Using MDA and a UML Profile integrated with international standards to model geographic databases

Filipe Ribeiro Nalon\(^2\), Jugurta Lisboa-Filho\(^1\), Karla A. V. Borges\(^2\),
José Luís Braga\(^1\), Marcus V. A. Andrade\(^1\)

\(^1\)Departamento de Informática – Universidade Federal de Viçosa (UFV)
36.570-000 – Viçosa – MG – Brasil

\(^2\)Prodabel – Empresa de Informática e Informação do Município de Belo Horizonte
Av. Pres. Carlos Luz, 1275 – 31230-000 – Belo Horizonte – MG – Brasil
{frnalon,marcus,zeluis}@dpi.ufv.br, jugurta@ufv.br, karla@pbh.gov.br

Abstract. In the last 20 years, several conceptual data models specific for Geographic Information Systems (GIS) have been proposed. However, so far there isn’t a consensus model, which has generated several problems for the GIS area, such as the lack of interoperability among CASE tools that give support to these models. A UML profile, called GeoProfile, was proposed to standardize the task of geographical data modeling. This article shows the integration of GeoProfile with the international standards of ISO 19100 series, which are addressed to geographical information. This integration is presented through the different abstraction levels of the approach Model Driven Architecture (MDA).

1. Introduction

The activity of software development is a task that requires more and more the use of standardized methodologies and techniques that are widely known. Currently, the main concern of the designer is to understand well the problem domain in order to generate solutions that suit the real necessities of the users.

In order to help in this task of understanding the problem and reducing the system complexity to be developed, the main technique that is used is modeling. A model is a reality simplification [Erikson et al. 2004]. In database design, the construction of models helps to describe the data without having to worry about the implementation details. The Model Driven Architecture (MDA) approach [OMG 2003] enables the development of systems using models in different levels of abstraction.

In the classic databases design [Elmasri and Navathe 2003], the most abstract model level is called conceptual model. The data conceptual modeling is done with languages which syntax and semantics have its focus turned to the conceptual and physical representation of a system [Fuentes and Vallecillo 2004]. Currently, one of the most used and accepted languages is the Unified Modeling Language (UML), which is extensible, in order to attend to some specific domains, using for that, a mechanism called profile.

An application domain that has been calling attention at present is the geographic domain, due to the current availability of the spatial data sets. In the last 20
years, several researches have been done aiming to create or adapt conceptual data models for geographic applications.

The existence of several models has brought a problem for the area, which is the lack of a modeling standard. Tools have been created for the several models and there is difficult to obtain interoperability among the created solutions, thus making it impossible the reuse of solutions in other projects. Besides that, there are certain modeling requirements of geographic applications that some models support while others do not.

For the standardization of these models, a UML profile, called GeoProfile [Lisboa Filho et al. 2010], was proposed. The GeoProfile is a UML profile for conceptual modeling of geographic database (geoDB), which puts together the characteristics of the main existing conceptual models. As an effort for the geographic information standardization, some organizations, such as the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC), have published international standards to help in the construction of standardized geographic applications.

This article describes the integration of the GeoProfile with the standards of the ISO 19100 series, using the different abstraction levels of the MDA approach. Section 2 describes UML profiles. Section 3 describes the MDA approach. Section 4 describes some requirements for geoDB modeling and the standards from the ISO 19100 series which can be used in the geoDB modeling. Section 5 describes the GeoProfile. Section 6 compares the abstraction levels between the GeoProfile and the ISO standards, presenting the correspondence between both the concepts. In Section 7 some final considerations as well as some future works are presented.

2. UML profile

UML is a modeling language which can be used in several application domains [OMG 2007]. However, there are situations in which the UML designers are not able to express all the concepts of certain domains. Thus, as it is mentioned by Erikson et al. (2004), in order to avoid that UML became too complex, its creators made them extensible, that is, it is possible to adapt it to a domain or specific platform, through its extension mechanisms, which are stereotypes, tagged values and constraints.

The set of these extension mechanisms can be grouped in a UML profile. The intent of the UML profile mechanism is to supply a direct path to adapt an existing metamodel with the constructors that are specific to a particular domain, platform or method. The profile mechanism is consistent with the Meta Object Facility (MOF) specification [OMG 2007].

A well-specified UML Profile would 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 2010] and Rational Software Modeler [RSM 2010] are examples of CASE tools that support 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 UML2, favoring standard acceptance and reducing time for training in new languages.

