An example of database generalization workflow: the Topographic Database of Catalonia at 1:25.000

Blanca Baella, Maria Pla
Institut Cartogràfic de Catalunya
Parc de Montjuïc
E-08038 Barcelona
http://www.icc.es
blancab@icc.es
mariap@icc.es

Introduction

Since their foundation in 1983, the Institut Cartogràfic de Catalunya has been producing and continuously updating three vector topographic databases covering Catalonia at scales 1:5.000, 1:50.000 and 1:250.000. The 1:250.000 database is used for general applications that need a full coverage of the country at small scale. The 1:50.000 is used for most of the GIS applications in the public administration and the 1:5.000 scale is used for public work planning and GIS.

In the last years, ICC costumers were asking for more and more a database for new GIS applications and mapping. This database should be manageable, like the 1:50.000 database, but with the topographic information as detailed as possible, like the 1:5.000. Moreover, elevation data is becoming essential for visualization and analysis applications. A new topographic database at 1:25.000 scale and 2.5D vector data should fulfill these user requirements.

The 2.5D nature of the new database implies the use of photogrammetric methods for data compilation, but the availability of the 2.5D Topographic Database at 1:5.000 and the previous experiences at the ICC implementing generalization workflows, offered us the possibility to produce the new database using generalization methods.

The original data: the Topographic Database of Catalonia 1:5.000

The 1:5.000 scale Topographic Base of Catalonia is a topographic database in 2.5D vector format covering Catalonia at the largest scale. The project started in 1985. Digital vectors were compiled with analog and analytical photogrammetric systems, but the information was never structured to create a geographical database for GIS purposes. It was only "spaghetti" data for using on automatic plotters. Using the collected topographic information, a digital terrain model (DTM) was produced in grid format with one elevation point every 15 meters. This DTM was used for orthophoto rectification and for shading maps at smaller scales. This first version was completed in 1995.

In 1996 we started the updating process using digital photogrammetric systems, which allow superimposition of stereo images and vector data. Some changes had to be introduced in the data structure for obtaining a GIS oriented database and to facilitate further generalization at smaller scales. The limitations of the too simple data model of the first version became clear after the unsuccessful attempt to use automatic generalization to 1:25.000. This has been reported several times since 1993. The new database includes polygons, hydrographic and communication networks and blocks in urban areas that define the street network. Additional information is also included in order to enhance the data, for example kilometric benchmarks of the roads and an improved classification of geographical names.

