A Methodology for GIS Databases Integration

Julia Celia Mercedes Strauch [(1,2)]   Jano Moreira de Souza[(2)]   Marta Mattoso[(2)]

[(1)]*EMBRAPA/CNPS,*
Brazilian Agricultural Research Corporation

[(2)]*Computer Science Department,*
Federal University of Rio de Janeiro

{julia.jano,marta}@cos.ufrj.br

Abstract

The effective development of a Multidatabase (MDB) environment for integration of independent Geographic Information System (GIS) databases, requires a methodology to guide the users and Database Administrators, as well as the provision of a foundation for the development of tools to help the integration. The existing methodologies do not consider characteristics of the cartographic representation in the schemas nor do they take into account the semantic richness of geographic data. They are also limited to a given application domain or to specific information communities. This work presents MMultiGIS, a methodology for the integration of heterogeneous and distributed GIS databases. It has been developed for the MultiGIS System, which aims at providing semantic interoperability among GIS databases. MMultiGIS disciplines the activities during schema integration taking into account the spatial context, the application areas and the semantic and syntax of database schemas. This methodology has wide application, contemplates all the characteristics of representation of geographic data and may be applied to any system that proposes an open federation of GIS.

1. Introduction

The state of the art of database technology for sharing data in conventional databases has been to substitute closed environments by open solutions, characterised by different platforms, applications and databases that interoperate in real time. These solutions are found under several denominations: multidatabases (MDB), mediators, federated databases and interoperable systems [2], [4], [5], [16], [21]. They have the objective of preserving the existent systems, maintaining its autonomy, through solutions that possesses means to access and share the existing data.

The extension of these solutions for GIS provides the effective sharing of geographic data. These solutions offer an environment that allows two or more heterogeneous GIS to interoperate with one another, and exchange geographic data and services, independent of the structure and format of the data, of the data model, of the modelling, of the application, and of the platform.

The development of this environment presupposes a distributed architecture that encapsulates the involved GIS and that enables the interconnectivity among multiple databases. In the literature we find several interoperable architectures [7], [8], [9], [10], [20], [22] that could be used. However, most of then are restricted to pre-established information communities that in general have the same spatial and application context. Therefore, considering that GIS are: **i)** stable systems, reluctant to the evolution of their schemas, due to the high cost of the data, and **ii)** the GIS users belong to specific application areas and often lack the knowledge about other areas with which they may have to interact, we specified the MultiGIS architecture that provides semantic interoperability using a global schema.

To accomplish an effective sharing of geographic data through this architecture it is necessary to adopt a methodology that disciplines activities that accommodate the local schemas into a global one, and maps distributed and heterogeneous resources into global applications. The adoption of a conventional database integration methodology [3], [6], [17], [18] does not lead to a solution for GIS integration. They do not consider the graphic aspects for the representation of schemas nor the diversity and richness of semantic representation of the geographic data stored by the GIS.

This work presents the MMultiGIS methodology to be incorporated into the MultiGIS System. This methodology, currently in a test phase, supports a modelling that integrates the database schema of different information communities of GIS into a global schema. It proposes a set of activities that capture the semantics of the GIS database schemas, to point out their differences and similarities.

Related proposals are found in the literature for integration of GIS databases. Aguiar [1] proposes a

solution to integrate urban applications of GIS using a data model proposed by Pires [13]. Her methodology emphasises the mapping of the local data models into a global data model and considers the semantics of the geographic data. However, it was developed for a restricted class of schemas, in the same application domain and does not consider the spatial aspects, the differences of application areas nor the graphic representation of the objects as a whole. Parent *et al.* [12] propose a methodology for schema integration of GIS databases. This methodology is composed by three phases: schema preparation, correspondence investigation and integration. In the first phase, the schemas are described and enriched with complementary information. This methodology also emphasises the main types of semantic and structural conflicts that can be found among the GIS database schemas. In spite of these positives characteristics it is difficult to apply, because it does not take into account the similarities among schemas that have the same application area. Yaser [22] proposes a solution for interoperability among GIS. The author does not use a common data model to describe the local schemas. He uses a specification to describe the syntax and extends it to describe the semantics of the geographic objects. This description allows the user to consider the proximity of the application areas during the query process. Thus, Yaser contemplates the application context, but semantic context conflicts are left to the users. In general, the three methodologies apply to specific architectures and applications. They do not contemplate many semantic aspects of the geographic data stored by the GIS databases, such as, localisation, cartographic parameters, application area and semantic and syntactic representation of the schema.

