i.

We then say that S_b is *built_from* S_a. The aggregation of n instances of S_a is an instance $ta_i = \langle val(b_1), \dots, val(b_p) \rangle$ of S_b such as:

$$\forall j \in [1,..p], val(b_j) = \phi_{i=1}^{n}(t_i.a_1, \dots t_i.a_k) \text{ with}$$

$t_i \in \mathcal{I}(S_a)$ and $a_r \in \mathcal{A}(S_a)$ where $r \in [1, \dots k]$.

The concept of aggregation mode is provided in order to support measure as complex entity or complex object. The idea is to get one entity representing the detailed measure and another one representing the measure after the aggregation process and to link their instances through aggregation functions, one for each attribute of the "aggregated entity". These functions are ad-hoc user defined functions and they establish how to calculate the value of the attribute of the "aggregated entity", associated to that function, from attributes values of the original measures.

Example 6. Let us define $\Theta_1 = \langle S_{dept}, S_{agg\_dept}, \Phi_1 \rangle$

The schema S_{agg_dept} is *built-from* S_{dept}. Its attributes are: b_1 = `agg_population`, b_2 = `agg_geometry`, b_3 = `agg_areaclass`, b_4 = `agg_nbHospitals`

$\Phi_1 = \{\phi_1, \phi_2, \phi_3, \phi_4\}$
ϕ_1 : dom(S_{dept}.`population`)$^n \rightarrow$ dom(S_{agg_dept}.`agg_population`) *// sum of populations*
ϕ_2 : dom(S_{dept}.`geometry`)$^n \rightarrow$ dom(S_{agg_dept}.`agg_geometry`) *// geometric fusion*
ϕ_3 : (dom(S_{dept}.`population`) $\times$ dom(S_{dept}.`areaclass`))$^n \rightarrow$ dom(S_{agg_dept}.`agg_aeraclass`) *// ratio*
ϕ_4 : dom(S_{dept}.`nbHospitals`)$^n \rightarrow$ dom(S_{agg_dept}.`agg_nbHospitals`) *// sum*

Examples of aggregated instances are:

ta_1 = ⟨`14,p1,'Cities and Services',11`⟩ ∈ $\mathcal{I}$($S_{dept\text{-}agrégé}$) // *result of aggregation on Rhône and Ain departements (figure 5)*

ta_2 = ⟨`13,p2,'Coastal and Countryside',6`⟩ ∈ $\mathcal{I}$($S_{dept\text{-}agrégé}$) // *result of aggregation on Isère and Ain and Pyrénées-Atlantiques departements*

Example 7. Let us define $\Theta_2 = \langle S_{incidence}, S_{agg\_incidence}, \Phi_2 \rangle$
The schema $S_{agg\_incidence}$ is *built-from* $S_{incidence}$.
$\Phi_2 = \{\phi_5\}$
ϕ_5 : dom($S_{incidence}$.`incidence`) $\rightarrow$ dom($S_{agg\_incidence}$.`agg_incidence`) // *average*

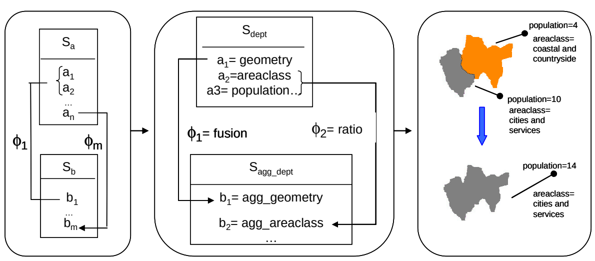

Figure 5. Aggregation of Rhône and Ain departments.

Definition 8. (*Cube Schema*)

A Cube Schema SC_c is a tuple $SC_c = \langle SBC_{bc}, \mathcal{L}, \Theta_f, \gamma \rangle$ where

- $SBC_{bc} = \langle S_{b1}, \ldots S_{bm}, S_f, \delta \rangle$ is a schema base cube
- $\mathcal{L}$ is a set of entity schemas S_i such as $\forall S_i \in \mathcal{L}$, $\exists\, S_{bi} \in \{S_{b1}, \ldots S_{bm}\}$ such as $S_{bi} \not\Uparrow S_i$ or $S_{bi} = S_i$ and $\exists$ at least one $S_k \in \mathcal{L}$ so that $S_{bi} \not\Uparrow S_k$

i.e. the schemas S_i of $\mathcal{L}$ are such that they appear in the same hierarchy schema as the entity schemas S_{bi} of SBC_{bc} and at least one of them is not a bottom

- Θ_f is an aggregation mode $\langle S_f, S_{af}, \Phi \rangle$
- γ is a boolean function from $\mathcal{I}(S_1) \times .. \times \mathcal{I}(S_m) \times \mathcal{I}(S_{af})$ to $\{0, 1\}$

Example 8. $SC_{mortality1}$ implements our second application model with department being the spatial measure, and incidence being a dimension:

$SC_{mortality1} = \langle SBC_{mortality\_base}, \{\langle S_{month}, S_{sexage}, S_{disease}, S_{incidence} \rangle\}, \Theta_1, \gamma_1 \rangle$