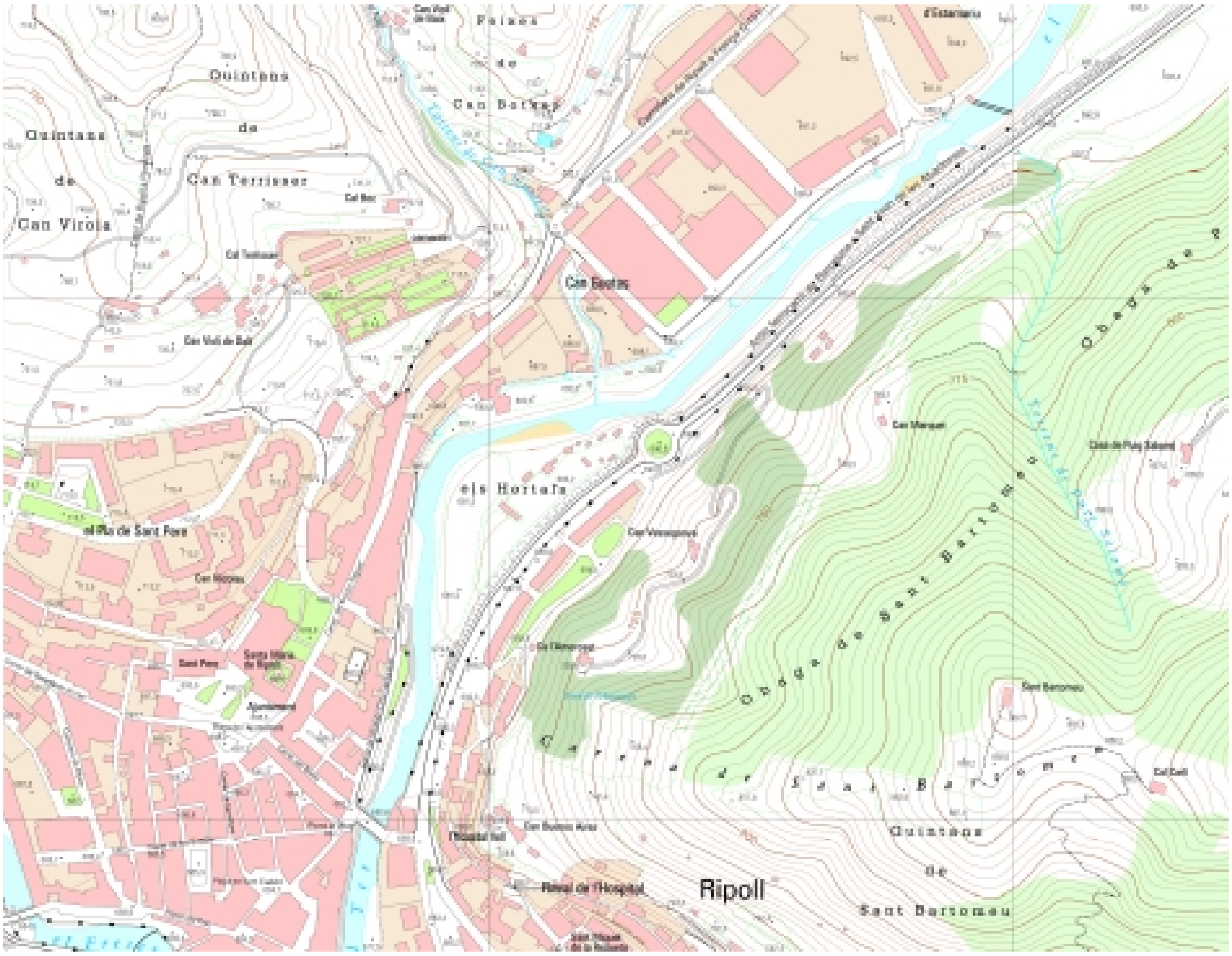


Figure 1.- Detail of the printed map of the Topographic Database of Catalonia at scale 1:5.000.

A significant improvement of the new version is the complete set of documentation, that includes information about the product, such as the technical specifications, detailed guidelines for data capture and metadata.

The guidelines for data capture contains the full description of the entity, its sources, the geometry, the topology and relationships with other entities, and the rules and parameters for compilation. The rules describe with special care the elevation of the elements because the 2.5D nature of the database.

The entities are represented by point, line and polygon geometry. All the vertices are defined by three coordinates. Lines representing hydrographic objects store also the orientation. Each polygon is composed by lines representing the boundary and a point representing the centroid. Although the polygons have 3D coordinates in each vertex of the boundary, they are not 3D surfaces.

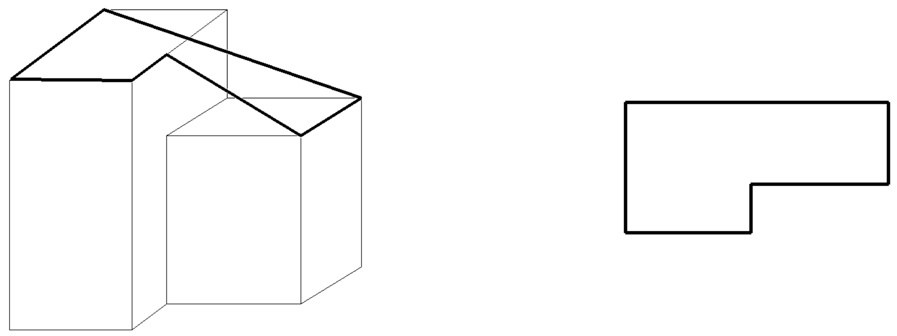


Figure 2.- The figures show the building data capture. Left figure shows that for each vertex only the top value is collected and that the intermediate vertices are not collected to avoid alignment problems. Right figure shows the top view of the same building.