MMultiGIS is different from the above proposals due to its amplitude, since its objective is to assist the construction of an open federation, where any schema, independent of the requirements of modelling and the application, can at any time join or leave the federation. MMultiGIS considers the aspects involved in the whole process of constructing a GIS database.

This work is organised as follows. Section 2 presents the elements used by this methodology, section 3 describes the five steps of the methodology, and section 4 concludes this work.

2. Elements used by MMultiGIS

MMultiGIS adopts a Common Data Model (CDM) that specifies the necessary metadata to describe the export schemas, uses a thesaurus to reduce the subjectivity of the semantic description and proposes a simpler data model for the external schema (Figure 1) based on *Themes* and *Regions*.
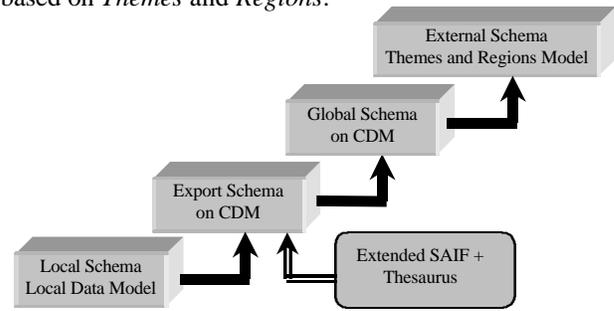

Figure 1. Schema used by MMultiGIS

2.1. Common Data Model

The CDM represents the information derived from the heterogeneous local schemas in a neutral way, helping the administrator of the MultiGIS environment in understanding and interpreting explicit semantic information, as well as to identify the existing similarities and conflicts. The CDM chosen is based on an extension of SAIF [15], a format for data interchange, independent of hardware/software and of application area.

SAIF uses an object oriented data model (Figure 1) to represent a set of basic classes to describe the content of the GIS databases and also offers an object definition language with two components: *Class Syntax Notation* (CSN) and *Object Syntax Notation* (OSN).
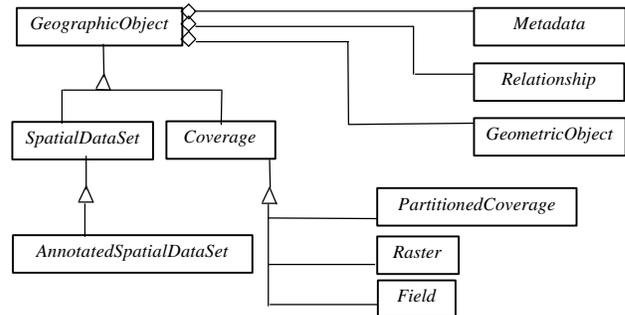

Figure 2. SAIF classes

Thus, this methodology uses SAIF's Metadata class, which has the following subclasses: *SpatialReferencing*, *TemporalReference*, *GeneralLocation*, *Quality*, *Lineage*, *Source*, *Production Description* and *UpdateOperation*. However, to assist the integration, the MMultiGIS specialises the *GeneralLocation* and *Source* classes further inserting properties as shown in Table 1.

SAIF SPECIALISATION	MMULTIGIS PROPERTIES
GeneralLocationMM	*GeographicRegion*

SAIF SPECIALISATION	MMULTIGIS PROPERTIES
SourceMM	*DBInfo.queryLanguage*
	DBInfo.operationSystem
	DBInfo.hardwareType
	DBInfo.platformName
	Project.projectObjective
	Project.applicationArea

Table 1. MMultiGIS extensions to SAIF

The *GeneralLocationMM.GeographicRegion* property measures the degree of spatial relationship with other bases of the federation. The property extensions for the class *SourceMM.DBInfo* describe the location of the database in the network and the characteristics of the local schema management. The property extensions to the class *SourceMM.Project* aims to identify the schemas that may have some degree of semantic proximity. All the above properties are of the type *Enumeration*.

2.2. Thesaurus

A great part of the database knowledge is acquired during the modelling process and this knowledge is subjective. To handle this subjectivity and to capture the implicit semantics in the schemas the translation process uses a vocabulary. This vocabulary offers a group of fixed terms to express the characteristics of the application domain, and the semantic associated to the description of each element in the intentional level.