3. Model Driven Architecture (MDA)

With the promise of improving the software development, the Object Management Group (OMG) has adopted the MDA approach, whose main characteristic is the emphasis given to the models. In this approach, the software development process is directed by the modeling activity of the system. A system model is a description using a specific notation. The artifacts produced in MDA are formal models, that is, models that can be understood by computers [Erikson et al. 2004].

In MDA, the system requirements are modeled using a Computation Independent Model (CIM). This model is called domain model or business model and it uses a familiar vocabulary to the domain experts. A CIM does not show details of the systems structure, but the environment in which the system will operate, being useful to understand the problem [OMG 2003].

In the second level of abstraction we find the Platform Independent Model (PIM). This is a model with an abstraction level relatively high and independent from any implementation technology [OMG 2003].

Later, a PIM should be transformed into a Platform Specific Model (PSM). A PSM is customized in order to specify the system in terms of implementation constructors which are available in a specific implementation technology. For instance, a relational database PSM should include terms such as “table”, “column”, “foreign key”, among others. A PIM can be transformed into one or more PSMs. For each specific technology platform a separated PSM is generated. Next step is the transformation of each PSM to source code. This transformation is relatively direct since the PSM is adjusted to a specific technology.

A MDA key element is that the transformations should be automatically executed. Traditionally, the transformations from model to model or from model to code are manually made. In the MDA approach, in the other hand, transformations are executed preferably by tools [Kleppe et al. 2003].

The OMG also provides some ways of transforming models into MDA, one of them is the transformation using UML profiles. A CIM and a PIM can be prepared using a UML profile independent from platform. This model can be then transformed into a PSM using a second profile, of specific platform [OMG 2003].

4. Geographic database modeling

Geographic databases (geoDB) belong to the category of the non conventional databases. Geographic data have, besides the descriptive attributes, a geometric representation in the geographic space; these data are known as geo-spatial or geo-referenced data.

The geoDB modeling holds some particularities that cause the development of specific solutions for this domain. Friis-Christensen et al. (2001) describe a survey of the geographic data modeling requirements, which were classified into five groups: space-temporal properties, roles, associations, constraints and data quality.
Another requirement list is exhibited in Lisboa Filho and Iochpe (1999). In this study, eight groups of requirements are mentioned, five of them equivalent to the ones presented by Friis-Christensen et al. (2001): modeling possibility of the phenomena in the field and object views, spatial aspects, spatial relationships, temporal aspects and quality aspects. The other requirements, which are not explicitly mentioned in the previous paper, are: possibility of differentiation between geographic phenomena and objects without spatial reference, necessity of organizing the phenomena by theme and possibility of modeling of phenomena with more than one spatial representation (multiple representations).

Currently there are several proposals of modeling for geographic data, among the most known are: GeoOOA [Köster et al. 1997], OMT-G [Borges et al. 2001], MADS [Parent et al. 2008], UML-GeoFrame [Lisboa Filho and Iochpe 2008] and Perceptrory’s model [Bédard and Larrivée 2008]. Each of these models presents particular characteristics and try to implement the requirements of geographic applications modeling.

4.1. International standards for geographic information

The efforts for the international standardization in the area of geographic information have been taking place since the last decade through organizations such as ISO and OGC. The Technical Committee ISO/TC 211 is the one responsible for the preparation of the ISO 19100 series, which define the international standards regarding the geographic information field. These standards aim to promote the usage of geographic information in an efficient, effective and economical way, thus contributing to the solution of global problems, such as the humanitarian and ecological problems.

These standards can contribute in several levels of abstraction, from abstract modeling through implementation aspects. In this article some standards related to data models for geographic information, more specifically the ISO 19107 Spatial Schema [ISO/TC211 2003], ISO 19108 Temporal Schema [ISO/TC211 2002] and ISO 19123 Schema for Coverage Geometry and Functions [ISO/TC211 2005] standards, are analyzed.

The ISO 19107 Spatial Schema standard specifies schema to describe and manipulate the spatial characteristics of the geographic features. A feature is an abstraction of a real world phenomenon. This abstraction is a geographic feature if it is associated to a relative localization in the Earth [ISO/TC211 2003]. The standard consists of class diagrams that can be used in a application schema, profiles and implementation specifications. It also defines spatial operations, standards for use in the access, query, management, processing, and data exchange of geographic objects. The ISO 19107 standard defines in details the geometric and topological characteristics that are necessary to describe the geographic features.

