Interoperability In Practice: Problems in Semantic Conversion from Current Technology to OpenGIS

Gilberto Câmara[1], Rogério Thomé[2], Ubirajara Freitas[1] and Antônio Miguel Vieira Monteiro[1]

[1]Image Processing Division - DPI[1]
National Institute for Space Research - INPE
P.O.Box 515
12201-027 São José dos Campos , Brazil
email: {gilberto,bira,miguel}@dpi.inpe.br
[2]Geoambiente Sensoriamento Remoto
Rua Anésia N. Matarazzo, 60 – Vila Rubi
12245-160 São José dos Campos , Brazil
email: geoamb@tecsat.com.br

Abstract. This work investigates the practical issue of mapping existing GIS to the OpenGIS standards. We describe the data models used in three systems (MGE, ARC/INFO and SPRING) and analyse the problems involved when mapping them to OpenGIS. Our conclusion is that the OpenGIS standard has not been defined in a formal and unequivocal way, and therefore, there are indefinitions and competing alternatives for mapping existing GIS systems into the proposed standard.

1. Introduction

The issue of interoperability is currently subject to substantial efforts, both from an academic and an industry perspective. In this issue, academia and industry have taken different, if complementary perspectives: whereas there is a major effort in the industry towards a consensus-based solution [1], researchers have concentrated efforts in theoretical issues, such as abstract models for semantic interoperability [2][3].

This work investigates the problem from a different perspective, aiming to describe and analyse the practical barriers to interoperability. We start from the pragmatic consideration that most end-users, which will eventually build interoperability frameworks, already have a large geographical database, organised around a commercial system and based on a proprietary data model. These users will probably soon face a decision as regards the introduction of technology which will support the OpenGIS standards, and will probably have a choice between different commercial implementations and migration paths from their existing environment. Therefore, in our assessment of interoperabilty in practice, we aim to understand issues such as:

- Is the OpenGIS proposal a truly generic model, which is able to provide semantic equivalents to concepts on existing proprietary data models ?
- What do real-world systems teach us about the problems of semantic interoperability and possible limitations of the OpenGIS approach ?
- How effective and easy will be the migration from proprietary frameworks to environments such as OpenGIS ?
- What sort of tools would simplify the migration from existing GIS to the OpenGIS framework ?
- What lessons can be learned, from the academic perspective to interoperability, from considering the interoperability challenges to today's technology ?

In order to address these questions, we have examined three existing GIS solutions: MGE [4], ARC/INFO [5] and SPRING [6]. We have chosen these systems because the first two are representative of existing technology and claim a significant proportion of GIS market share. The choice of SPRING is based on two reasons: this system has been developed by INPE, and is therefore well known to the authors and it represents an attempt to devise a conceptual model for spatial data, which explicitly includes the notions of *fields* and *objects.*

The work is divided in three parts. In Section 2, we briefly examine the semantic models used by MGE, ARC/INFO and SPRING. In Section 3, we describe a possible mapping between these systems and OpenGIS, which could be used in real-world migration to OpenGIS. The work concludes with Section 4, where we consider the theoretical and practical consequences of our findings.

2. Semantic Models of Existing Systems

The semantic models of existing systems are a clear demonstration of the barriers faced by the interoperability issue in GIS. In the vast majority of cases, their semantic models have been derived based on practical implementation considerations, related to the data structures used for representing geographical data on a computer.

In order to represent the data model of existing systems and of the OpenGIS model, we have used Rumbaugh's [7] OMT diagrams, which capture the notions of *specialisation* ("is-a") and *aggregation* ("has-a").

2.1 A Generic Reference Model for Geographical Data

Our working hypoteses for comparing the semantic models of different systems is that, given the great differences between them, a generic reference model is necessary to establish a common base into which the system concepts will be refferred to. Modelling in each system will be first expressed in terms of these concepts, for later conversion to the OpenGIS model.

We will use an abstract formulation as a reference for comparing the semantic concepts: the notions of *fields* and *objects* [8]. The *field model* views the geographical reality as a set of spatial distributions over the geographical space. Features such as topography, vegetation maps and LANDSAT images are modelled as fields. The *object model* represents the world as a surface occupied by discrete, identifiable entities, with a geographical location (with possible multiple geometric representation) and descriptive attributes. Human-built features, such as roads and buildings, are typically modelled as objects. For a more detailed discussion on these issues, the reader should refer to [9] [10] and [11].