To describe the characteristics of the application domain we use the domain analysis for GIS applications proposed by Ramirez [14]; and to describe each element in the intentional level we use a thesaurus.

When describing each element of the schema using the data model of SAIF the user selects the terms that best describes the element. These terms complement the descriptions of the elements in the schemas and help to identify conflicts and establish mappings.

2.3. An external schema for geographic data

The MultiGIS System external schema offers a common perspective of the available data in the federation, regardless of the data location, to understand the schemas and the query languages of each local system. This schema presents only one definition that comprises all the incomplete and heterogeneous representations of the GIS databases schemas.

The use of the SAIF data model also as a external schema would add complexity to represent the data of the federation, for it is not a model that offers enough simplicity for the non-specialist user to understand and use its concepts. Therefore the methodology restricts the use of SAIF model to the mapping and the integration

process creating a global schema. In the external it uses a higher-level model.

For the establishment of this higher-level model it was taken into consideration the spatial and/or thematic characteristics present in GIS database queries. Basically it should allow the user to answer two questions:

1. What themes are available in the federation, and how are they represented?
2. What areas are treated in the federation, and how do they relate to the themes?

Thus, the methodology for the MultiGIS System proposes an object oriented modelling formed by two abstract classes to represent the geographic objects in terms of the geographic area and the thematic aspects of the reality that they represent: *Themes* and *Regions*. This data model presents two functionalities: **i)** to present the data, and their relationships, which are available to the federation users; and **ii)** to encapsulate the necessary functionality to locate the geographic objects and their mappings and then to translate a query to the language of the local database.

3. MMultiGIS steps

The methodology establishes five well-defined steps that apply a sequence of operations to help the integration: **i)** *pre-integration*, **ii)** *schema analysis*, **iii)** *schema conformation*/*integration*, **v)** *external schema creation*, and **vi)** *validation of the integration*.
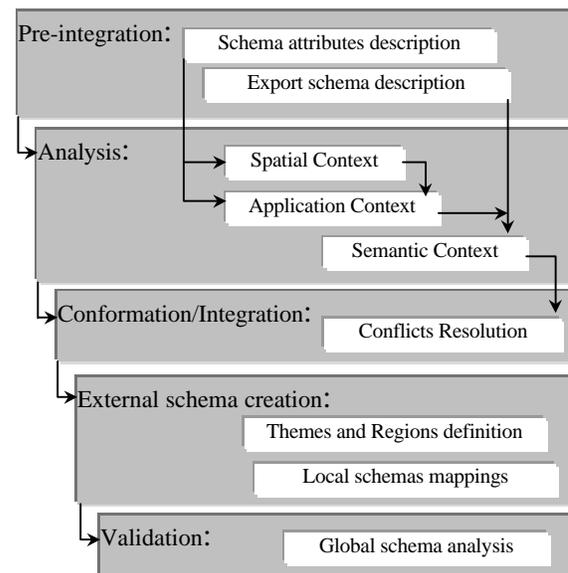

Figure 3. MMultiGIS

3.1. Pre-integration

The first step of MMultiGIS consists of representing the GIS database schemas using CDM. However, the methodology decomposes this description into two levels: **i)** *description of the schema attributes* and **ii)** *description of the export schema.*

3.1.1. Description of the schema attributes. This step describes the schema attributes using SAIF's OSN with the *Metadata* subclasses (Figure 2). The schema attributes represent the implicit attributes in the construction of the GIS database. These attributes gather information on the application domain, which are not parts of the schema modelling.

For each schema, the local administrator creates a specialisation of the SAIF *GeographicObject* class named **Generic Geographic Object** (GGO) class. Then the *Metadata* subclasses are created instantiating a set of objects from the *Metadata* subclasses. To represent the local schemas, subclasses of GGO will be created.

The values of the *Metadata* attributes are the same for almost all the sub-objects that will inherit properties from GGO. To inherit the *Metadata* values, SAIF defines some attributes through a default clause and that is the case of *Metadata* attributes. However, some sub-objects may have specific *Metadata* properties such as *Quality*, *Lineage*, *ProductDescription* and *UpdateOperation*. In this case these sub-objects must create new *Metadata* attributes that will replace the inherit ones.