The ISO 19108 Temporal Schema standard defines the concepts regarding the temporal characteristics of geographic information, showing how these characteristics are abstracted from the real world. Jensen (1994) considers two kinds of time: the valid time and the transaction time. The first one is the time when a fact is true in the observed reality and it is generated by the user. The second one is the time when a fact is stored in a database from which it can be recovered. This international standard emphasizes the valid time instead of the transaction time. The standard consists of a
class hierarchy that considers the geometric and topological aspects of the temporal characteristics [ISO/TC211 2002].

The ISO 19123 Schema for Coverage Geometry and Function standard, on the other hand, defines a scheme for the spatial characteristics of coverage. Coverage is a feature that has multiple values for each type of attribute and can represent a simple feature or a set of features. They integrate discreet and continuous geographic phenomena [ISO/TC211 2005]. Examples of coverage include raster, TIN, point coverage and polygon coverage. They are used in several specific areas such as, for instance, remote sensing, meteorology, soils and vegetation.

5. GeoProfile – a UML profile for geoDB conceptual modeling

The UML profile proposed for geoDB conceptual modeling, called GeoProfile [Lisboa Filho et al. 2010], puts together the characteristics of the main conceptual data models of the area, previously mentioned, thus seeking to achieve the requirements of geographic applications modeling.

The GeoProfile was specified following the guidelines for specification of UML profiles discussed in Fuentes and Vallecillo (2004) and Selic (2007). The first step was the construction of the domain metamodel [Lisboa Filho et al. 2010], in which the concepts present in the geoDB modeling and the basic requirements were approached.

The way each considered conceptual model in this proposal (GeoOOA, MADS, UML-GeoFrame, OMT-G and Perceptory’s model) meets 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 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.

Table 1. Comparison between requirements and models presented, and major contributions to the GeoProfile

<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>
The second step was to extend the UML metaclasses to create the profile itself. In this step the stereotypes, tagged values as well as the constraints were defined. The GeoProfile stereotypes are shown in Figure 1.

Most of the GeoProfile stereotypes extend the metaclass Class. Both the GeoObject and GeoField stereotypes represent the geographic phenomena perceived in the objects and fields views, respectively. Since these stereotypes were defined as abstracts, as well as the NetworkObj and Arc stereotypes, they will not be included in the schema classes during the usage of the GeoProfile, but their corresponding subclasses will.

To deal with temporal aspects, the TemporalObject stereotype, that also extends the metaclass Class, was included. The two enumerations that were included (TemporalPrimitive and TemporalType) are used to list the possible values that the meta-attributes (tagged values) temporalPrimitive and temporalType may assume, which are: instant and interval.

Besides the extensions to the metaclass Class, extensions to the metaclass Association were included. These extensions had the aim of creating stereotypes to serve the topological relationships, which are: Touch, In, Cross, Overlap and Disjoint. 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.

Besides the stereotypes, some constraints were also added, which are useful for the conceptual schema validation. The constraints were defined using the Object Constraint Language (OCL) and they have as context the created stereotypes. Details about the constraints specification can be obtained in Lisboa Filho et al. (2010).
6. GeoProfile adequacy to ISO standards using MDA

GeoProfile was designed having in mind higher abstraction levels, helping the designers in the first steps of a geoDB project. This abstraction level, in the classical approach of database design, is called conceptual level, in which only the aspects related to the problem domain are taken into account, without dealing with implementation details. In the MDA approach, this more abstract level is the CIM. According to OMG (2003), such model uses vocabulary that is familiar to the domain experts. The GeoProfile also acts as a CIM, since it represents the geoDB in a more abstract way.

These models of higher abstraction levels should be transformed into models of lower levels, enriched with elements of a more technical order until they achieve implementation details. In the classical approach, this transformation is called logical-conceptual mapping. It is what happens, for example, in the transformation of a schema made in the Entity-Relationship Model for the Relational Model. In the MDA approach, on the other hand, a CIM is transformed into a PIM.

The international standards of the ISO 19100 series, which were analyzed in the previous section, act in a lower level of abstraction as a PIM due to the fact that they present some technical details. Despite the fact they are still in a conceptual level and do not present implementation details, these standards are not in the same GeoProfile abstraction level. Table 2 shows the correspondence between the elements of GeoProfile and elements of the ISO 19100 series standards.

