ICC Topographic Databases: Design of a MRDB for data management optimization

Blanca Baella, Anna Lleopart, Maria Pla

Institut Cartogràfic de Catalunya
Email: blanca.baella@icc.cat; anna.lleopart@icc.cat; maria.pla@icc.cat

1. Introduction

Since the foundation of the Institut Cartogràfic de Catalunya (ICC) in 1982, one of the main duties was the production of topographic data. Databases at different resolutions were produced and maintained ranging from 1:1000 scale to 1:250.000 scale. The most detailed information covers the urban areas at 1:1000 scale (20cm of accuracy), while data at 1:5000 scale (1m accuracy) and 1:25.000 (2,5m accuracy) covers Catalonia. These databases are compiled in 2.5D using stereoploting on top of digital photogrammetric systems. Smaller scales such as 1:50.000 and 1:250.000, covering also all the country, are collected in 2D, digitizing on top of orthophotoimages.

After a first generation of spaghetti data collected using CAD systems, more complex models were designed and implemented for allowing new data exploitations using GIS systems. Although the data models were designed to preserve the semantic coherence between scales, there are no explicit relationships between the representations of the same geographical object in the different databases, and each topographic database is produced independently. Because of the lack of commercial tools and the different updating cycle of each database, the implementation of a MRDB in the production environment never was considered.

The huge demand of updated information and the high pressure to obtain derived products for visualization in internet and mobile devices has introduced new requirements costly to achieve with the current models. The implementation of a MRDB that integrates the topographic data could be a solution to fulfill these requirements. The MRDB, in addition to a better management of the updating processes, will allow optimizing the generation of derived products using generalization, including on-the-fly processes for intermediate resolutions, used mainly for visualization.

The paper will explain how the ICC is designing and starting the implementation of a topographic MRDB, taking into account the current situation of the ICC databases, the existing limitations in commercial software, the requirements to be implemented in the ICC production environment and the problems to achieve reasonable productivity ratios.

2. Current situation of the ICC topographic data

The first design of the ICC Topographic MRDB will include two strongly related databases with complete coverage of Catalonia: the Topographic Database at 1:5000 scale (BT-5M) and the Topographic Database at 1:25.000 (BT-25M).

The first version of the BT-5M started in 1985. Digital vectors were compiled using analog and analytical steroplotting on top on a CAD system, and the information was never structured to create a geographical database for GIS purposes. Using the collected topographic information, a digital terrain model (DTM) and a digital surface model (DSM) was produced. This first version was completed in 1995. For the updating process, started in 1996, some changes were introduced in the data model for allowing GIS exploitations and for facilitating further generalization at smaller scales. The new datamodel was improved by adding polygons and ancillary elements for creating hydrographic and communication networks, and by defining a more detailed classification of geographical names.

The need of updated base data at 1:25.000, the existence of a production program for the BT-5M, and the experiences at the ICC in implementing generalization workflows, allowed to start producing the first version of the BT-25M using generalization processes in the year 2003. Comparing with previous ICC generalization experiences, the workflow entailed two challenges: to obtain a topographic database, not only a map, and to generalize 2.5D data instead of 2D data. The main difference between generalization for obtaining a map or for creating a database comes from having to preserve the topological structure of the data and their attributes. The 2.5D characteristic of the generalized data required a new software development and a careful editing process. It is worth noting that the efforts devoted to derive a database instead of a map were small compared with ones spend in preserving the 2.5D nature of the topographic database.

Although the BT-25M was derived by generalization from the BT-5M, there was no explicit mechanism linking the original and the generalized datasets. In some sense, it might consider that the ICC lost an opportunity to build a MRDB. The main reason was the lack of tools for compiling photogrammetric data in a GIS production environment, which forced to keep using a photogrammetric CAD system. The desired improvements in the datamodel, such as unique identifiers for the objects and metadata at object level, were not implemented at that time because they could not be managed efficiently in a CAD system.

In the last few years, GIS based photogrammetric tools ready for production environments and delivering reasonable productivities compared with those obtained using CAD systems have started to become common. The availability of these tools has encouraged the ICC to think in a new version of the datamodel that would implement the main aspects of a MRDB, linked objects and different levels of detail, where the links between the objects will be established through the unique identifiers.