3.1.2. Description of the export schema. In this step the modelling of each local schema is translated to CDM using the SAIF's CSN. However, the rules of schema creation in the SAIF data model are not appropriate for the local schema translation. These rules do not represent the schema appropriately, since they create classes with new denominations to match the specialisation of the geographic object, which are not foreseen in the schema. This leads to the replication of classes of each schema increasing the complexity in the comparison process.

Each different GIS offers its own data model with its associated structures to represent objects. The description of the local data schemas involves the semantic reconstruction and mapping of a native schema in the GIS data model to an object oriented schema in SAIF. This requires the identification of the best SAIF classes to describe each element in the GIS data model. Thus, this must be performed in agreement with the mappings ($\Rightarrow$) of the elements to SAIF, presented below, in that order:

M1: Entities without graphic attributes $\Rightarrow$ specialised classes of *AbstractObject* in which the user will transcribe all the properties of the entity;

M2: Entities with graphic representation $\Rightarrow$ specialisation of the GGOs, for the local export schema in agreement with its representation structure in the native schema;

M3: Information layer $\Rightarrow$ SAIF classes that contain group of objects, i.e. *PartitionedCoverage* or *AnnotedSpatialDataSet*;

M4: Geometry of each geographic object $\Rightarrow$ SAIF geometry attribute of *GeographicObject* class. It may be further specialised for each GIS, using the structures offered by its corresponding data model;

M5: Name of each class $\Rightarrow$ its name in the local schema, prefixed by the name of the export Schema, concatenated with the string "::";

M6: Term that describes the geographic object chosen in the Thesaurus $\Rightarrow$ the *comments* field of the SAIF's syntactic class notation;

M7: Properties defined for the geographic objects in the original schema $\Rightarrow$ attributes of GeographicObject, including the properties pre-established by GIS, such as perimeter, length, area, etc;

M8: For each property it should be provided its name, a term, the format and the domain:
Atomic domains $\Rightarrow$ basic types *Integer*, *Real* or *String*; and
Classification domain $\Rightarrow$ type *Enumeration*;

M9: In case an entity or some of its properties in the original schema presents some integrity restriction $\Rightarrow$ the clause *restriction* of CSN;

M10: The conventional relationships among classes must be instantiated through the OSN and will be represented by specialised classes of the Relationship type, having properties that describe their cardinalities:
Temporal relationship $\Rightarrow$ specialisation of the *TemporalRelationship* class.
Spatial relationships $\Rightarrow$ specialised using the *SpatialRelationship* class.

3.2. Schema analysis

To determine the semantic proximity among different schemas, it is necessary to establish the correspondence of their descriptions. Thus, the schemas are analysed in an intermediate stage. This stage establishes an order for the schema integration, to correlate classes in each schema, and finally to identify conflicts among them. The MMultiGIS suggests that this analysis should be divided in three distinct phases as shown in Figure 3.

3.2.1. Spatial context analysis. The MMultiGIS accepts schemas that do not describe the same geographical space. Thus, it is necessary to establish the relationships among these schemas. Hence, this step accomplishes the following:

- *Analysis of the cartographic parameters:* This step solves conflicts among cartographic parameters used in the graphical representation of *GeographicObject*. These parameters describe properties such as the coordinate system, the projection system, the horizontal and vertical data and the scale.

  To detect conflicts among the cartographic parameters the *SpatialReferencing* metadata are analysed. To simplify this comparison the information for each schema (s_1, s_2,..., s_n) is displayed in a cartographic parameters table.

  In case of detecting differences in these parameters, system transformations are applied to the *GeneralLocation* properties. Besides, a flag will be created to indicate that specific transformations may be needed in other activities such as query output.

- *Analysis of the spatial relationship among schemas:* To analyse the geographical space treated by each database, the *GeneralLocation* properties, describing the geographic location of the schemas, are compared. To simplify this comparison a localisation table (LT) is created with a line for each schema (s_1, s_2,..., s_n).

