# Revisiting "Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels"

Robson do Nascimento Fidalgo, Valéria Cesário Times, Joel da Silva,
Fernando da Fonseca de Souza, Ana Carolina Salgado

Center for Informatics, Federal University of Pernambuco, P.O. Box 7851, Recife-PE, Brazil, 50.732-970
{rdnf,vct,js,fdfd,acs}@cin.ufpe.br

Categories and Subject Descriptors: Information Systems [**Miscellaneous**]: Databases

Keywords: DW, OLAP, GIS

## 1. INTRODUCTION

Extensive research has been conducted regarding the integration of Geographical Data Warehouse (GDW) and Spatial OnLine Analytical Processing (SOLAP) technologies. The former can be defined as an extension of the traditional DW approach by adding a geographical component. Basically, this consists of extending the star model through the inclusion of geographical properties (including descriptive and geometrical data), which are defined as GDW dimensions and/or measures. The latter permits processing data extracted from GDW and analysing them from different perspectives and with different degrees of details. Differently from existing proposals that mix concepts related to GDW (e.g. dimension table and fact table) and SOLAP cubes (e.g.hierarchy and level), we proposed the GeoDWFrame framework for guiding the design of geographical dimensional schemas and a pair of MOF based metamodels, namely GAM and GeoMDM [Fidalgo et al. 2004]. The motivation behing using separated metamodels for GDW and SOLAP cubes is to consider a GDW as a database that may be used by a range of query tools, including Crystal Reports packages and GIS softwares that do not deal with the concepts of levels and hierarchies. A specific aim of the work presented here is to gain an understanding of what else was done after having proposed these metamodels, by discussing the extended versions of our original approach and detailing our current and further work.

In this paper, a presentation of our work on providing data models for GDW (section 2 ) and SOLAP cubes (section 3) is given.

## 2. A METAMODEL FOR GEOGRAPHICAL DATA WAREHOUSE

We first proposed GeoDWFrame which is a high-level specification and defines a set of guidelines aiming to direct the definition of the dimensional and geographical schema of a DWG [Fidalgo et al. 2004]. Following this, some MOF based metamodels that are related to the CWM OLAP and OGC/ISO specifications were defined [Fidalgo et al. 2004]. Then, GeoDWM was specified, which is a metamodel based on GeoDWFrame, UML (Unifed Modeling Language), CWM (Common Warehouse Metamodel) and OGC standards and provides the conceptual modelling needs for the developing of GDW applications. GeoDWM defines how the concepts (e.g., measures and conventional and geographical dimensions) of a GDW schema can be organized and are related to each other. It also provides a set of stereotypes with pictograms that are meant to assist and guide the project designer in the GDW modelling activity and serves as a basic metamodel for CASE tools aimed at the conceptual modelling and automatic generation of logical GDW schemas. In addition, GeoDWM extends the original GeoDWFrame definitions to take into account spatial measures and to remove the Hybrid dimension specializations.

