GEOPROFILE: 
UML profile for conceptual modeling of geographic databases

Gustavo Breder Sampaio, Filipe Ribeiro Nalon, Jugurta Lisboa-Filho
Departamento de Informática – Universidade Federal de Viçosa (UFV)
36570-000 – Viçosa – MG – Brasil
gustavobreder@gmail.com, frnalon@gmail.com, jugurta@ufv.br

Keywords: UML profile; Conceptual model; GIS; Geographic database

Abstract: After many years of research in the field of conceptual modeling of geographic databases, experts have produced different alternatives of conceptual models. However, still today, there is no consensus on which is the most suitable one for modeling applications of geographic data, which brings up a number of problems for field advancement. A UML Profile allows a structured and precise UML extension, being an excellent solution to standardize domain-specific modeling, as it uses the entire UML infrastructure. This article proposes a UML profile developed specifically for conceptual modeling of geographic databases called GeoProfile. This is not a definite proposal; we view this work as the first step towards the unification of the various existing models, aiming primarily at semantic interoperability.

1 INTRODUCTION

One of the current concerns in software development is to better understand the domain of the problem, about which it is intended to create solutions that meet satisfactorily the real needs of users. To aid in this task, one of the techniques used is the conceptual modeling, which consists in to extract from the real world only those essential elements observed, leaving out implementation aspects.

The process of conceptual modeling allows a better understanding of the system being designed and is performed with the aid of specific modeling languages, which are languages whose syntax and semantics are focused toward the conceptual representation of a system (Booch et al., 2005). Among other, the Unified Modeling Language (UML) has been widely used and accepted by the scientific community and industry, as a tool for design and specification of systems (OMG, 2007).

One area that has currently received much attention includes the geographic applications domain, given its wide range of usefulness to society and the scientific community and whose systems have particular characteristics that need to be taken into account in developing such applications.

Parent et al. (2008) emphasize that the conceptual modeling has several advantages for the design of geographic applications. It allows, for instance, users to express their knowledge on the application using concepts that are closer to them, without the need to use computational expressions.

For the past 20 years, a number of research groups have been studying the requirements for conceptual modeling used in GIS applications (Bédard et al., 2004). A large number of conceptual models specific to this area were proposed. OMT-G (Borges et al., 2001), MADS (Parent et al., 2008), GeoOOA (Kösters et al., 1997), UML-GeoFrame (Lisboa Filho and Iochpe, 2008) and the Perceptory’s model (Bédard, 1999) are important among these models.

Despite the maturity of this research field, to date, there is no consensus among designers and users as to which model best meets the requirements for modeling a geographic database (GeoDB). The lack of a standard model brings up serious problems in the development of the field, as for instance, communication difficulties among different projects. For example, considering CASE tools that support conceptual models specific to GeoDB, data conceptual schemas cannot be migrated between different tools, as it happens with conventional database designs.

These problems would not exist if there were a standard for modeling such applications that incorporated the main features of the existing models. The creation of a UML profile is one option to standardize this type of models. UML profile is a feature that allows for a structured and precise
extension of the UML elements so that it can fit into a specific domain (Fuentes and Vallecillo, 2004).

This paper aimed to initiate the specification of a UML Profile for the conceptual modeling of geographic databases taking into account the requirements imposed on this application domain. Some models in the literature provided the basis for this task.

The remaining of the paper is structured as follows. Section 2 presents the concept of UML profile. Section 3 describes the requirements of GeoDB conceptual modeling, as well as the main current models, while Section 4 details the proposal to the GeoProfile and usage examples. Section 5 presents the final considerations and future work.

2 UML PROFILE

Despite being a general purpose language, which can be used in different application domains, there are situations in which the UML elements are not able to express all the peculiarities of a given domain. Therefore, to prevent the UML became too complex, it was specified as an extensible language (Erikson et al., 2004).

The OMG defines two ways of extending the UML. The first is based on the modification of the UML metamodel, thereby creating a new language, in which the syntax and semantics of the new elements are adapted to the intended domain. The second way is to adapt the UML to specific domains or platforms using the mechanism of profiles. In this second alternative, the elements of language are specialized, but respecting the UML metamodel and maintaining the original semantics of the elements unchanged (Fuentes and Vallecillo, 2004).