There are no duplicated lines in the data model; therefore adjacent polygons share the common line geometry. In the line intersections there is a vertex for each line, but not a node because it can be generated automatically.

The generalized data: the Topographic Database of Catalonia 1:25.000 data model

The first step was the definition of the data model, which was designed taking into account the data model of the existing ICC vector databases for larger and smaller scales. The reason was to keep as much as possible the semantics of the objects across the different scales. The database documentation includes detailed information about the product and the production process. For each object, the guidelines describe in detail both the generalization and the photogrammetric data capture procedures. For generalization, the guidelines provide the sequence of operations with their parameters and some considerations about the related objects. As mentioned before, the availability of the 1:5.000 Topographic Database and the previous experiences at the ICC implementing generalization workflows led us to create the first version of the 1:25.000 database using generalization methods.

The methods of updating the generalized database will be dependent on the possibility to link the original database with the generalized one, in such a way that any update of the original database can be propagated automatically to the generalized one. Assuming that this can be done, it forces that any update at 1:25.000 be performed first at 1:5.000. This conflicts with the updating frequencies because medium scale databases are usually updated more often than large scale ones. The alternate solution is to update both databases separately. In this case every database has its own life but updates are duplicated and the consistency between the databases gets lost after a short period of time.

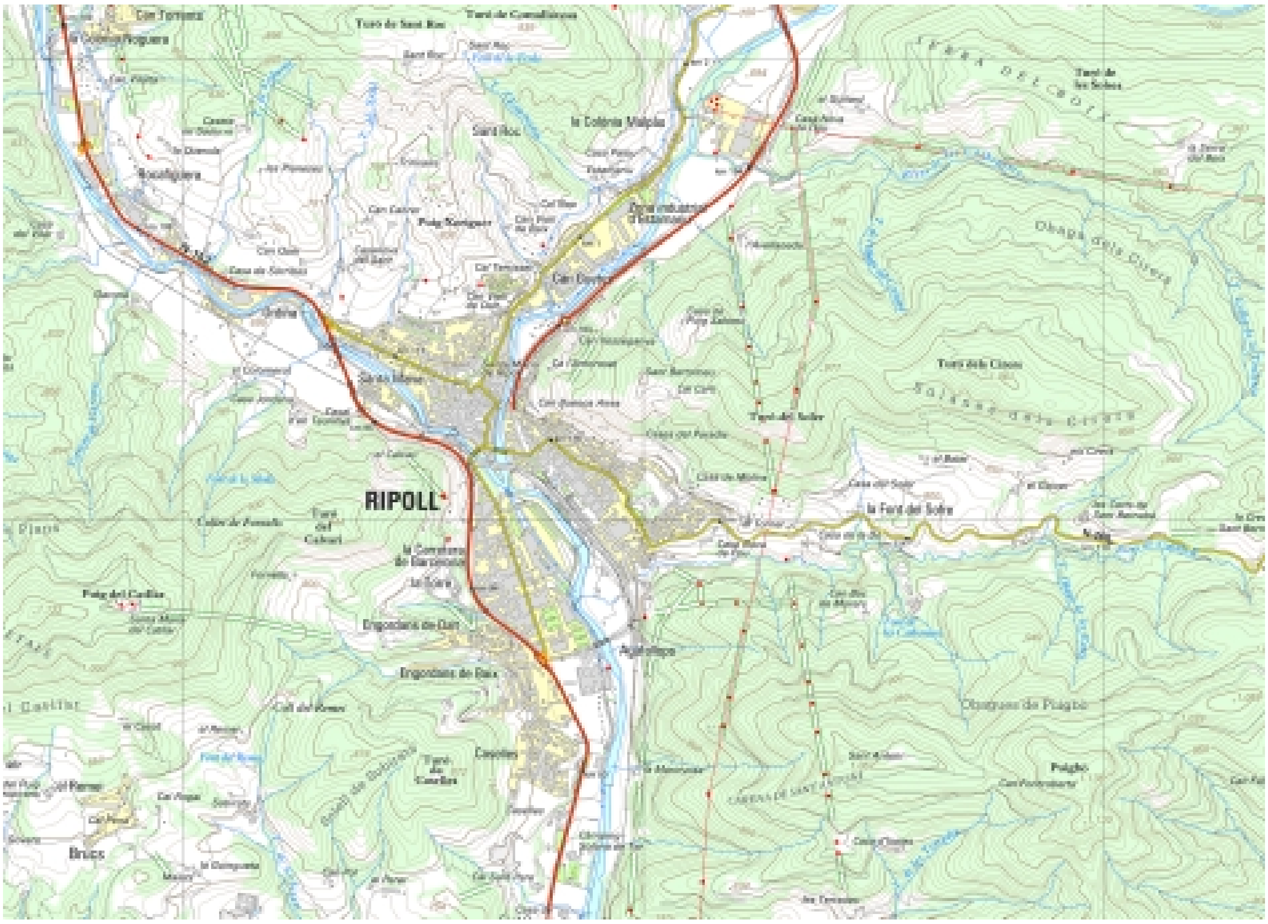


Figure 3.- Detail of the printed map of the Topographic Database of Catalonia at scale 1:25.000.

The generalization software