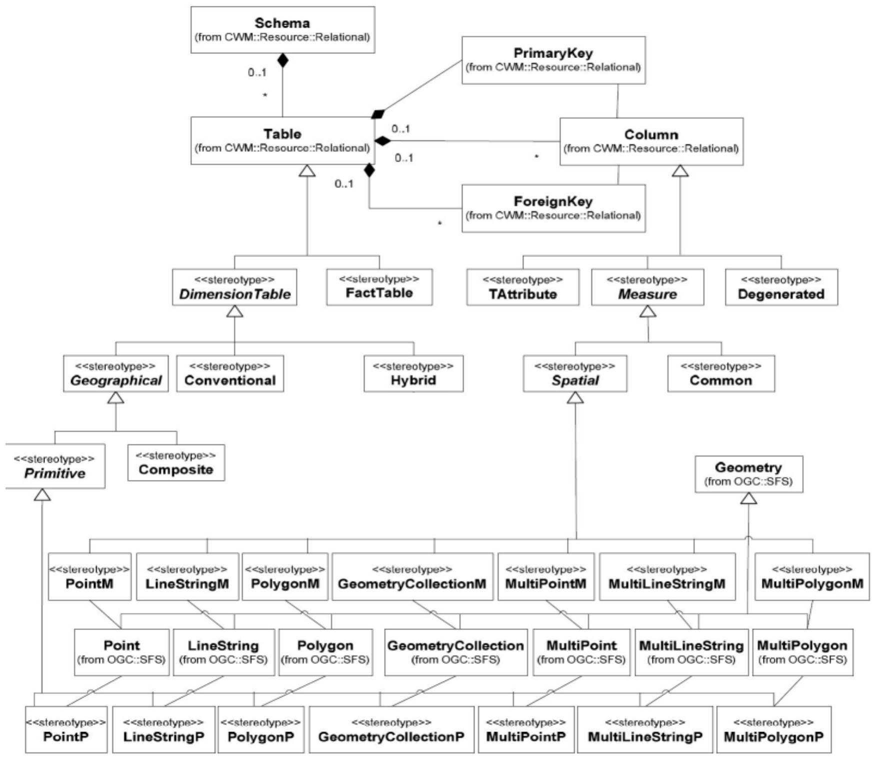


Fig. 1. The Class Diagram for the Proposed Metamodel

Figure 1 displays the UML class diagram for our GeoDWM, which is detailed in [da Silva et al. 2010] A dimension table (DimensionTable) can be specialized in three different dimensions: conventional, hybrid and geographical. In the first case, as in a traditional DW, a GDW also provides support for dimensions that store only conventional data. The other two types of dimensions model the concepts of the GeoDWFrame proposal [Fidalgo et al. 2004]. While the hybrid dimension deals with location descriptions and conventional data (e.g. client's address plus age and gender), geographical dimensions are specialized in composite (Composite) and primitive (Primitive) dimensions. The geographical composite dimensions only stores location descriptions (e.g. country names), whereas the geographical primitive dimensions maintain the coordinate data (e.g. geometries of countries) related to the corresponding location descriptions kept in composite or hybrid dimensions. While a primitive geographical dimension represents the geometrical data used to handle (e.g. draw, query and index) spatial objects, a composite geographical dimension contains attributes to represent its primary key, the location description of the geo-objects and its foreign keys to primitive geographical dimensions. It is worth noting that, without the primitive geographical dimensions, a dimension with spatial data would have to store its location descriptions and geo-references together. Due to the intrinsic redundancy of dimensional data and the high costs of storing the geo-references (compared to foreign keys to primitive dimensions), keeping these data together in a single dimension table has proven to have a high storage cost [Siqueira et al. 2009]. All the GeoDWM concepts presented in this section were used in the implementation of GeoDWCASE [Fonseca et al. 2007]. An example of a data model built using this modeling tool as well as further details on its implementation can be found in [da Silva et al. 2010].