3. ICC MRDB Data model

3.1 Objects

The first decision in the design of the new data model has been defining the meaning of the concept "object" and how to model it. It was decided to use the definition of feature included in ISO standards: "A feature is an abstraction of real world phenomena [ISO 19101] and a fundamental unit of geographic information [ISO

19109]”. Figure 1 describes the most abstract level of defining and structuring geographic data, from the real-world phenomena to feature instances.

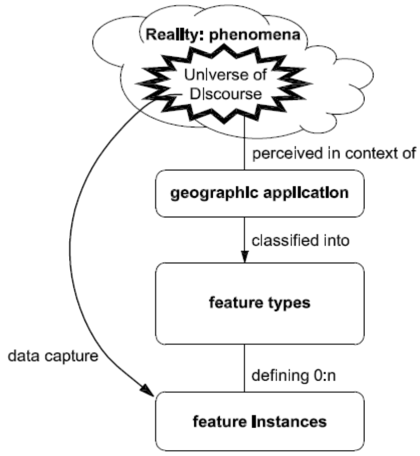
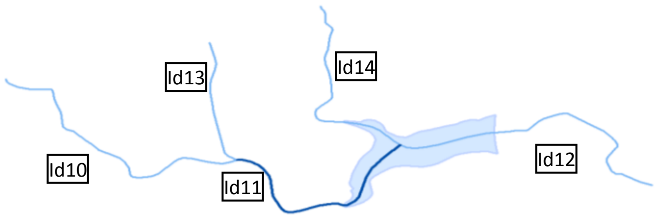


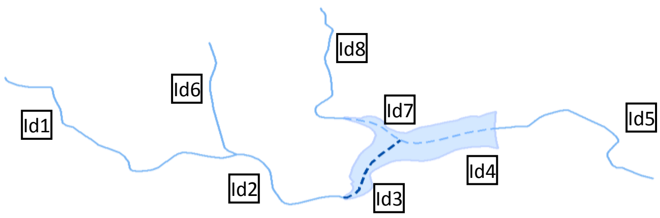
Figure 1: The process from universe of discourse to data [ISO 19109].

Depending on the use cases to which a model must respond, real world phenomena can be modeled in different ways. For instance, the feature that models the real world phenomenon “river” may refer to the whole river (in dark blue in Figure 2a), to the river stretch between two intersections (Figure 2b) or to the river stretch with common attributes between two intersections (Figure 2c).

(2a) The whole river (in dark blue).



(2b) A river stretch between two intersections (in dark blue).



(2c) A river stretch with common attributes between two intersections (in dark blue).

Figure 2: Examples of real world phenomenon “river” modeling.

In addition, as a result of the use cases requirements, the conceptual model may include several modelling of the same real world phenomenon. Features may be aggregated in “complex features”, which are, in turn, features.

For instance, if “RiverStretch” feature type is defined as “a river stretch with common characteristics between two intersections” and “River” feature type is defined as “the whole extent of the river”, then “River” feature type may be a “complex feature” defined as an aggregation of “RiverStretch” feature type. Figure 3 shows a “River” feature instance (Id100) as an aggregation of “RiverStretch” feature instances (Id1,Id2,Id3,Id4,Id5).

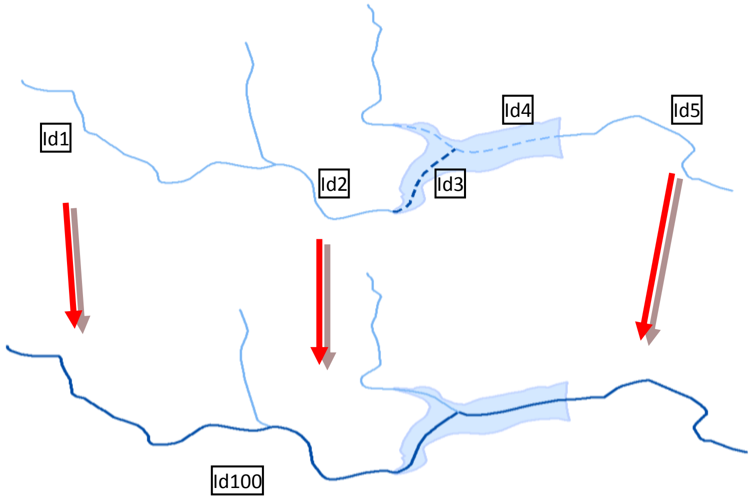


Figure 3: A“RiverStretch” feature and a“River” complex feature (in dark blue).