In this first form of UML extension, the new language is created using MOF. In the second alternative, the language elements will be specialized by using the extension mechanisms provided by UML, which are: stereotypes, tagged values and constraints.

A UML profile is a set of extension mechanisms grouped in a UML package stereotyped as <profile>$. As mentioned earlier, these mechanisms allow the extension of the syntax and semantics of the UML elements, but without violating the original semantics of UML and, therefore, consistent with MOF.

The idea of extending the UML for specific purposes is not new. UML 1.1 could already easily assign stereotypes and tagged values to model elements. However, the notion of profile was defined to provide a more structured and precise extension (OMG, 2007). UML profile is already adopted as a standard modeling in some domains, such as CORBA architecture (OMG, 2008). Other profiles are in the process of being adopted by the OMG or are being created by private organizations, software companies and research centers.

OMG(2007) emphasized that there is no simple answer to the question of when to create a new metamodel? or when to use the mechanism of profiles? Each alternative has its advantages and disadvantages, but the use of UML profiles provides a better cost-benefit ratio, by utilizing the entire structure of the UML tools and training materials. Fuentes and Vallecillo (2004) mention that the benefits of using UML profiles undoubtedly exceed their limitations.

A UML Profile allows a structured and precise extension of UML constructors to customize UML for a particular domain. A well-specified UML Profile will have direct support of CASE tools. In other words, once the Profile is defined there is no need to implement new CASE tools. Enterprise Architect (EA, 2008) and Rational Software Modeler (RSM, 2008) are examples of CASE tools with support for UML Profiles.

Hence, the development of a UML Profile has proven an excellent method to standardize modeling of specific domains, as it uses the language’s popularity and tools compatible with UML 2.0, favoring standard acceptance and reducing time for training in new languages.

3 CONCEPTUAL MODELING OF GEOGRAPHIC DATABASE

The term Geographic Information Systems (GIS) is applied to systems that perform a computational analysis of geographic data. The main difference between GIS and a conventional information system is the ability of GIS to store both the descriptive attributes and the geometries of different types of spatial data.

GIS use has grown and continues to grow rapidly throughout the world due to advances in hardware and software and the increasingly easier access to these technologies. Among the main components of a GIS is the storage component, which is called geographic database. Its function is to structure and store data in order to enable carrying out the analysis with spatial data. Applications developed with GIS are highly complex and a major problem in developing these applications has been designing the GeoDB (Lisboa Filho et al., 2004).

The classical approach to project database is to divide the process into three stages: conceptual design, logical design and physical design (Elmasri
and Navathe, 2003). In conceptual design, the conceptual database is drawn up on the basis of conceptual models that provide high-level abstraction builders to describe the requirements for application data.

One of the principles of conceptual modeling is that a conceptual schema should only contain the elements of the domain, discarding implementation aspects. The process of database conceptual modeling includes a description and definition of possible contents of data, as well as structures and rules that apply to them (Lisboa Filho and Iochpe, 2008). In the case of GeoDB, the specific nature of geographic information led to the development of specific solutions for modeling spatial data.

Fris-Christensen et al. (2001) describe a survey of requirements for modeling spatial data. These requirements are classified into five groups, as follows: spatiotemporal properties; roles; associations; constraints; and data quality.

Another list of requirements is shown in Lisboa Filho and Iochpe (1999). This study mentions eight groups of requirements, five of which are equivalent to those presented by Fris-Christensen: possibility of modeling phenomena in the field and object view, spatial aspects, spatial relationships, temporal aspects, and quality aspects. The other requirements, not explicitly mentioned in the previous work, are: possibility of differentiating between geographical phenomena and objects without spatial reference; the need to organize the phenomena by theme; and the possibility of modeling phenomena with more than one spatial representation (multiple representations).

Fris-Christensen et al. (2001), compare some models with these requirements to show advantages and disadvantages of each model. One of the conclusions of this study shows the importance of balancing ease-of-use of the model notation with its comprehensiveness. The posed challenge is to balance these two characteristics or improve them, and the development of a standard model provides the basis for data exchange.

The profile proposed in this paper is based on contributions from a number of models existing in the literature, as well as the concepts defined in Goodchild et al. (2007). The models that have contributed most significantly to the GeoProfile development are cited below, but certainly other predecessor models also had their contribution.