## 3. A METAMODEL FOR SOLAP DATA CUBES

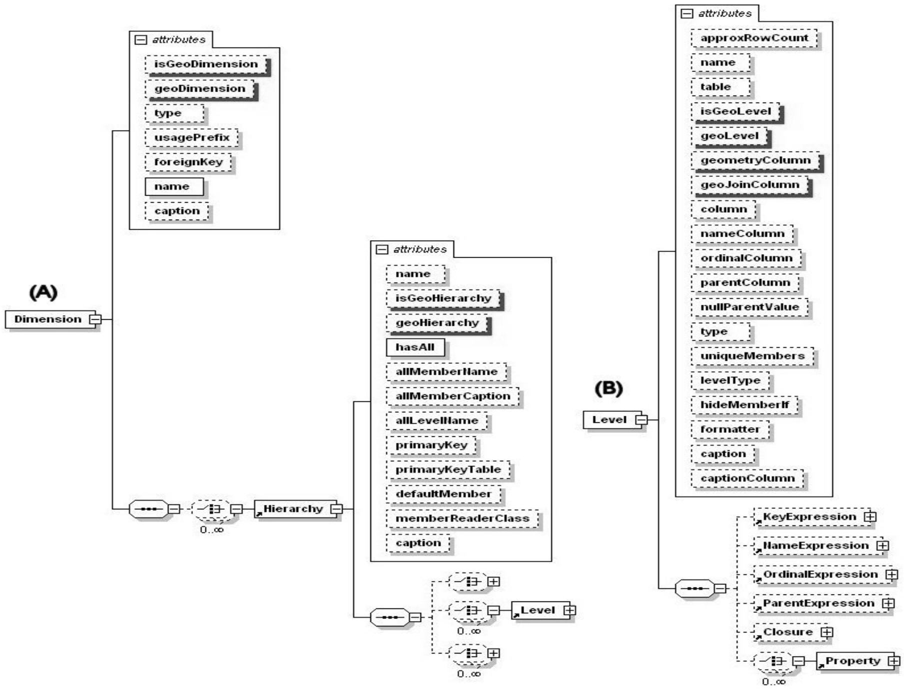


Fig. 2. *Dimension*, *Hierarchy* and *Level* elements of the Mondrian XML Schema

This section presents our metamodel used to define schemas of SOLAP data cubes. Our metamodel is based on the main GeoMDM metadata (i.e. *GeoCube*, *GeoDimension*, *GeoHierarchy* and GeoLevel) and was defined in XML because it extends the XML schema of the Mondrian OLAP server. For this, we have modified the Mondrian XML elements that define a *Schema*, a *Cube*, a *Dimension*, a *Hierarchy* and a *Level*. For each of them, new attributes and elements were added to the Mondrian XML schema to allow the manipulation of geographical data. For the element *Schema*, the following attributes were added: *isGeoSchema*, to indicate whether the current cube schema is geographical and *geoMDSchema*, to define the geographical schema name. Similarly, for the element *Cube*, the next two attributes, *isGeoCube* and *geoCube*, were added to specify whether the data cube is geographical and to store the geographical cube name, respectively. This was also the case with regard to the elements *Dimension* and *Hierarchy*, as shown in Figure 2. Regarding the element *Level* of the Mondrian Schema, the following four attributes were included: *isGeoLevel, geoLevel, geometryColumn* and *geoJoinColumn*. These were respectively used in our prototype system [da Silva et al. 2010] to: 1) indicate whether a certain level is geographical; 2) represent the name of the GDW dimension table that is associated with the level; 3) specify which column of the dimension table has the geometry of the level members; and 4) maintain the primary key of the dimension table that has the geographical data to be used in the level definition.

## 4. CONCLUSION AND FUTURE WORK

This paper reviews the literature on our data modelling approaches for representing GDW and SOLAP data cubes. A comparative analysis between our work and other studies on both multidimensional and geographical data modeling and SOLAP queries is given in [da Silva et al. 2010; Ruiz and Times 2009]. The following are some of our planned approaches for future work: to assess the potential application of SOLAP tools regarding to a user-needs perspective and to handle complex hierarchies that represent the partial containment of spatial objects as well as to represent spatial-temporal DW.

## REFERENCES

da Silva, J., de Oliveira, A. G., Fidalgo, R. N., Salgado, A. C., and Times, V. C. Modelling and querying geographical data warehouses. *Information Systems* 35 (5): 592–614, 2010.

Fidalgo, R. N., Times, V. C., da Silva, J., et al. Geodwframe: A framework for guiding the design of geographical dimensional schemas. In *Proceedings of the Data Warehousing and Knowledge Discovery*. Zaragoza, Spain, pp. 26–37, 2004.

Fidalgo, R. N., Times, V. C., Silva, J., et al. Providing multidimensional and geographical integration based on a gdw and metamodels. In *Proceedings of the Brazilian Symposium on Databases*. Brasília, Brazil, pp. 148–162, 2004.

Fonseca, R. L., Fidalgo, R. N., Silva, J., and Times, V. C. Geodwcase: Uma ferramenta para projeto de data warehouses geográficos. In *Proceedings of the Brazilian Symposium on Databases, Demos Session*. João Pessoa, Brazil, 2007.

Ruiz, C. V. and Times, V. C. A taxonomy of solap operators. In *Proceedings of the Brazilian Symposium on Databases*. pp. 151–165, 2009.

Siqueira, T. L. L., de Aguiar Ciferri, C. D., Times, V. C., de Oliveira, A. G., and Ciferri, R. R. The impact of spatial data redundancy on solap query performance. *Journal of the Brazilian Computer Society* 15 (2): 19–34, June, 2009.