As “complex features” are strongly dependent on the use cases, and moreover, because it is a complex structure, collection and maintenance of “complex features” reduces the performance of the topographic database production process, it was decided that the new data model will be based on simple features with common attributes. In the example, it would correspond to the Figure (2c).

3.2 Attributes

In the new data model, each feature instance will be characterized by an identifier that must be unique and persistent along the whole life-cycle of the feature instance and never reused.

Moreover, each feature is characterized by four types of attributes: spatial, temporal, thematic and metadata attributes.

The spatial attribute carries the geometric representation of the features.

Life-cycle information describing temporal characteristics of the feature in the DBMS makes possible feature versioning, the propagation of changes across scales and the propagation of changes between features of related databases. In the new data model, life-cycle information will be managed by four attributes. They are the insertion date, assigned when the feature instance is created; the modification date, assigned when the feature instance is modified; the deleting date, assigned when the feature instance not longer exists; and the revision date, corresponding to the last date when the instance was revised with a new data source. If the feature has not changed in the new data source, the modification date will be maintained and only the revision date will be updated. The unique feature identifier and the life-cycle information allow obtaining, for each feature instance, its successor/predecessor information.

Metadata attributes contain the information related to lineage and quality. Lineage metadata attributes stores the data sources used to collect the information, such as photogrammetric flights, field survey, map names sources, etc, and also their dates.

Finally, features have a set of thematic attributes as classification attributes or other information as geographical names or road numbers.

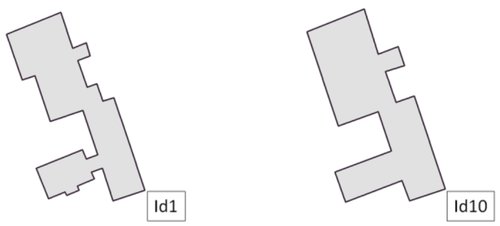
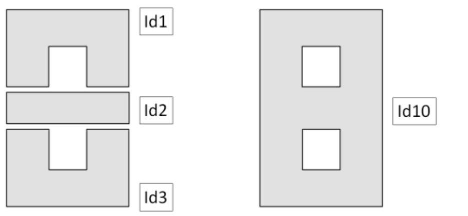
3.3 Standards

In the design of the new data model, national and international standards have been taken into account to guarantee interoperability, ensuring that the new model can be translated fully automatically to the Base Topográfica Armonizada (BTA), the Spanish standardized model for topographic data, and to the European INSPIRE data models.

3.4 MRDB modelling

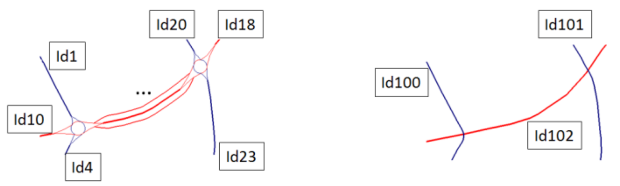
The ICC MRDB data model will be based in one single schema with linked data, where one feature belongs to one single resolution and has a link with one, or more, features of the other scale. Information about the generalization operation used to derive smaller resolutions will not be stored.

The links between corresponding feature instances at different level of detail will be determined by the various cardinalities that exist in the relations between them. In the case of the 1m and the 2,5m resolution data, where the lower accuracy data has been obtained applying generalization operations, the possible cardinalities are “one to one” (1:1), “many to one” (n:1), “one to many” (1:n) and “many to many” (n:m). Thus, it will be necessary to model (n:m) relationships (Table 1). Other cardinalities (0:1, 0:n), corresponding to the appearance/creation of new feature instances, or the derivation of one feature type from other ones are not considered in the ICC MRDB modelling, due to the small ratio between resolutions.