Example 9. The following cube implements our first application model with incidence (i.e. number of deaths) being the measure and department being a dimension:
$SC_{mortality2} = \langle SBC_{mortality\_base}, \{\langle S_{month}, S_{sexage}, S_{disease}, S_{dept} \rangle\}, \Theta_2, \gamma_2 \rangle$

Definition 9. (*Cube Instance*)

The instance of a Cube Schema $SC_{bc} = \langle SBC_{bc}, \mathcal{L}, \Theta_f, \gamma \rangle$ is a set c_c such as:

$c_c = \{\langle t_1^j, \ldots, t_n^j, f_m^j \rangle, j=1,..p\}$ where $S_i \in \mathcal{L}$, $t_i^j \in \mathcal{I}(S_i)$ and $f_m^j \in \mathcal{I}(S_{af})$ and $\gamma(t_1^j, \ldots, t_n^j, f_m^j) = 1$

in other words, c_c is the set of tuples composed of entity instances for which γ is true. Consequently measure values are represented by instances of the "aggregated entity" (S_{af}) belonging to the aggregation mode (Θ) associated to the cube.

Example 10. The instance of the cube $SC_{mortality1}$ is:
$c_{mortality1}$ = {⟨⟨`jan2001`⟩,⟨`'F40.45','Female','40-45'`⟩, ⟨`201,'cancer'`⟩,⟨`2,'20-30'`⟩, ⟨`14,p1,'Cities and Services',11`⟩ ⟩, …}

Example 10 shows the relation between base cube and cube. The instances of a cube are the instances of the base cube after the aggregation process, as for example a roll-up operation, using a particular aggregation mode. The first tuple of $c_{mortality1}$ shows that the investigating for the areas where 20-30 female persons died from cancer in January 2005, gives an area, made by Ain and Rhône department, with 11 hospitals, 14 thousands inhabitants and qualified `'Cities and Services'`.

5. DISCUSSION AND ON-GOING WORKS

Spatial OLAP aims to integrate spatial data into multidimensional databases. Supporting and exploiting the particular nature of spatial data into multidimensional analysis implies a re-thinking of basic OLAP concepts, as for example spatial dimensions or spatial measures, at formal, logical and physical levels. If some efforts have been done to handle spatial dimensions in OLAP, spatial measure concept has been re-formulated but no multidimensional formal model support it.

In this work we have defined measures and dimension members as complex objects. The proposed multidimensional data model (Spatial Multidimensional Model) satisfies the following requirements: support of n to n relationships between fact and dimension, usage of every kind of spatial and alphanumeric attributes as one single measure, support of inter-dependent attributes and aggregation functions, use of ad hoc aggregation functions, in order to handle spatial measures as geographical object. Our Spatial Multidimensional Model provides concepts of attribute and entity (schema and instances) to model information involved in the multidimensional application. Exploiting these concepts the model gives the concepts of hierarchy, base cube, aggregation mode, and cube to organize data into a multidimensional way. A base cube projects dimensions and measure on the axes of a multidimensional space allowing the symmetrical treatment of measure and dimension and n to n relations between fact and dimension. Aggregation mode, relating two different entities through aggregation functions permits to support measure as complex object. As the base cube, a cube represents a multidimensional space where measures are aggregated and dimensions are not all at most detailed levels. Our model handles measures described by complex objects and, even if it does not adopt any specific spatial data model, it is well suited for geographical data as defined in section 1. However, some specific problems due to the spatial nature of the measure must be further investigated. For instance, pre-aggregation of the measure in the cuboids lattice, additivity question, or topological relations in spatial dimensions are not yet covered by our model.

At the moment, we work to define a multidimensional algebra for the model. This algebra must support all typical OLAP operators and some new SOLAP ones in order to introduce spatial analysis operators in multidimensional databases.

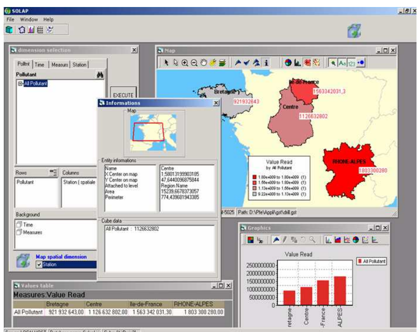

Figure 6. GéOlap, a Spatial Olap tool.

In previous works, we have implemented a prototype named GéOlap (figure 6) that handles spatial dimensions and allow OLAP navigation into synchronized tabular, cartographic and diagram representations [23]. This prototype authorizes the management of numeric measures. We are now working in order to integrate spatial measures in GéOlap.

6. ACKNOWLEDGEMENTS

This work has been supported by the CORILA's Research Program.

7. REFERENCES

[1] Abelló, A., Samos, J., and Saltor, F. YAM² (Yet Another Multidimensional Model): An extension of UML. In *Proceedings of International Database Engineering & Applications Symposium*. (Canada, July 2002). IEEE Computer Society Press, 2002, 172-181.