In what follows, we will consider the following definitions:

- A *geographical field* is defined by a relation $f = [R, V, l]$, where R is a geographical region, V a set of attributes and $l: R \rightarrow V$ is a mapping between points in R and values in V (In OpenGIS, l is called the *coverage function*). When V is a finite denumerable set, we call such fields *thematic*, and when V is the set of real values, we call such fields *numeric*, representing - respectively - thematic maps and digital terrain models.
- Given a set of geographical regions $R_1,...R_n$ and a set of attributes $A_1,...A_n$ with domains $D(A_1),...,D(A_n)$, a *geographical object* is defined by a relation $[a_1,...a_n, S_1,...,S_m]$, where a_i are its descriptive attribute ($a_i \varepsilon D(A_i)$) and S_i its geographical locations ($S_i \subseteq R_i$).

2.2 The MGE Data Model

The MGE ("Modular GIS Environment") data model uses three main notions: CATEGORIES[1], FEATURE CLASSES and FEATURES[4]. A geographic element is represented as a FEATURE. Features are instances of FEATURE TYPES, which may, in turn, be further grouped into CATEGORIES, as shown in Figure 1. Each FEATURE TYPE is associated to an ATTRIBUTE TABLE.

MGE does not include an abstract notion of *fields*, and uses the concepts of REGULAR GRID, TIN and IMAGE to deal with raster representations of geographical reality.

Vector representations of a THEMATIC FIELD use the notions of CATEGORIES and FEATURE TYPES. There are two options to represent a thematic map: (a) the thematic field may be considered as part of a single FEATURE TYPE, and its values stored as attributes in the ATTRIBUTE TABLE; or (b) the thematic field may be considered as a higher-level entity (a CATEGORY) and each of its values (thematic classes) is modelled as a different FEATURE TYPE.

Raster representations of thematic fields are modelled as REGULAR GRIDS and constitute a separate entity from their vector representation. Numerical fields (DTMs) are stored separately as TIN or REGULAR GRID, depending on the chosen representation.

In resume, the MGE data model can be considered as an *object-centered* model, with a *one-level hierarchy*; fields are modelled directly through one of their representations.

---

[1] In what follows, semantic constructs of the different data models are marked in SMALLCAPS.

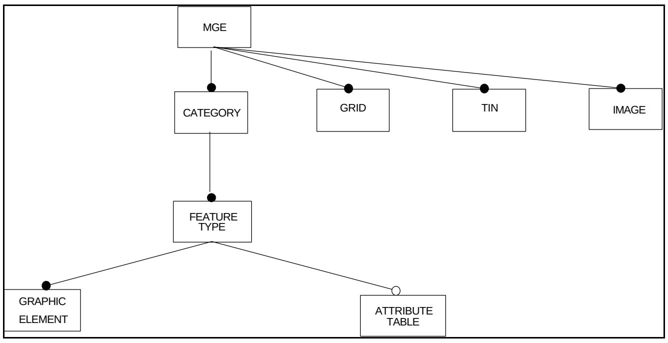

Fig. 1. MGE's data model

2.3 The ARC/INFO Data Model

The ARC/INFO data model [5] has four basic concepts*:* COVERAGE, GRID, TIN and ATTRIBUTE TABLE. A COVERAGE is a vector representation of geographical data, associated to an ATTRIBUTE TABLE which describe the map elements (points, arcs or polygons). In this model, the notion of *object* (or *feature*) does not exist explicitly; objects are implemented as rows of the ATTRIBUTE TABLE, which is required to maintain a unique index.

Thematic fields have two possible representations: their vector representation is mapped to a COVERAGE, where one or more fields in the ATTRIBUTE TABLE indicate the attributes associated to each geographical location. The raster representation of thematic fields uses an INTEGER GRID, associated to an ATTRIBUTE TABLE, which indicates, for each value in the grid, the corresponding attributes. Numerical fields can be mapped either as an "FLOATING-POINT GRID" or as a triangular mesh (TIN).