<table>
<tr><td>1:1
Simplification</td><td>
</td><td><table><tr><th>1m resolution</th><th>2,5m resolution</th></tr><tr><td>Id1</td><td>Id10</td></tr></table></td></tr>
<tr><td>n:1
Geometric aggregation</td><td>
</td><td><table><tr><th>1m resolution</th><th>2,5m resolution</th></tr><tr><td>Id1</td><td>Id10</td></tr><tr><td>Id2</td><td>Id10</td></tr><tr><td>Id3</td><td>Id10</td></tr></table></td></tr>
</table>

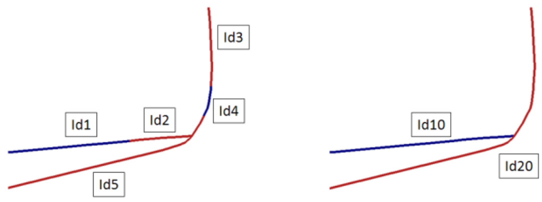
n:1

Geometric and semantic aggregation



1m resolution	2,5m resolution
Id1	Id100
...	Id100
Id4	Id100
...	...
Id20	Id101
...	Id101
Id23	Id101

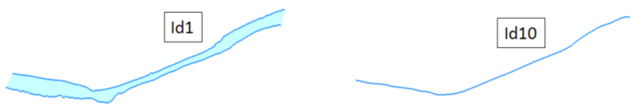
Semantic aggregation



1m resolution	2,5m resolution
Id1	Id10
Id2	Id10
Id3	Id20
Id4	Id20
Id5	Id20

1:1

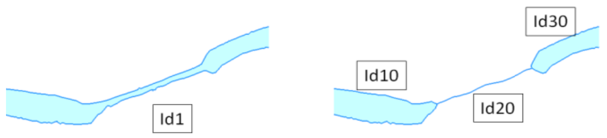
Collapse



1m resolution	2,5m resolution
Id1	Id10

1:n

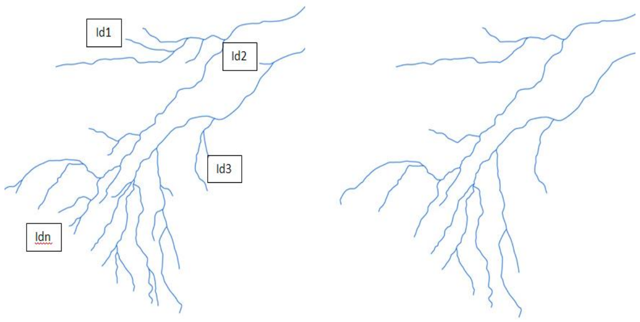
Partial collapse



1m resolution	2,5m resolution
Id1	Id10
Id1	Id20
Id1	Id30

1:0

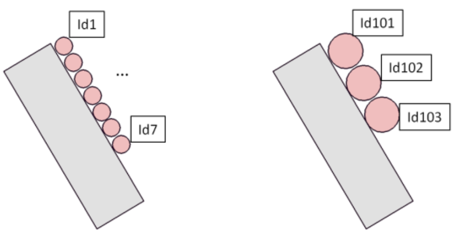
Prunning



1m resolution	2,5m resolution
Id1	-
Id2	-
Id3	-
...	-
Idn	-

n:m

Typification



1m resolution	2,5m resolution
Id1	Id101
Id1	Id102
Id1	Id103
...	-
Id7	Id101
Id7	Id102
Id7	Id103

Table 1: Examples of different types of relation cardinalities.

Due to the complete coverage of the current version of the BT-5M and the BT-25M, the links between the feature instances at different resolutions will be established using matching techniques that will be applied in the migration of the existing data to the MRDB data model. For later updating processes on the higher accuracy data, the lower resolution data and the links will be managed in the generalization process.

4. ICC MRDB Implementation

The implementation of the new ICC topographic datamodel includes two main aspects, the migration of the photogrammetric data collection from a CAD system to a GIS system based on a DBMS, and the management of the data and the MRDB relationships.