After the database was designed and after deciding that the first version would be generated with generalization, the next step was the implementation of the workflow. The basic software requirements were 2.5D generalization, good building simplification and easy integration in our environment. Another wish was to find an object oriented software to guarantee that the generalized database would preserve the relationships defined in the data model. In addition, the same system should provide a stereoplotting interface for direct updating of the database. Even though an object oriented system would mean a radical change in the way the ICC structures data and, of course, in the production environment, the benefits for both databases, the generalized and the original one, would largely justify the migration.

The for long time desired requirement to link the original and the generalized databases was discarded from the very beginning because any commercial software at hand did not offer this capability in a straightforward way. Nevertheless DYNAGEN from Intergraph and LAMPS2 GENERALIZER from Laser Scan were tested. The first one manages 2.5D data but the automatic building generalization requires too much manual editing to achieve good results, and the stereoplotting interface was not available. LAMPS2 GENERALIZER was not selected because the generalized data becomes 2D. The high cost per seat of the LAMPS2 environment was hardly justifiable considering that the system did not provide a complete solution. Because of these reasons we decided to proceed with another solution and delay the use of an object oriented software until these functionalities would be implemented.

The selected option was to use CHANGE from the University of Hannover for building generalization – although it produces 2D generalized data – and ICC software for other generalization operations. The excellent results of CHANGE in 2D building generalization were well known at the ICC from the 1:5.000 to 1:10.000 generalization workflow. The problem of assigning Z values to generalized buildings was solved with own software that computes a new Z value for each generalized vertex. The results have been tested and validated by superimposing the generalized data with the stereoscopic images on a digital photogrammetrical system.

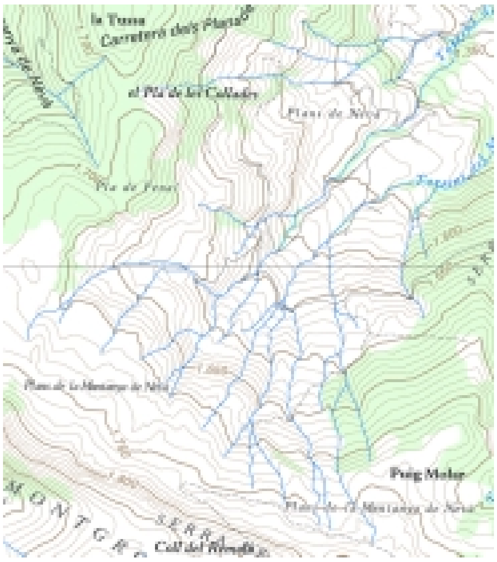
An important percentage of the development resources has been invested in the implementation of interactive tools so to optimize manual generalization and editing. The 2.5D nature of the data has required to improve existing 2D interactive generalization tools developed at the ICC.