The OMT-G (Object Modeling Technique for Geographic Applications) model (Borges et al., 2001) has a rich collection of conceptual constructors, the strong point of which is modeling spatial relationships, including spatial aggregation. The GeoOOA model (Kösters et al., 1997) supports the abstraction of spatial classes, whole-part topological structures, network structures and temporal classes. MADS (Modeling of Application Data with Spatio-temporal Features) (Parent et al., 2008) approaches objects and relationships in its diagram, with structures very similar to the Entity-Relationship model. Its main feature is the orthogonality, in which spatial and temporal characteristics can be added either to objects or attributes or relationships. The Perceptory’s model was the pioneer in the use of pictograms. These pictograms are grouped into the languages Spatial PVL and Temporal PVL (Plug-in for Visual Languages), which allow the addition of spatial-temporal characteristics not only to UML, but also to other visual modeling languages. The UML - GeoFrame model is based on a structured hierarchy of classes that make up the GeoFrame, providing the basic elements present in any geographic database (Lisboa Filho and Iochpe, 1999). The proposal of ISO-191xx Standard (Brodeur and Badard, 2008) differs from the models above mentioned for addressing more the logical level (records) than the conceptual level (abstractions).

Finally, Clementini et al. (1993) formally describe a small set of relationships capable of reproducing all the possible topological relationships that can occur between spatial elements with the representation of point, line or area. Although not proposing a model, this work has considerable importance in the scope of the GeoProfile design. Defining a minimum set of relationships, one eliminates the possible use of two relationships with different names, but having the same meaning. This set includes the following relationships: touch, in, cross, overlap and disjoint.

4 GEOPROFILE

GeoProfile is a UML profile built for the conceptual modeling of geographic databases. According to the proposed methods to guide the construction of a UML Profile (Fuentes e Valecillo, 2004) e (Selic, 2007), two artifacts are generated during profile development: the domain metamodel and the profile itself. While the first is useful to understand the addressed problem, the second presents the extensions received by the UML metaclasses.

In order to check the validity of the GeoProfile specification, this profile has been implemented in Rational Software Modeler (RSM, 2008) which is produced by IBM® and supports UML 2.1. Mechanisms for creation of stereotypes were successfully tested, as well as automatic validation of schemas by checking OCL constraints.
Section 4.1 defines a metamodel for the geographical domain. Section 4.2 proposes a set of stereotypes for the proposed profile. Section 4.3 shows a way to specify additional integrity constraints, and Section 4.4 presents examples of GeoProfile use.

### 4.1 Defining a metamodel for geographical domain

At the beginning of the metamodel specification, elements are identified in a conceptual schema, observing the requirements of this type of conceptual modeling.

The way each considered conceptual model in this proposal (GeoOOA, MADS, UML-GeoFrame, OMT-G and Perceptory’s model) meets the found requirements was examined. The inclusion of the main mechanisms present in each of these models into the GeoProfile allows it to meet most requirements of a geographic database (GeoDB).

Table 1 summarizes the results obtained in the comparative analysis between requirements and conceptual models, but also displays in its last column the models that most influenced GeoProfile construction in each requirement.

Among the discussed conceptual models, the UML-GeoFrame shows the closest organization to a metamodel. GeoFrame is defined in a class hierarchy representing the elements present in a GeoDB. Thus, the metamodel development started from a GeoFrame adaptation (Figure 1).

A GeoDB comprises a number of themes, which is characterized by the metaclass Theme. A theme can be formed by the aggregation of other themes or objects with or without spatial representation, characterized by the classes GeoPhenomenon and ConventionalObj respectively.

When one chooses to associate a spatial representation with objects of a class, it is possible that the phenomenon is perceived in the geographic field view (GeoField) or object view (GeoObject). Depending on the technique used in geographic information acquisition in the field, its representation be selected from six options as described in Goodchild et al. (2007): AdjPolygons, Isolines, TIN, GridOfPoints, GridOfCells or IrregularPoints. Representation of geographic objects can be of the types point, line, polygon or complex (the object geometry consists of other geometries).