The migration to a GIS environment will imply the customization of the commercial system. The goal is to achieve productivity similar to the one of the existing CAD environment. The customization includes the development of a set of tools that preserves the 2.5D nature of the data model, for example:

- The adaptation of existing tools to the needs of the data capture. For example, the rubber banding management, the use of 2D and 3D snapping, the editing tools that work with the XY or Z values locked, the automatic management of metadata attributes, or the optimized selection of classification attributes.
- The implementation of geometric primitives not available on the system. For example, curve or ellipse.
- The development of complex data capture tools. For example, draw an element using curve or arc geometry and store it as a stroked polyline; draw an element squaring automatically the outline if the angle of consecutive segments is close to 90º; or the use of auxiliary elements associated to the cursor to visualize distances during data capture.

Another important task to accomplish is the training of the production teams in the GIS system and in the customized tools.

The management of the MRDB links will be based on tables that store the relationships between features instances through unique identifiers, as is showed in Table 1. The first implementation of the MRDB will be done from the two existing databases, 1m and 2,5m of accuracy, which do not have any explicit link between the feature instances representing the same geographical object. The matching processes to find the related feature instances must fill the tables of relationships. In further updating processes of the database, the generalization operations to derive the smaller scale must establish and manage automatically these links.

At the time of this writing, the design of the datamodel has been completed, the programming of data capture tools is in progress, including the unique identifier and the life-cycle information management, and the operator training on GIS has also started. Next step will be the in depth analysis of the MRDB implementation taking into account their impact on the performance in the production environment. The analysis will be focused in the management of the MRDB links, its maintenance in the updating operations and also the aspects related to the matching processes.

5. Conclusions

The ICC MRDB should optimize the generation of the wide range of products required for visualization and data dissemination, the spreading of other exploitations and, finally, the high amount of resources devoted to producing and managing topographic data.

The new workflow implementation requires a high effort to customize the commercial GIS system used in the production environment in two main aspects: the photogrammetric data collection and the management of the data and the MRDB relationships. Moreover, this implementation will represent a huge change for a production environment based until now on CAD systems.

After the complete implementation and deployment of the MRDB production workflow for 1m and 2,5m resolution data, if reasonable productivity ratios are achieved, the workflow will be extended to ICC topographic data at other resolutions.

References

Baella, B., Pla, M., An example of database generalization workflow: the Topographic Database of Catalonia at 1:25.000, *ICA Workshop on Progress in Automated Map Generalization, Paris, 2003.*

Baella, B., Pla, M., Reorganizing the Topographic Databases of the Institut Cartogràfic de Catalunya applying generalization, *ICA Workshop on Progress in Automated Map Generalization, A Coruña, 2005, http://aci.ign.fr/Acoruna/Papers/Baella_Pla.pdf*

Balley, S., Vers la représentation multiple: le project MurMur, *Bulletin d'Information de l'IGN*, nº 73, 2002.

Hampe, M., *Integration einer multiskaligen Datenbank in eine Webservice-Architecture*, Bayerischen Akademie der Wissenschaften in Kommission beim Verlag C.H. Becl, München, 2007.

Mustière, S., Smaalen, J., "Database Requirements for Generalisation and Multiple Representations", in *Generalisation of Geographic Information: Cartographic Modelling and Application* (Mackaness, W.A., Ruas, A., Sarjakoski, L.T. (Ed.)), Elsevier Ltd., 2007, pg. 113-136.

Parent, C., Spaccapietra, S., Zimányi, E., *Conceptual Modeling for Traditional and Spatio-Temporal Applications, The MADS Approach*, Springer-Verlag, Berlin Heidelberg, 2006.

Sester, M., Sarjakoski, L. T., Harrie, L., Hampe, M., Koivula, T., Sarjakoski. T., Letho, L., Elias, B., Nivala, A.-M., Stigmar, H., *Real-time generalisation and multiple representtation in GiMoDig mobile service*, 2004.

Vangenot, C., Parent, C., Spaccapietra, S., Modelling and Manipulating Multiple Representtions of Spatial Data, *Symposium on Geospatial Theory, Processing and Applications*, Otawa, 2002.