The last step has been to implement the "on demand" distribution of the generalized database in digital vector format for GIS applications and in paper or raster format for automatic database symbolization.

The generalization process

The generalization operations applied are:

- Orography:
    - Contour lines. The contours multiple of 10 meters are selected and simplified using Douglas-Peucker with a tolerance of 25 cm. No smoothing is applied. The "light" parameter values used in simplification and the not use of any smoothing avoid conflicts thus preserving the quality of the generalized data. Short length contour lines delimiting small areas without any spot height inside, are eliminated.
    - Spot heights. The selection is done manually from the digital data of the Topographic Map 1:10.000. This map is obtained also from the Topographic Database 1:5.000 by generalization. The use of this intermediate generalized data speeds up the selection process.
    - Embankments. The main generalization operations are simplification, aggregation and selection, but the hardest task is the conflict resolution with communications and hydrographic elements.
- Hydrography:
    - Coast line. The lines are automatically simplified and interactive aggregation, exaggeration and typification operations are applied.
    - Watercourses. Simplification and typification are applied on watercourses represented by one line. For the two margins courses, in addition to the collapse of the margins and then aggregation, exaggeration and typification is applied.



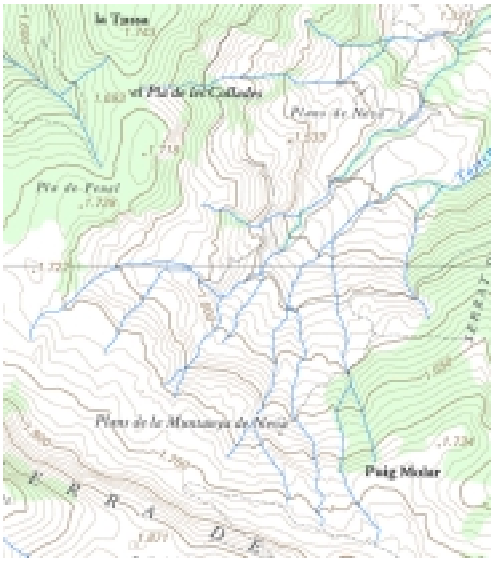


Figure 4.- Left figure shows the results after automatic generalization. Right figure shows the results after the manual typification.

- Lakes and reservoir. For the boundaries, the operations are simplification and partial collapse. In the regions with a high density of lakes, typification can be necessary. Aggregation, exaggeration and typification is applied to the islands.
- Wharfs and dams. Only collapse is applied.
- Water channels. Simplification is applied. After generalization, the channels represented by two margins are re-classified according to their width, and some of them become channels by centerline.
- Pools and swimming pools. Isolated and small are eliminated. In areas with high density they can be aggregated, exaggerated and typified.

- Communications
  - Roads. In the original database, roads are compiled by both margins and the centerline. Although in the final database only the centerline is preserved, its generalization introduce conflicts with almost all the elements surrounding it because they must be at a distance to the centerline greater than a certain threshold. The generalization operators applied to the centerline are simplification, collapse and conflict resolution. In order to detect the conflicts, a buffer zone is generated for each centerline showing the threshold between the road and the other elements.