In resume, the ARC/INFO data model is *representation-oriented*: instead of describing the world in terms of objects and fields, it allows the user to define and manipulate geometrical representations. The user will therefore be responsible for externally defining the abstract entities and for mapping those entities to the most appropriate representation. The ARC/INFO data model is shown in Figure 3.

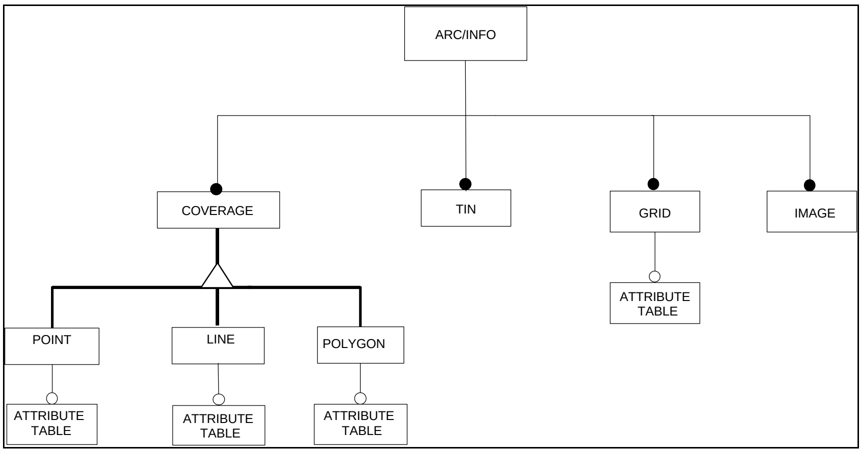

Fig. 2. ARC/INFO's data model.

2.4 SPRING Data Model

SPRING is a public-domain GIS developed by INPE [6], available on the Internet (http://www.dpi.inpe.br/spring), whose data model (shown in Figure 3) deals with the two basic classes: GEO-FIELD and GEO-OBJECT. The definitions of GEO-FIELD and GEO-OBJECT (and the specialisation of GEO-FIELDS into THEMATIC and NUMERIC classes) correspond to the generic data model description given in Section 2.1. Moreover, the model distinguishes between these abstract definitions and their geometrical representations, since:

- GEO-FIELDS can be associated simultaneously to vector and raster representations. THEMATIC GEO-FIELDS can be represented as a vector (polygon map) or as raster (integer grids). NUMERIC GEO-FIELDS can be represented as vectors (contour maps, samples or TINs) or in raster format (floating-point grids). In other words, the relation between a GEO-FIELD and its representation is one of aggregation ("has-a") and not a specialisation ("is-a").
- GEO-OBJECTS can be mapped into different geometrical vector representations, with different topologies (polygon maps, networks and point maps). For this purpose, SPRING uses the auxiliary concept of GEO-OBJECT MAP, as explained below.

Since most applications in GIS do not deal with isolated elements in space, it is convenient to store the graphical representation of geo-objects together with its neighbours. For example, the parcels of the same city borough are stored and analysed together. These features lead us to introduce the concept of a GEO-OBJECT MAP, which groups together geo-objects for a given cartographic projection and geographical region. Therefore, the representations for *geo-objects* are maintained in instances of the class GEO-OBJECT MAP. The relation between GEO-OBJECTS and GEO-OBJECT MAP is one of "is-represented-by".

Use of the model concepts has enabled the design of an user interface and a query and manipulation language for SPRING which allows manipulation of geographical data at an abstract level. [12]

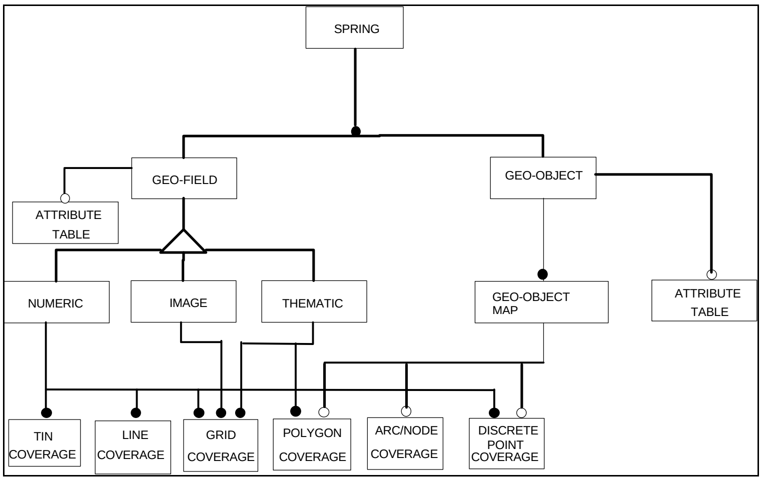

Fig. 3. SPRING's data model.