  The spatial relationship determines how each local schema will take part into the global schema. The identification of this spatial relationship is determined by the operation of spatial intersection of the properties described in the LT. This comparison can produce the following types of spatial relationship:

  - *Empty* - occurs when the information communities are willing to take part in the federation, although their schemas cover different geographical spaces of the federation. In this case, specialised *Regions* classes are created to allocate the new geographical space that is not recognised by the existing classes and its corresponding schema is joined to the global schema. This join specialises the geographic objects treated by the local schema for the *Themes* classes, along with the respective mappings for the local schema;

  - *Partial* - occurs when the databases present spatial intersection areas, but this intersection is not complete. In this case, the schemas are treated in the federation according to the diagnosis of spatial relationship. The geographical spaces that have no spatial relationship will be treated as those of the first case. For the intersected areas the proposed methodology for the integration of the schemas is applied; and

  - *Total* - occurs when the geographic spaces treated by the local databases are coincident. In case it is a region that has already been contemplated by the classes belonging to the *Regions* class, the respective instance will be created and the *Themes* resulting in the integration process will be associated to it.

3.2.2. Application context analysis. Since each application area has a particular vocabulary, this phase has been proposed to analyse the application context to identify the schemas that are semantically closer and to establish a hierarchy for the application areas.

The MMultiGIS determines that the schemas having a partial or total spatial relationship should be analysed relatively to their application contexts, to identify their semantic proximity. The purpose of this comparison is to establish a priority order for the integration of schemas. This priority is established according to the application area and the project objective.

Therefore, this analysis is done through the comparison of the syntactic description of the metadata provided by the properties *Source.project.project Identifier.projectPurpose*, and *Source.project.project Identifier.applicationArea*, which describe the projects associated to the schema relatively to the purpose and application area. Hence, an application context table should be created showing the properties that describe the projects for each schema (s_1, s_2,..., s_n).

This analysis will enable the determination of a hierarchy in the context of application areas. This hierarchy is materialised by a tree, which allocates the application areas in the higher nodes. These nodes can be subdivided in one or more nodes, according to the application. This tree materialises the generic relationship among the schemas relatively to their application. This hierarchy is used to establish a priority order for the semantic analysis involving the schemas and, as a result, this determines an order where the sequences of pairs of schemas are integrated.

3.2.3. Semantic context analysis. The purpose of this phase is to establish the degree of semantic proximity among the classes of schema pairs (s_m, s_n), identifying their similarities and differences. This stage uses the conflict taxonomy proposed by STRAUCH et al.[19] and applies the concept of semantic context to help identifying the degree of semantic proximity of the classes. The semantic context is defined by an abstract concept for a geographical object. Each semantic context

will be treated as a collection of classes of objects that possess some association among them through relationships and common properties.

Thus, this phase aims at the co-ordination of the following analyses:

- *Analysis of class denominations:* This analysis consists of an observation of the intentional level of the specialised classes of *GeographicObject* and *AbstractObject* for each pair of schemas (s_m, s_n). The classes are arranged in a class table (table 2) and we look for the terms used to describe the classes, to determine their semantic proximity.

S_m			S_n		
CLASS$_i$	TYPE$_i$	LIST(TERM$_{ik}$)	CLASS$_i$	TYPE$_i$	LIST(TERM$_{ik}$)

Table 2. Comparison of class denominations

Denomination conflicts are frequently found. However, the terms used to describe the geographic object helps the identification of the classes that actually have semantic proximity. Therefore related classes are organised in thematic groups, in an ascending order of similarity.

These thematic groups determine a comparison terminology that helps the class analysis. They must be sent to the local database administrators, who must confirm the thematic grouping and establish the relationship among their classes and those belonging to other schemas. Each thematic group contains a semantic set and is assigned a general denomination.

- *Analysis of the semantic relationships:* For each thematic group (TG_g) detected in the previous analysis, we must create a table containing the relationships (table 3). In each table, we observe the semantic context of the relationship relatively to the denomination and to the respective cardinalities. This analysis aims at the detection of those classes of the schemas, which, besides having the semantic proximity, also have the same relationship context.

S_m:TG_g	CLASS$_i$	RELATIONSHIP$_{ij}$	CLASS$_j$

S_n:TG_g	CLASS$_o$	RELATIONSHIP$_{op}$	CLASS$_p$

Table 3. Comparison of TG_g relationships

This analysis detects the abstraction conflicts caused by the different forms of reality representation and the different components used to model it.