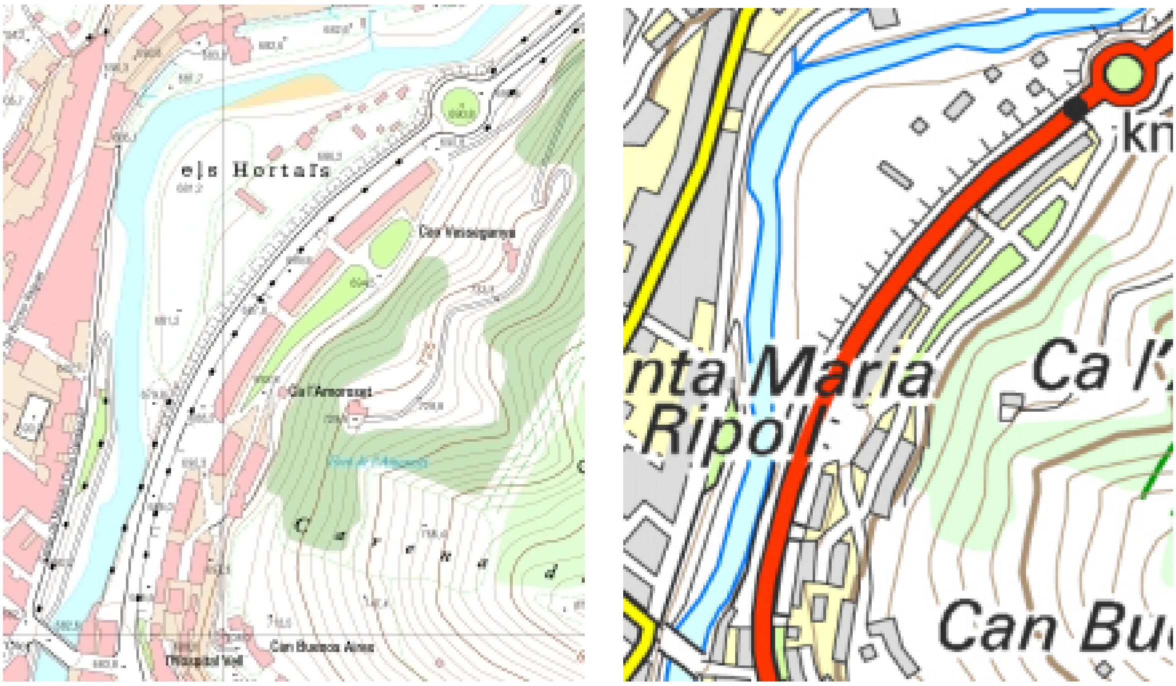


Figure 5.- Left figure shows some roads in the original data. Right figure shows the results after the generalization.

- Streets. In both databases the streets are stored by the centerline, but implicitly their margins are defined by the elements nearby. Whereas in the original database the street width is measured on the terrain, in the generalized database most of them must be exaggerated. This transfers the conflicts to the nearby elements, mainly blocks and buildings.

Figure 6.- Left figure shows the streets in the original data. Central figure shows the results after the automatic generalization and right figure shows the results after exaggerating the street width.

- Railways. The main generalization operations are simplification, collapse. Typification is also applied in railway stations with multiple rail tracks.

- Bridges. Only the very large bridges are selected. For each eliminated bridge, an attribute is assigned to the elements passing over it.

- Built up areas:
  - Blocks. Block boundaries are composed by the buildings outlines and the block lines.

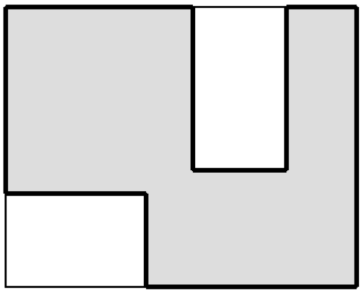


Figure 7.- The figure shows a block. The block boundary is composed by the block lines and some lines of the building outline.

The original block lines cannot be generalized together with buildings because CHANGE generalizes feature by feature. So the connections between block lines and buildings are lost and the blocks must be rebuilt by modifying manually the block lines.