3. Mapping into OpenGIS Semantic Model

3.1 The OpenGIS Semantic Model

The OpenGIS model [1] is based on an abstract class (FEATURE) which has two specialisations: FEATURE WITH GEOMETRY and COVERAGE. The definition of FEATURE WITH GEOMETRY supports the idea of *geo-object* (as given in Section 2.1) and allows for complex geometrical representations to be associated to the same feature and for different features to share the same geometrical representation. The locational support for each feature is modelled by the ideas of OpenGIS GEOMETRY and CORNER, whereby geometry structures (such as lines, points and polygons) which describe the geographical locations of the feature are related to reference extent in a given projection.

In OpenGIS, COVERAGES are metaphors of continuous phenomena over the Earth's surface, whose spatial domain is a *c-function*, which associates each location to the spatial phenomenon being represented. This definition is analogous to the notion of *geo-fields*, as described in our reference model (section 2.1). A COVERAGE may be specialised into one of several geometrical representations, including*: IMAGE, GRID, LINES* and TIN. Figure 4 illustrates the OpenGIS semantic model.

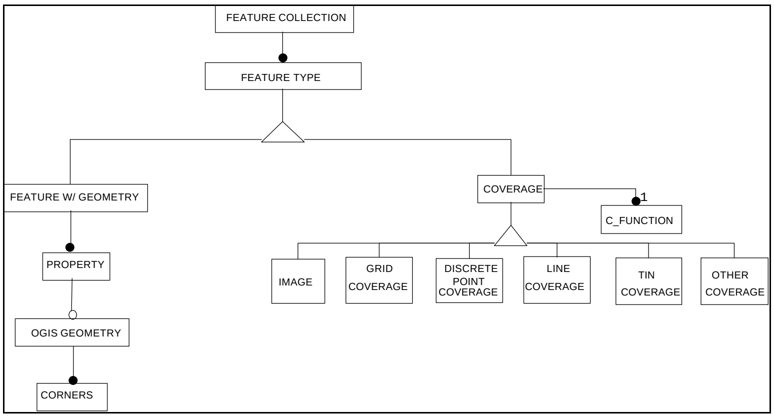

Fig. 4. OpenGIS data model.

OpenGIS is an evolving standard and, as of September 1998, the consortium had not published a definitive position on the notion of FEATURE COLLECTIONS. This concept would allow the expression of complex features, such as feature hierarchies.

It should be noted that there is a semantic mismatch between the definitions of FEATURE WITH GEOMETRY and COVERAGE in OpenGIS. The former definition is abstract and generic, and its relation to the geometric representations is one of aggregation ("a feature has many geometric representations"). The latter concept is directly related to the geometric representation, by *specialisation* ("a grid coverage is a coverage").

In resume, the OpenGIS model approximates the notions of *fields* and *objects*, but abstains from formal definitions and from the use of these theoretical notions, and adopts industry-established terms, such as *feature* and *coverage*. In our opinion, this choice, instead of simplifying the migration of existing systems, is likely to cause significant problems in the adoption of OpenGIS in real-life situations, as discussed in later sections.

3.2 Mapping Existing Databases into OpenGIS

The scenario envisaged in this paper is a situation where an institution, which has an established geographical data base in a proprietary format, would like to migrate to the OpenGIS model. In this process, they will need to find approximate equivalents in the OpenGIS concept to their semantic models.

One important consideration here is that the mapping should be able to benefit, as much as possible, from the features and tools provided by OpenGIS. This will require an extra abstraction level which is not present in most semantic models of existing systems: that of establishing whether the data represent *objects* or *fields*.

3.3 Mapping MGE into OpenGIS

The mapping of MGE concepts into OpenGIS definitions faces a meaningful issue: the two-level feature hierarchy of MGE semantic model (CATEGORIES and FEATURE TYPES) requires the concept of FEATURE COLLECTIONS in OpenGIS to be fully defined; otherwise, a significant part of MGE's semantic richness will be lost in the translation.