- *Geometry analysis:* Classes with semantic proximity, which are also specialisation of *GeographicObject*, must have their geometry analysed to identify the conflicts. Geometrical representation may use different territorial units of data acquisition and graphical representation. Thus, classes that have both the same denomination and the same relationship type should be arranged in a new table (table 4) containing the class geometry to be analysed.

S_m : TG_g		S_n : TG_g	
CLASS$_i$	GEOMETRY$_i$	CLASS$_j$	GEOMETRY$_j$

Table 4. Comparison of TG_g geometry

- *Identification of common properties:* Classes that have semantic proximity are also analysed to identify common properties. Thus, the class pair (S_m:C_k; S_n:C_k) having both denomination and relationship proximity will be inserted in a Not Normalised Properties Table (NNPT). This table contains for each property (P_i) theirs respective term (T_i). A search on NNPT (table 5) finds equivalent properties or properties that are implicit in other properties.

C_k	S_m::P_1.T_1	...	S_m::P_x.T_x
S_n::P_1.T_1	X		
.....			
S_n::P_z.T_z			

Table 5. Analysis of the common proprieties

This comparison consists of detecting for each line a column with a corresponding term. The matching pair should be marked with an "X" to identify the common properties or with some semantic proximity. This proximity found among the properties of the classes should be sent to the local administrators for confirmation.

- *Analysis of property descriptions:* The detection of proximity among class properties is refined through the creation of a Properties Correspondence Table. In this table (table 6), each class with an indicative of semantic proximity is described relatively to its name, type, constrain, format and domain.

S_m: TG_g: CLASS$_i$				
NAME	TYPE	CONSTRAIN	FORMAT	DOMAIN
S_n: TG_g: CLASS$_j$				
NAME	TYPE	CONSTRAIN	FORMAT	DOMAIN

Table 6. Correspondence between proprieties

- *Analysis of the semantic relationship of schemas:* The context analysis reveals the correspondence of similar classes. The methodology proposes two strategies to help establishing the global schema.

Firstly, each pair of elements encountered in the analysis stages is assigned a semantic proximity degree (SPD). This degree consists of a numeric value belonging to the [0,10] interval, which is established through weights defined heuristically in agreement with the similarity found among the terms, the relationships, the geometry and the properties of the class pair.

The second strategy consists of separating the classes in three large groups according to their degree of semantic proximity:

I) Identical classes - in this group, we find the rare class pairs whose SPD equals 10. In this case, the class definitions match exactly, with no information loss;

II) Similar classes - in this group we find the class pairs class whose SPD belonging to the [6,8]. This case occurs when, although there is not an exact semantic match there are some transformations that can be applied to make them compatibles, and

III) Heterogeneous classes - in this group, we find the class pairs whose SPD is inferior to 6. The classes of this group have no semantic proximity.

The classes belonging to group I and II are arranged in tables 7 and 8, respectively. Corresponding classes are represented by their denominations and properties along the same line to aid the conformation step.

S_m		S_n	
CLASS$_i$	LIST (PROPERTIES)	CLASS$_j$	LIST (PROPERTIES)

Table 7. Identical classes - Group I

S_m		S_n	
CLASS$_i$	LIST (PROPERTIES)	CLASS$_j$	LIST (PROPERTIES)

Table 8. Similar classes - Group II

The classes belonging to group II are sent to the local administrators to be analysed in the next step.

3.3 Conformation and integration

The objective of this step is to solve the semantic and structural conflicts of the schemas. This is done by the determination of solutions for the conflicts encountered in group II, to conform the geographic objects of each schema into a global schema.

Firstly, denomination conflicts are solved, then it deals with abstract conflicts followed by structural conflicts and finally geometry conflicts are solved.

The denomination conflicts are solved by the standardisation of names for the global schema classes. The methodology proposes that classes should use a generic denomination, readily understood by every federation member. This denomination is created using a common term selected by the administrators to describe the class.

To deal with abstraction conflicts among the properties, one must initially deal with the properties that refer to the same object in both schemas. One must recognise the properties that appear in the pair of schemas.

Next we check the presence of structural conflicts. In case there is no structural conflict, a class is created by the simple union of compatible properties. On the other hand, when one detects structural conflicts, the solutions are established through a consensus among the local administrators. These solutions lead to the creation of classes with the common properties or to a hierarchy of classes, or else to the specification of transform functions to be applied to formats, domains, units, etc.