Table 2. Variables to be considered on the evaluation of interaction techniques

<table>
<thead>
<tr>
<th>Requirements of BDGeo modeling</th>
<th>GeoProfile</th>
<th>Classes in the international standards</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical objects in the object view</td>
<td>Point</td>
<td>GM_Point</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>Line</td>
<td>GM_Curve</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>Polygon</td>
<td>GM_Surface</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>ComplexSpatialObj</td>
<td>GM_Complex</td>
<td>ISO 19107</td>
</tr>
<tr>
<td>Geographical objects in the field view</td>
<td>TIN</td>
<td>CV_TINCoverage</td>
<td>ISO 19123</td>
</tr>
<tr>
<td></td>
<td>Isolines</td>
<td>CV_SegmentedCurveCoverage</td>
<td>ISO 19123</td>
</tr>
<tr>
<td></td>
<td>AdjPolygons</td>
<td>CV_DiscreteSurfaceCoverage</td>
<td>ISO 19123</td>
</tr>
<tr>
<td></td>
<td>GridOfPoints</td>
<td>CV_DiscreteGridPointCoverage</td>
<td>ISO 19123</td>
</tr>
<tr>
<td></td>
<td>GridOfCells</td>
<td>CV_GridCell</td>
<td>ISO 19123</td>
</tr>
<tr>
<td></td>
<td>IrregularPoint</td>
<td>CV_DiscretePointCoverage</td>
<td>ISO 19123</td>
</tr>
<tr>
<td>Network elements</td>
<td>Node</td>
<td>TP_Node</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>Arc</td>
<td>TP_Edge</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>UnidirectionalArc</td>
<td>TP_DirectedEdge</td>
<td>ISO 19107</td>
</tr>
<tr>
<td></td>
<td>BidirectionalArc</td>
<td>TP_DirectedEdge</td>
<td>ISO 19107</td>
</tr>
<tr>
<td>Temporal objects</td>
<td>Temporal Object</td>
<td>TM_Object</td>
<td>ISO 19108</td>
</tr>
<tr>
<td></td>
<td>Instant</td>
<td>TM_Instant</td>
<td>ISO 19108</td>
</tr>
<tr>
<td></td>
<td>Interval</td>
<td>TM_Period</td>
<td>ISO 19108</td>
</tr>
</tbody>
</table>
The execution of these correspondences can be made as a transformation between a CIM, that is a schema using the GeoProfile, and a PIM, that is a schema enriched with elements from the ISO 19100 series standards. For instance, the phenomena perceived in the objects views modeled with the GeoProfile will be mapped to a PIM enriched with the ISO standards in the following way: the classes that were stereotyped as Point will be mapped to a class that will have an attribute called geometry of GM_Point type. In the ISO 19107 standard, GM_Point is a kind of basic data for objects with 0-dimension. The same will be done with the other three classes, Line, Polygon and ComplexSpatialObj, which will be mapped to a class with geometry attribute of GM_Curve, GM_Surface and GM_Complex types, respectively.

It is important to highlight the fact that these standards offer several ways to model the same geographic information. The correspondence made here was the closest possible to the GeoProfile concepts. The ISO 19100 series standards used above are the ones that come closer to the requirements for geoDB conceptual modeling. For example, the ISO 19107 standard was used to build the correspondence with the GeoProfile stereotypes which represent the geographic objects perceived in the objects view and also with the network elements. The standard is divided into two parts. In the first, which deals with the geometric aspects of geographic information, the correspondence with the geographic objects perceived in the objects view was done. In the second, that deals with the topological aspects, the correspondence was done with the GeoProfile network elements. The ISO 19108 standard was used to build the correspondence with the elements that represent the geographic objects temporal aspects and the ISO 19123 standard was used to make the correspondence with the GeoProfile stereotypes which represent the geographic objects perceived in the fields view.

Regarding the topological relationships, in which the GeoProfile are represented by the Touch, In, Cross, Overlap and Disjoint stereotypes and extend the Association metaclass, in the standard they are dealt with as operations. The ISO 19107 standard is what specifies these operations, which are inherited by all the geometric classes defined in the standard. Therefore, the correspondence with the GeoProfile will not be made, since these operations may be accessed by all the geometric classes, from which the correspondences were made.