<table>
<thead>
<tr>
<th>Models X Requirements</th>
<th>GeoOOA</th>
<th>MADS</th>
<th>OMT-G</th>
<th>Perceptory</th>
<th>UML-GeoFrame</th>
<th>Contribution for GeoProfile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical phenomena and conventional objects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Perceptory</td>
</tr>
<tr>
<td>Field visions and objects</td>
<td>Partial</td>
<td>Partial</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>OMT-G</td>
</tr>
<tr>
<td>Spatial aspects</td>
<td>Partial</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>OMT-G, UML-GeoFrame</td>
</tr>
<tr>
<td>Thematic aspects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>UML-GeoFrame</td>
</tr>
<tr>
<td>Multiple representations</td>
<td>Partial</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>UML-GeoFrame</td>
</tr>
<tr>
<td>Spatial relationships</td>
<td>Partial</td>
<td>Yes</td>
<td>Yes</td>
<td>Partial</td>
<td>Partial</td>
<td>MADS, OMT-G</td>
</tr>
<tr>
<td>Temporal aspects</td>
<td>Partial</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Partial</td>
<td>MADS, Perceptory</td>
</tr>
</tbody>
</table>
To specify multiple representations, it is possible to use more than one stereotype in the same class of the conceptual schema, as in the Perceptory’s model.

The requirements related to the roles and metadata are not considered in the GeoProfile proposal. Despite representing important information relating to spatial data, it is believed that they need not necessarily be demonstrated during the conceptual modeling of a GeoDB.

Topological and composition are the main types of spatial relationships to be represented in a conceptual schema. There was no need to add new constructors to the GeoProfile to characterize composition, as the UML can indicate whether an association is a composition or aggregation. However, it was necessary to add new constructors to model topological relationships, including the capacity to represent networks.

With basis on GeoOOA and OMT-G models, which provide more detailed solutions for network representation, Stempliuc et al. (2009) proposed an extension of GeoFrame to address the requirement. This extension was incorporated into the metamodel.

The other types of topological relationships are directly defined in the creation of stereotypes and OCL constraints. This is because a large number of possible relationships between spatial objects of the type point, line and polygon would overburden the metamodel.

The MADS and Perceptory approaches stand out among temporal aspects. Although they do not consider transaction time, icons added at different positions of the class diagram can indicate that the object’s existence time, its spatial evolution or the evolution of values of certain attributes in that class should be kept in the database. Despite being an good solution, it can visually overload the schema. Another solution adopted by GeoProfile is indicated only whether a class is considered temporary or not, as in the GeoOOA model. In this case, it is implied that both the attributes and spatial data of an object can vary, and these changes must be maintained in the database.

In this way, the metaclass TemporalObject was added to the metamodel. This metaclass has two attributes that characterize temporal information. One of these attributes indicates the temporal type (validity time, transaction time or bitemporal time), whereas the other defines the used temporal primitive type (instant or interval). There are two enumerations (TemporalType and TemporalPrimitive) for the possible values these attributes can assume.
4.2 GeoProfile stereotypes

After creating the domain metamodel, the next step is to extend the UML metaclasses to create the profile itself. Figure 2 illustrates the stereotypes of GeoProfile.

It is worth noting that not all metaclasses of the domain metamodel have a corresponding stereotype, as it happens with Theme and ConventionalObj. Themes can be represented by packages. Classes of conventional objects are, however, modeled by UML classes without addition of stereotypes. Therefore, the UML constructors themselves can reproduce these two concepts.

Another important observation is that some stereotypes are abstract (GeoObject, GeoField, NetObject and Arc). During GeoProfile use, these stereotypes are not available to be used. They are, nevertheless, useful for organizing profile elements, allowing addition of constraints common to all the other stereotypes created as their specialization. For example, a constraint that is common to the stereotypes UnidirectionalArc and BidirectionalArc can be added to Arc.

Geographic phenomena, extending the metaclass Class, are defined in a similar hierarchy to that found in the domain metamodel. The stereotype Network directly extends the metaclass Class, since there is no stereotype defined for representation of conventional objects.

To deal with temporal aspects, the stereotype TemporalObject was added to GeoProfile, as well as two enumerations (TemporalPrimitive and TemporalType). In addition, designers are allowed to indicate that an association between two objects is only valid for one period and this history should be kept in the database.

