NEW PERSPECTIVES IN THE GENERALIZATION OF MEDIUM-LARGE SCALE DATABASES IN ITALY

Sandro Savino, Department of Information Engineering, University of Padova,

via Gradenigo 6/B, 35132 Padova, Italy

Abstract

Recently the Italian government signed a bill that is going to affect deeply cartography in Italy. Formalizing the adoption of a single national data model for the large scale Italian topographic databases, it will have a broad range of consequences on the whole field of cartography in Italy, impacting the way that data are created and also the way that data can and will be generalized.

This paper will relate some of the latest developments in the field of cartographic generalization in Italy, discussing the consequences brought by this recently signed bill and describing the latest results of the research on this field.

The paper will briefly list the contents of the bill and describe the new mandatory national data model for topographic data that it defines, and will discuss the implications of its approval, especially from the point of view of cartographic generalization.

This paper will also illustrate the latest developments in the research on automated cartographic generalization in Italy, describing the advances in the CARGEN project that recently evolved in two main directions, expanding its research boundaries to the generalization of the 1:50000 scale and doing some more steps forward the transformation of its software into a generalization framework.

Introduction

In February of 2012 the Italian government finally published a bill, signed the 10th of November of 2011, that is of great importance in the definition of a national standard in the production of cartographic data in Italy.

The bill, hereafter called DM10/11/2011 addressed four topics, that are:

- the adoption of the national geodetic system;
- the definition of the contents for geo-topographic databases;
- the definition of the contents of the national geo-spatial data repository and of the rules for its creation and update;
- the definition of rules for the creation, documentation and exchange of digital orthophotos at 1:10000 scale.

To comprehend the full extent of such an event, it is important to understand the present situation of the production of cartography in Italy that is going to change.

As a brief recall, we can identify two main actors in the production of official topographic maps in Italy: the IGMI, or Istituto Geografico Miltare Italiano, a corps of the Italian Army acting as National MA, and each of the 20 local administration into which Italy is divided acting as Regional MA. Other agencies are entitled to the production of other type of maps (e.g. the cadastre), but their activities go beyond the scope of this article.

The IGMI is the historic NMA of Italy, its map production dating back to soon after the unification of Italy in 1861; among many other activities, it produces topographic maps in a various range of scales, from medium-large to small (1:25000 to 1:1000000).

Since 1977 the production of topographic maps of larger scales (1:1000 to 1:10000) has been officially delegated to the 20 regional administrations that constitute Italy. At that time each regional administration, called Regione, was left autonomous in deciding how these topographic maps had to be done, leading to a situation in which each RMA developed its own map (called CTR, Carta Tecnica Regionale), adopting, for instance, different codes to describe the objects on the map or choosing different sets of objects to be represented in it.

Maps produced by different Regione's were semantically different and thus not interoperable; furthermore, with the Italian reference system dividing the territory in two zones (EPSG 3003 and EPSG 3004), East and West of the meridian passing by Rome, the data of two Regione's being on different sides of this meridian could not even be geometrically matched.

The DM10/11/2011 will provide a solution to this chaotic and fragmented situation: as it will be illustrated in this paper, the bill will dramatically transform the way that geographic data is produced and exchanged in Italy and, to a certain extent, even generalized.

The DM10/11/2011

As described in the introduction, after 1977 the large scale cartography in Italy was a field with many official players: the IGMI covered the maps at 1:25000 scale, the CTR made by each Regione were at 1:10000 and 1:5000 scale while at even larger scales, 1:2000 and 1:1000, maps were usually produced, inside each Regione, by the municipalities. Each map had legends and specifications that, although similar, were different from each other, leading to a situation, from a national point of view, very fragmented and chaotic.

Already in 1996 a working group called IntesaGIS was created, formed by IGMI, the 20 Regione’s and the municipalities, with the official task to plan a set of steps leading, in the span of 6-8 years, to the creation of a national SDI. After many years of debates the outcome of their work became finally a bill, signed the 10th of November 2011 and published the 27th of February 2012.

The bill addresses three main topics concerning topographic maps: the reference system, the content and the structure.