In the case of *objects*, they are defined in MGE using the CATEGORY-FEATURE TYPE hierarchy, which would require an equivalent in OpenGIS, namely, the FEATURE COLLECTIONS-FEATURE WITH GEOMETRY hierarchy.

The issue is further complicated in the case of *thematic fields*. As discussed in Section 2, there are two possible ways of mapping vector representations of thematic fields in MGE: (a) using the CATEGORY-FEATURE TYPE hierarchy, where the CATEGORY is specialised into the type of thematic map (e.g. "Land Cover") and there are as many FEATURE CLASSes as different themes in the map (e.g. "Urban Area", "Forest", "Agriculture"); (b) collapsing the CATEGORY-FEATURE TYPE notions into a single concept (e.g. "Vegetation") and using the ATTRIBUTE TABLE to store the values of the theme associated to each geographical area. When mapping to OpenGIS, situation (a) requires the associate notion of FEATURE COLLECTION in OpenGIS for a direct mapping to take place, and situation (b) is best handled by using the GEOMETRY COVERAGE definition in OpenGIS. This is an ambiguous situation, caused both by MGE's way of modelling fields and by OpenGIS' choice of avoiding formal definitions and terms and opting instead for established (but loosely defined) industry terminology.

Numeric fields in MGE (represented as TIN or GRID) are mapped in a straightforward fashion to OpenGIS' GRID COVERAGE and TIN COVERAGE.

3.4 Mapping ARC/INFO to OpenGIS

When mapping ARC/INFO to OpenGIS, the user will first have to define whether the data being mapped refer to *fields* or to *objects*. Since ARC/INFO does not provide an explicit way of representing objects (they are implemented as unique indexes of the ATTRIBUTE TABLE), such a translation cannot be automatic but will require user intervention.

The issue is further complicated by OpenGIS's choice of the terminology COVERAGE to refer (loosely) to *fields*. Some users will be tempted to automatically map ARC/INFO's COVERAGEs into OpenGIS' COVERAGEs, when in fact these are not exact equivalents. In the case of *objects*, ARC/INFO's COVERAGEs are best mapped into OpenGIS's FEATURE WITH GEOMETRY concept, in order to benefit from the query functions defined by OpenGIS (which include topological operators).

Thematic fields in ARC/INFO are mapped directly to OpenGIS. Vector representations of such maps (which are ARC/INFO COVERAGEs) can be translated to OpenGIS LINE COVERAGEs, and raster representations (which are ARC/INFO GRIDs) are mapped into OpenGIS GRID COVERAGEs. Numeric fields are mapped directly to their OpenGIS equivalents (ARC/INFO’s GRIDs and TINs to OpenGIS’s TIN COVERAGE and GRID COVERAGE).

3.5 Mapping SPRING to OpenGIS

Since the SPRING data model is based on the notions of fields and objects, its mapping into OpenGIS is somewhat simplified. OBJECTS in SPRING correspond directly to FEATURES WITH GEOMETRY in OpenGIS. The notion of object maps in SPRING can be mapped to the concepts of OPENGIS GEOMETRY and CORNERS in the OpenGIS model.

In the case of fields, the situation is more complicated. As discussed earlier, the relation of an OpenGIS COVERAGE to its subtypes is one of specialisation. In SPRING, THEMATIC and NUMERIC field can have multiple representations (raster and vector), a notion which is consistent with the abstract definition of a field. Therefore, SPRING THEMATIC data will be mapped to many OpenGIS COVERAGES, depending on how many representations are associated with it. As a consequence, an important part of SPRING's semantic expressiveness will be lost in the process of translation to OpenGIS.

4. Conclusion: Practical Challenges to Interoperability

The main conclusions of the above discussion on the issues of mapping existing systems to OpenGIS are:

- Some systems have semantic models which are richer in content then OpenGIS one (e.g. *object* definition in MGE and *field* definition in SPRING).
- The use of industry-established terminology in OpenGIS is a mixed blessing. Instead of simplifying the migration process, it may rather be a source of misunderstanding (e.g., the mapping of ARC/INFO COVERAGES to OpenGIS FEATURE WITH GEOMETRY).