This is done by simply assigning the stereotype Temporal, which extends the metaclass Association to an association of the schema.

Finally, stereotypes were created to represent the topological relationships that were not considered during drawing up of metamodel. We chose to use the set of five relationships proposed by Clementini et al. (1993), as they are capable of representing any topological relationship between objects of type point, line or polygon. Thus, the stereotypes Touch, In, Cross, Overlap and Disjoint, all extending the metaclass Association, were added.

4.3 OCL constraints

The constraints included in the GeoProfile focus on the validation of the designer’s conceptual schema. Consequently, they always have a stereotype of the GeoProfile as context, as well as being invariants.

Those constraints basically prevent the occurrence of three error types: addition of incompatible stereotypes with a same element, poor network construction and addition of impossible topological relationships between two elements (e.g. Cross relationship between two geographic objects.
with point representation). These three constraints groups were analyzed and a set of OCL expressions was specified. There is no limitation to the inclusion of the stereotype <<TemporalObject>> in classes of the schema or <<Temporal>> in their associations. Because of space limitation, this article describes only one of the OCL constraints as example.

The constraint (a) evaluates the use of incompatible stereotypes. Each class that receives a stereotype of geographic field (context GeographicField) must have all its applied stereotypes captured (getAppliedStereotypes). Stereotypes of the geographic object type are selected from the result using the select method, and the returned set must be empty (isEmpty), since a class cannot have object and field representation at the same time.

```ocl
class GeographicField
  inv: self.getAppliedStereotypes() -> select(s | s.name = 'Point' or s.name = 'Line' or s.name = 'Polygon' or s.name = 'ComplexSpatialObj') -> isEmpty()
```

4.4 Usage examples of GeoProfile

Considering the results obtained after the establishment of the GeoProfile, this section presents examples of conceptual modeling using GeoProfile. To allow a comparison between the GeoProfile and conceptual models that were the basis for its definition, each example also shows the corresponding conceptual schema in the other model.

Figure 3 illustrates part of the conceptual modeling of a system for pollution control in parcel (or plot) of land. The diagrams (a) and (b) display the model developed using the model GeoOOA and GeoProfile, respectively. The parcels have a polygonal representation. A non-geographic object providing information on the owners of each parcel must be stored. In addition, each parcel may contain several pollution controls, which are geographically represented by points. In the association between parcels and points of pollution control, the restriction that each control point must be contained in the area of the parcel with which it is associated is represented in the conceptual schema.

Another example of topological relationship involving parcels (or plots) of land is showed in Figure 4, which compares (a) the MADS model with (b) GeoProfile. In this schema each plot may contain several buildings, and both classes have polygonal representation. Furthermore, a restriction is imposed that the buildings belonging to a particular plot must have their geographical area within the area of the plot. In the case of the topological relationship «In», it may be important for a correct interpretation of the schema, to state which of the objects involved in a particular association must have its geometry contained in the geometry of the other object that participates in the association. In Figure 4-b, roles of the association were used for this purpose. Another option is to indicate the navigability of the association. Both solutions use resources of the UML specification.
5 CONCLUSIONS

This article proved a standard conceptual model for geographic database modeling to be feasible. The existence of several alternative conceptual models of geographical databases prevents users and designers to migrate their projects from a CASE tool to another. Other major problem brought up by the lack of standardization is the difficulty in training designers, since although the models produced for the same purpose; each one has its differences and particularities. Users who are familiar with a model (and its respective CASE tool) show strong resistance to accept a new one.

The use of a UML profile will solve these problems. Besides the wide UML acceptance by software developers, the availability of CASE tools with support for profiles rule out the need for implementing specific tools for a particular model.

A subject for future work is the logical-conceptual transformation of schemas produced with GeoProfile. The existence of logical standards, as defined by OGC and the series ISO 19100, will have a strong link with the level of conceptual modeling. Finally, the great challenge is to make authors of the existing conceptual models contribute to improve the GeoProfile. Moreover, to know the opinion of the users is important, because in many cases the database of a GIS application is designed by then. Thus, it is also important to measure the GeoProfile use’s facility and its learning curve.

Acknowledgements

This project was partially funded by CAPES, FAPEMIG and CNPq / MCT / CT-Info.

REFERENCES