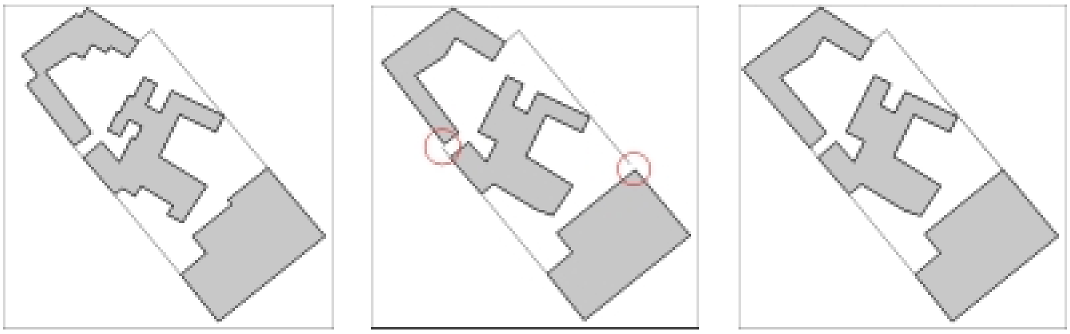


Figure 8.- The left figure shows the original block. The central figure shows the result after automatic building simplification. Right figure shows the results after the manual editing.

Because in the original database the polygons are stored sheet by sheet, the first operation is the aggregation of the original blocks in the 1:25.000 sheet, in order to merge the adjacent polygon blocks.

  - Buildings inside blocks. Building are generalized using CHANGE. The input is 2D polygons. To preserve the streets, which should be generalized manually, the buildings are generalized block by block, applying simplification to eliminate detail, and aggregation to merge buildings too close together. The last automatic process is the Z-coordinate assignment to each vertex. If the X,Y-coordinate is not modified, the original Z value is maintained, otherwise, the Z value is interpolated using the original Z values of the vertices of the buildings inside the block. Finally the automatic generalization is refined by applying aggregation, typification, collapse and exaggeration using assisted interactive tools.