6.1. Application example

Figure 2 shows an example of conceptual schema modeled with the GeoProfile. The schema uses a visual notation for the GeoProfile stereotypes. This is a possibility that is suggested by the OMG for UML profiles. In the geoDB modeling visual notation to represent the geographic objects spatial characteristics is used in several models. Some models use other denominations, such as the “pictograms” developed by Bédard and Larivée (2008). In these schemas a visual notation for the stereotypes <<Polygon>> and <<Point>> is used.

The schema shows four classes, three of them with spatial characteristics, which were represented by the GeoProfile stereotypes. In this level of abstraction only the “which” geographic representations and not “how” they were implemented were considered, as well as some basic attributes. Therefore, the schema is a CIM, which uses concepts that are the closest to the end users.
After the construction of the CIM using the GeoProfile, it should be transformed into a PIM, which will take into account some technical details. Figure 3 shows the PIM resulting from this transformation.

The spatial characteristics were transformed into attributes with the types according to the correspondence with the ISO 19100 series standards. For example, the class City, which was modeled using the stereotype <<Polygon>>, in this level of abstraction has a geometry attribute of the GM_Polygon type. The same thing was done with the other classes that possess spatial characteristics.

The next step is transforming the PIM into a PSM, that could be, for example, an object-relational data model. However, this level won’t be shown in this article. One of the main benefits of the MDA approach is the gain in productivity in the development of
software systems through the emphasis given to modeling and to the transformation of models from higher abstraction levels into models of lower abstraction levels in an automated way [Kleppe et al. 2003]. The geoDB project can follow these steps. For example, using tools that support model transformation languages, it will enable the generation of models of lower abstraction levels and, later on, models for specific platforms. An example of model transformation language is Atlas Transformation Language (ATL) [Jouault and Kurtev 2005].

In order to illustrate, a small part of transformation code from the CIM, shown in Figure 2, to the PIM presented in Figure 3, will be shown, using the ATL models transformation language.

The definition of transformations using ATL starts with the transformation module statement as well as the source and target models. The module is defined using the keyword “module” followed by the module name. The keyword “create” indicates the source and target models [Jouault and Kurtev 2005].

After this step, the transformation rules are then defined. Those rules are written using ATL syntax, are saved in files with the extension .atl and can use either the declarative or the imperative style. The code presented in Figure 4 shows one of the transformation rules for the CIM and the PIM previously shown. This rule is responsible for creating the classes that have geographic information, that in this case are represented by the GeoProfile stereotypes and for creating the elements that were not contained in the CIM such as, for example, the geometry attribute, which type will be the correspondent to the ISO standard.

``` ATL
1 rule stereotypedClass{
2  from
3    input : GeoProfile!Class( 
4      not thisModule.emptyGeometry(input.stereotype))
5  to
6    output : ISO!Class( 
7      name <- input.name, 
8      reference <- input.reference ->
9      collect(e | thisModule.getReferences(e)).asSet(),
10     attribute <- input.attribute ->
11     collect(e | thisModule.getAttributes(e)).asSet(),
12     attribute <- id, 
13     attribute <- geometry
14  );
15  id : ISO!Attribute( 
16    name <- 'id' + input.name, 
17    type <- thisModule.integerDataType() 
18  ),
19  geometry : ISO!Attribute( 
20    name <- input.name + 'Geometry',
21    type <- if(thisModule.isPolygon(input.stereotype)) then
22       thisModule.polygonDataType()
23     else
24       thisModule.pointDataType()
25     end if
26  )
27}
```

**Figure 4. Example of an ATL transformation rule.**

With the transformation application, the output model is generated in the XML Metadata Interchange (XMI) format, which is a standard format for the models
exchange and that can be imported by most of the CASE tools with support to the UML2.

7. Final considerations

The GeoProfile development had as main motivation the fact that it can use the UML, together with all its available resources, for example the CASE tools, to conceptually modeling a geographic database. GeoProfile gathers in its definition the main requirements for geographic applications and it uses characteristics of the main existing conceptual data models.

This article has shown the correspondence between the GeoProfile and the ISO 19100 series international standards. The use of standards is very important for the acceptance of the GeoProfile by the scientific community as well as by the geoDB designers.

By using the MDA approach, it was possible to show the difference of abstraction levels between the GeoProfile and the international standards and a possible model transformation, using the ATL. This automation of the transformations constitutes one of the main benefits of the MDA approach.

As future works, we can mention the definition of transformation rules for specific platform models, which have not been dealt with in this article, and also the source code generation, for example, the script for logical specification using SQL.

Acknowledgments

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

References