The conflicts due to geometry differences lead to a generalisation for the class pair, according to its geometry.

After establishing the denominations, the conformation of properties and the geometry handling for each class pair, the methodology defines primitive equivalent classes, establishing the proper theme classification with their corresponding properties. Then, the relationships are determined according to the identification of generalisation/specialisation and aggregation, creating the global schema.

3.4. External Schema creation

In this step, we perform an abstract modelling to create an external schema from the correspondence and transformations established in the previous step.

In the external schema model, the geographic objects are arranged in specialisation of *Regions* and *Themes*. The *Regions* are created using the analysis of the spatial context that identifies the territorial units dealt by the schemas. For each territorial unit we identify the associated *Themes*.

The *Themes* classes are created after the semantic analysis of the schemas. The stronger concepts found in semantic analysis will become a specialisation of the class *Themes*, whose denomination has been established in the previous step through a consensus among the local administrators. Therefore, each pair of classes of the groups defined in the semantic context analysis will be integrated in the external schema as a specialisation of the *Themes* class, in the following way:
- The classes pertaining to group I will be integrated in the external schema in specialised classes of *Themes*

and *Regions* with all their properties and respective mappings for the local systems;

- The classes pertaining to group II have two possibilities:
  - In the case of classes whose properties intersect, we create one or more *Themes* specialisation. The first specialisation has the common properties and the others are specialised from the first with the properties that do not have structural equivalence; and
  - In the case of classes that have a composition relationship, we create two specialisation *Themes* classes, with their respective properties. We assign for each of these classes the composition relationship materialised by a property of the type set or list.
- For the classes pertaining to group III, specialisation of the *Themes* and *Regions* classes, containing all their properties and mappings for the local schemas, are created.

Each specialisation of the *Themes* and *Regions* classes will be assigned an object identifier (ID), a external name (N), a list of properties (P_i), a set of functions that establish the mappings of the elements in the federate level onto the corresponding elements in the local schemas (Θ) and the *Regions* about which the *Themes* has any information. For example, *Themes.name* (ID, N, *list* (P)$_i$, Θ, *set* (*Regions.name*)).

3.5. Validation

The objective of this step is to analyse the integrated schema to guarantee **i)** that any federation user can readily understand the resulting schema; **ii)** that no information has been lost; and **iii)** that no redundancy or unnecessary complexity is present.

For this step the MMultiGIS uses the metrics proposed by ÖSZU & VALDURIEZ [11], namely completeness, exactness, minimality and comprehension.

4. Conclusions

The proposed methodology takes in consideration a broad number of the aspects involved in the construction and in the modelling of a geographical database. Although the MMultiGIS assists MultiGIS architecture, its main contribution is its wide applicability to any architecture that contemplates multiple GIS databases.

The MMultiGIS deals with the differences among GIS data models and the differences among schema modelling. The first goal is reached through the conversion of the schemas onto a CDM and the second is accomplished through the activities for the interpretation of geographic data represented in the export schema. Our CDM extends the SAIF model to contemplate geographic region specific properties and GIS application properties. We also refine SAIF's rules for schema mapping onto the common data model.

The data model proposed for the external schema is a simple model. It uses *Themes* and *Regions* to handle the geographic data in the federation. Previous experiences of using this methodology have shown its effectiveness. Since the Geoprocessing users work with information layers, they were benefited *Themes* and *Regions*, which from simplified the identification of the desired classes.

5. References

[1] AGUIAR, C. D. & MEDEIROS, C. B., 1996, "Construction of a unified basic model starting from systems stand-alone". In: *Anais do GIS 96*, pp. 503-515, Curitiba, (May), (In Portuguese)

[2] AHMED, R, SMEDT, P. D., DU, W., KENT, W., KETABCHI, M. A., LITWIN, W., RAFII A., SHAN, M., 1995, The Pegasus heterogeneous Multidatabase System, *IEEE Computer*, v. 24, n. 12 (Dec.), pp. 19-27.

[3] BONJOUR, M. & FALQUET, G., 1994, "Concept bases: a support to information system integration". In: *Proceedings of Caise*94 Conference*, Utrecht, pp. 242-245

[4] BUNEMAN, P, RASCID, L. & ULLMAN J., 1997, "Mediator language- a proposal for standard", *SIGMOD Record*, v. 26, n. 1 (Mar.), pp. 39-46.