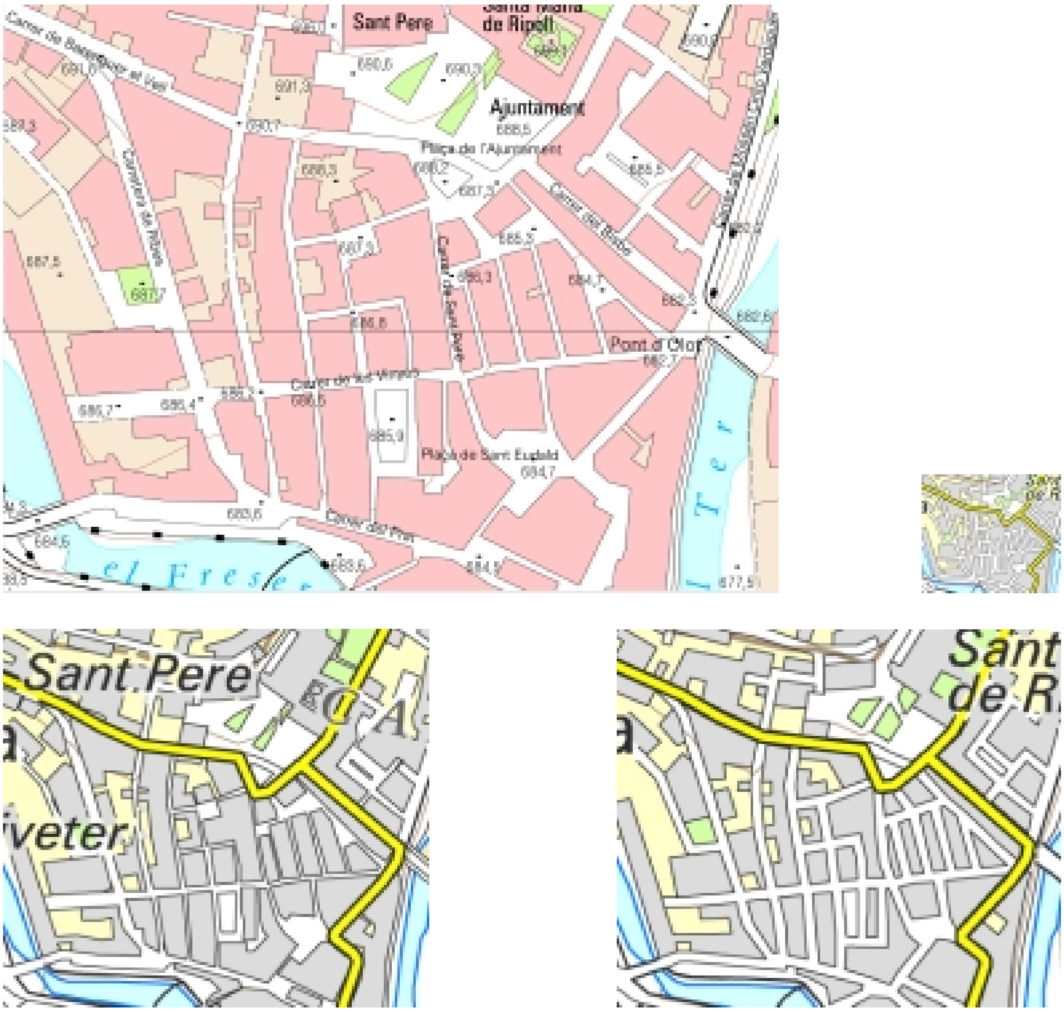


Figure 9.- The upper left figure shows, at 1:5.000 scale, the original data. The upper right figure shows, at 1:25.000 scale, the final result after automatic and manual generalization. The lower figures show in detail the differences between automatic and final generalization.

- Isolated buildings. For this kind of buildings, blocks are not defined and it is not necessary to maintain the streets. All the buildings are generalized together using the same tools than in the case of buildings inside blocks.
- Walls and containing walls. The main generalization operations are simplification, aggregation and collapse. But the hardest tasks are the selection in high density areas and the conflict resolution with built up areas or with hydrographic elements.

- Map names.
    - The automatic generalization tools include selection and cartographic scaling. The parameters for selection are geographical classification, typographical code or a combination of both. The cartographic scaling consists in a new map name placement with a new typographical code, a new intercharacter spacing and a new line spacing adapted to the smaller scale.
    - Assisted interactive generalization includes refinement of the selection, movements to avoid conflicts and changes in the typographical codes if the geographical element associated to the map name become more significant in the generalized data.

Conclusions

Comparing with the previous generalization experiences at the ICC, the new workflow has entailed two challenges: to obtain a database –not only a map– and to derive 2.5D data instead of 2D data. The main difference between generalization for a map or for a database is preserving the topological structure of the data and their attributes. The 2.5D nature of the generalized data has required a new software development and asks for a carefully interactive editing. In production, the efforts spent to the first challenge are not too many compared with maintaining the 2.5D nature of the database.

To date, six sheets have been produced using the first implementation of the workflow, and some software improvements are already in development. It is too early to publish figures on productivity, but from the first sheet on it has been consistently higher than compiling the data by photogrammetric stereoplotting.

Having started production it is now time to think about updating. As mentioned before, there are no links between the original and the generalized databases. Therefore there are only two possibilities: to update the original database and generalize again from the beginning or to update directly the generalized database. The first one guarantees the coherence between both databases but it is hardly defendable in economical terms, mainly due to the large cost involved in updating the original database in the required updating period of the generalized one. The second alternative honors the timing requirements and short term cost, but it leads to two completely separate databases that at long term would be more expensive to maintain. At this moment, the question is open.

By taking the approach described in the paper, we know that we have missed the opportunity to design and build the multi-scale topographic database of our Institute. But the challenge is there and we will do our best to achieve this goal.