[2] Bédard, Y., and Bernier E. Supporting multiple representation with spatial databases views management and the concept of VUEL. In *Proceedings of Workshop on Multi-Scale Representations of Spatial Data* (Ottawa Canada, 2002). ISPRS / ICA.

[3] Bédard, Y., Merrett, T., and Han, J. Fundaments of Spatial Data Warehousing for Geographic Knowledge Discovery. *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, London, 2001, 53-73.

[4] Blaschka, M., Sapia, C., Hofling, G., and Dinter, B. Finding Your Way through Multidimensional Data Models. In *Proceedings of the 9th International Conference on Database and Expert systems Applications*. Lecture Notes in Computer Science, 1490 (1998), Springer-Verlag, 198-203

[5] Cowen, D.J. GIS versus CAD versus DBMS: What are the differences ?. *Fotogrammetric Engineering and Remote Sensing*, 54 (1988), 1551-1555.

[6] Feng, L., and Dillon, T. Using Fuzzy Linguistic Representations to Provide Explanatory Semantics for Data Warehouses. *IEEE TKDE*, 15, 1 (2003), 86-102.

[7] Fidalgo, R.N., Times, V.C., Silva, J., and Souza, F.F. GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. In *Proceedings of Int. Conf. on Data Warehousing and Knowledge Discovery* (Zaragosa, Spain, 2004). 26-37.

[8] Franklin, C. *An Introduction to Geographic Information Systems: Linking Maps to databases.* Database. 1992, 13-21.

[9] Jensen C., Klygis, A., Pedersen, T. and Timko, I. Multidimensional data modelling for location-based services. *VLDB Journal*, 13, 1 (2004), 1-21.

[10] Kouba Z., Matoušek K., and Mikšovský P. On Data Warehouse and GIS integration. In *Proceedings of the 11th Int. Conf. and Workshop on Database and Expert Systems Applications* (Greenwich, UK, 2000).

[11] Longley, P.A., Michael. F., Goodchild, D.J., and Maguire, Rhind, D. W. *Geographic Information Systems and Science*. John Wiley & Sons (August, 2001).

[12] Lujan-Mora, S., Trujillo, J. and Song, I. Extending UML for Multidimensional Modeling. In *Proceedings of the 5th International Conference on the Unified Modeling Language* (Dresden, Germany, 2002). LCNS, 2460, Springer-Verlag, 290-304

[13] Lujan-Mora, S., Trujillo, J. and Song, I. Multidimensional Modeling with UML package Diagrams. In Proceedings *of the 21st International Conference on Conceptual Modeling* (Tampere, Finland, 2002). LNCS, 2503, Springer-Verlag, 199-213.

[14] Marchand, P., Brisebois, A., Bédard, Y., and Edwards G. Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis. *Journal of the International Society of Photogrammetry and Remote Sensing*, 59 (2003), 6-20.

[15] Malinowski, E., and Zimányi, E. Representing spatiality in a conceptual multidimensional model. In *Proceedings of the 12th annual ACM International*

workshop on Geographic information systems. (Washington DC, New York, USA, 2004), ACM Press 2004, 12-22.

[16] Malinowsky, E., and Zimányi, E., Spatial Hierarchies and Topological Relationships in SpatialMultiDimER model. In *Proceedings of the 22th British National Conference on Databases* (Sunderland UK, July 2005), LNCS 3567, Springer, 17-28

[17] Miquel, M., Bédard, Y., and Brisebois, A. Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie *Revue ISI-NIS, Special Issue Data Warehousing*, Paris, Edition Hermès Science, 7, 3 (2002).

[18] Pedersen, T.B. and Jensen, C.S. Multidimensional data modelling for complex data. In *Proceedings of the 15th International Conference on Data Engineering*. IEEE Computer Society Press, 1999, 336-345.

[19] Pourabbas, E. Cooperation with geographic databases. *Multidimensional Databases: Problems and Solutions*. M. Rafanelli, Idea Group Publishing, 2003, 393-432.

[20] Rigaux, P., Scholl, M., Voisard, A. *Spatial databases with application to GIS*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 2002.

[21] Rivest, S., Bédard Y., and Marchand P. Toward Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica*, 55, 4 (2001), 539-555.

[22] Stefanovic, N., Han, J., and Koperski K. Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE TKDE,* 12*,* 6 (2000), 938-958.

[23] Tchounikine, A., Maryvonne M., Robert, L., Taher, A., Bimonte, S. and Baillot, V. Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes*. Revue des Nouvelles Technologies de l'Information* (RNTI), Editions Cepaduès, numéro spécial, 2005.

[24] Trujillo, J., Palomar, M., Gomez, J., and Song, I. Designing Data Warehouses with OO Conceptual Models. IEEE Computer, special issue on Data Warehouses, 34,12 (2001), 66-75.

[25] Vassialiadis, P. Modeling multidimensional databases, cubes and cube operations. In *Proceedings of the 10th International conference on Scientific and Statistical Database Management*. IEEE Computer Science Press, 1998, 53-62.

[26] Worboys, M.F., *GIS, a computing perspective*, Taylor&Francis, 1995