[5] DOGAC, A., DENGI, C., KILLIC, E. ET AL., 1996, METU Interoperable Database System, ACM SIGMOD RECORD, vol24, n.3 or Technical Report 6-1, Software R&D Centre, Jun./1995

[6] FRANKHAUSER, P, MOTZ, R. & HUCK, G., 1995, "SIM Schema integration methodology. IRO/SPEC/D4-4.1-V1.0/PF950210

[7] IGRAS, E., 1994, "A framework for query processing in a federated database system: a case study". In: *Proceedings of Annual Conference URISA*, pp. 167-178.

[8] KRAMER, R, NIKOLAI, R, KOSCHEL, A., ROLKER, C., LOCKERMANN, P. & KEITEL, A., 1997, "WWW-UDK: a web-based environmental meta-Information System", *SIGMOD Record*, v. 26, n. 1 (Mar.), pp. 16-21

[9] KUMAR, K., SINHA, P & BHATT, P. C. P. P., 1997, "Do-GIS: A distributed and object oriented GIS". In: *Advances in GIS Research II: Proceedings of VII International Symposium on Spatial data Handling*. ED. Kraak & Molenar, Taylor & Francis, pp. 263-275

[10] LECLERCQ, E., BENSLIMANE, D. & YÉTONGNON, K., 1996, "A distributed object architecture for interoperable GIS", In: *IEEE Knowledge and Data Engineering Exchange Workshop,* pp. 73-80, New Port Beach, USA

[11] ÖSZU, M. T. & VALDURIEZ, P., 1991, "Distributed data management: unsolved problems and new issues", *Computer*, v. 24, n. 8.

[12] PARENT, C., SPACCAPIETRA, S. & DEVOGELE, T., 1996 "Conflicts in spatial database integration". In: *Proceedings of the International Conference Parallel and Distributed Computing Systems*, Dijon: France, v. 2, pp. 772-778, Set

[13] PIRES, F.; MEDEIROS, C.B. & SILVA, A. B.; 1993, "Modelling Geographic Information Systems using an Object Oriented Framework", In: Proceedings of the 13 Th. International Conference of the Chilean Computer Society, Chile, pp. 217-232.

[14] RAMIREZ, M. R., 1994, *Systems Database Managers for Geoprocessamento. Master* Dissertation of COPPE/UFRJ, Rio de Janeiro, Rio de Janeiro (In Portuguese)

[15] SAIF, Survey and Resource Mapping Branch, 1995 "Spatial Archive and Interchange format: Formal Definition, release3.2", Ministry of Environment, Lands an Parks, Province of British Columbia, Canada

[16] SHETH, A. P. & LARSON, J. A., 1990, "Federated Database Systems for managing Distributed, heterogeneous and autonomous databases", *ACM Computing Surveys*, v. 22, n. 3 (Sep.), pp. 183-236

[17] SOUZA, J. M, 1986, *Software tools for Conceptual Schema integration*, Ph.D. Dissertation, University of East Anglia, Local

[18] SPACCAPIETRA, S. & PARENT, C., 1994, "View integration: a step forward in solving structural conflicts", *IEEE Transactions on Knowledge and Data Engineering*, v.6, n. 2 (Apr.), pp. 240-274

[19] STRAUCH, J. C. M., MATTOSO, M. L. Q. & SOUZA, J. M. DE, 1996 Issues in interoperability between GIS databases with heterogeneity and distribution aspects". In: *Proceedings of GIS Brasil 96*, Sagres Editor, Curitiba, (In Portuguese)

[20] VCKOVSKI, A., 1996, "Java as a software system for distributed and interoperable Geoprocessing". In: *Proceedings of the International Conference Parallel and Distributed Computing Systems*, Dijon: France, v. 2 779-783, (Sep)

[21] WIEDERHOLD, G., 1995, "Mediation in Information Systems", *ACM Computing Surveys*, v. 27, n.2 (Jun.), pp. 265-267

[22] YASER, B., 1997, A mechanism for object identification and transfer in a heterogeneous distributed GIS. In: *Advances in GIS Research II: Proceedings of VII International Symposium on Spatial Data Handling*. Ed. Kraak & Molenar, Taylor & Francis, pp. 929-939