In each case examined, there were different mapping alternatives from the system to OpenGIS, in some circumstances, which indicate that automatic migration to OpenGIS is not a recommended option and that a higher level of semantic modelling is needed *before* the actual mapping to OpenGIS takes place. This higher level of modelling would be an abstract description in terms of fields and objects (or a more sophisticated approach), along the lines of [2].

In our opinion, one of the main sources of the problems we have described is the choice of the OpenGIS consortium to use industry terminology, which is already content-rich and are associated by the users with existing semantic concepts (FEATURE is already used in MGE and COVERAGE in ARC/INFO). Had OpenGIS chosen to describe its concepts in more abstract and theoretical terminology, some of these problems might have been avoided.

Another problem is the semantic mismatch between the notions of FEATURE WITH GEOMETRY and COVERAGE in OpenGIS, the first being an abstract definition and the second, directly linked to geometric representation. This could be improved if the OpenGIS definition of COVERAGE be changed to an abstract one, where its relation to the representations is one of *aggregation*.

In conclusion, the analysis we have conducted has lead us to believe that there will be major challenges in practice to achieve interoperability, even with the adoption of the extensive work which is being pursued by the OpenGIS consortium. It also calls for theoretical work to be carried out regarding the issue of deriving rich semantic models, which could provide a general framework into which existing semantic models could be mapped, without loss of content and allow the later conversion to other models.

References

1. OGC - OpenGIS Consortium. *Technical Specifications***.** [online] <http://www.OpenGIS.org/techno/specs.htm>. Mar. 1998C.
2. Yuan, M. "Development of a Global Conceptual Schema for Interoperable Geographic Information"**.** *International Conference and Workshop on Interoperating Geographical Information Systems.* Santa Barbara, UCSB, 1997.
3. Gahegan, M. "Accounting for the Semantic Differences between Various Geographic Information Systems". *International Conference and Workshop on Interoperating Geographical Information Systems.* Santa Barbara, UCSB, 1997.
4. INTERGRAPH. *GIS-The MGE Way: Intergraph White Paper*. Huntsville, Ala, Intergraph Corporation, 1995. <http://www.ingr.com/mge/mgegis.htm>.
5. ESRI. *Arc/Info Data Management Concepts, data models, database design, and storage***.** Redlands, CA, Environmental Systems Research Institute, 1994.
6. CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U. M.; GARRIDO, J. "SPRING: Integrating Remote Sensing and Gis by Object-Oriented Data Modelling". *Computers and Graphics*, vol.20, n.3, 1996.
7. Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, B., Lorenson, W. *Object-Oriented Modeling and Design*. Prentice-Hall, 1991.
8. GOODCHILD, M. Geographical data modeling. *Computers & Geosciences*, 1992, *18(4): 401-408*, 1992b.
9. BURROUGH,P.A.; FRANK, A. *Geographic Objects with Indeterminate Boundaries*. London, Taylor and Francis, 1996 (GISDATA Series).
10. COUCLELIS,H. "People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS". In: Frank,A.U.; Campari, I. and Formentini, U. (eds). *Theories and Methods of Spatial-Temporal Reasoning in Geographic Space*. Berlin, Springer, Lecture Notes in Computer Science, n. 639, pp. 65-67, 1992.
11. CÂMARA, G.; FREITAS, U.; SOUZA, R.C.M., CASANOVA, M.A.; HEMERLY, A.; MEDEIROS, C.B., "A Model to Cultivate Objects and Manipulate Fields". In: *Second ACM Workshop on Advances in Geographic Information Systems,* Proceedings, ACM, Gaithersburg, MD., 1994., pp. 20-27.
12. Câmara, G.; Freitas, U.; Cordeiro, J.P.C. Barbosa, C.C.F., Casanova, M.A., "A Query and Manipulation Language for Integrating Fields And Objects in a GIS". Submitted to *GeoInformatica.*
13. OGC - OpenGIS Consortium. *Topic 5: The OpenGIS Feature***.** <http://www.OpenGIS.org/techno/specs.htm>. Mar. 1998.
14. OGC - OpenGIS Consortium. *Topic 6: The Coverage Type and Its Subtypes***.** <http://www.OpenGIS.org/techno/specs.htm>. Mar. 1998