The DM10/11/2011 defines that all the topographic data shall now be produced using the UTM-ETRF2000 reference system; furthermore, all the existing official data should be converted accordingly from the previous reference systems (ROMA40, ED50, ETRF89) using standard tools provided by IGMI to all the public administrations. The DM10/11/2011 also defines the RDN (Rete Dinamica Nazionale, translated to Dynamic National Network) listing the GPS stations realizing the National dynamic reference grid.

The DM10/11/2011 defines also that topographic data must be now structured in a geodatabase (hereafter called DBT) and provides a feature catalogue listing classes and attributes that form the database. Two subsets of features are specially defined in the catalogue, one for 1:2000, 1:1000 and larger scales and one for 1:5000, 1:10000 and smaller scales, respectively called National Core 1 and National Core 5: the bill defines as mandatory, for anyone producing official topographic data from now on, to populate all the classes and attributes that are NC1 or NC5, depending on the scale of the data produced; also in the case of updating existing data, these must be converted to the new schema.

It is left to the mapping agency to decide whether to add some other feature classes or attributes to the model, given that all the mandatory content of the DBT is present.

In the catalogue, a standard code is defined for each class, attribute and attribute value; each class and attribute is also associated to a name, composed only by uppercase letters, eventually separated by the underscore sign ("_"), no longer than 10 characters and suggesting the content of the class (eg. the class 020201 EDIFC actually stores the buildings, "edifici" in Italian).

Classes are identified by a 6-digit code and are grouped in layers and themes (each identified by a 2-digit code: the layer code, the theme code and the class number form the 6-digit class code); the following table lists the layers together with their translation:

Layer	Description
Informazioni geodetiche e fotogrammetriche	Geodetic and photogrammetric data
Viabilità, mobilità e trasporti	Transportation
Immobili ed antropizzazioni	Buildings
Gestione viabilità e indirizzi	Addresses
Idrografia	Hydrography
Orografia	Orography
Vegetazione	Vegetation
Reti di sottoservizi	Utilities
Località significative e scritte cartografiche	Names of places and labels
Ambiti amministrativi	Administrative boundaries
Aree di pertinenza	Areas of appurtenance

Table 1: layers of the DBT

Attributes of a class are identified by the class code plus a 2 digit number. Similarly, each attribute value is coded by the attribute code followed by some digits representing the value. Attribute values do not have a fixed length, as the definition uses hierarchical attribute values (e.g. 08 is the value for power plants, 0801 identifies thermoelectric power plants, 0802 hydroelectric power plants, 0803 nuclear power plants and so on).

The spatial component is an attribute of the class; it is interesting to note that, depending on the size of the object to represent or the scale of the representation, the geometry of some instances of a class might be collapsed to a lower dimension (eg. a polygon collapsed to a point).

The definitions in the catalogue also concern topology: it is defined which classes concur to form a planar partition of the space, and the spatial relations between some specific classes (e.g. Sidewalks should be part of the road area). All the constraints are formalized using the GeoUML syntax.

To assist the mapping agencies in the transition to these new standards, the working group behind the DBT commissioned to the Politecnico di Milano the development of two softwares, the GeoUML Catalogue and the GeoUML Validator. The GeoUML Catalogue is used to create a custom implementation of the DBT, adding or removing feature classes and attributes from the catalogue and producing the actual data schema, verifying its compliance with the mandatory content; the software can create the implementation as GML, Oracle SQL, PostGIS SQL and ESRI

Shapefile. The GeoUML Validator is able to test whether the data contained in a DBT implementing a data schema produced by the GeoUML Catalogue do conform to the semantic and topological constraints defined by the specifications.

As a conclusive remark, the DM10/11/2011 marks a profound evolution of the large scale cartography of Italy; the shared contents, the constraints and the tools to validate them, once entered in the production flow of every MA, will provide and guarantee high quality data that will affect positively everyone using spatial data in Italy; as shown in the next chapters, also generalization will benefit from it.

IGMI

The effects of the DM10/11/2011 will impact the way that large scale data are produced in Italy, as they now have to follow a precise set of specifications. This transformation had a reflection on IGMI, that is now thinking to change the way they produce their own maps.

The history of the IGMI dates back to the 1872, soon after that Italy had been united; as the first NMA of Italy, IGMI produced the first cartography of Italy: started in 1878, the “Nuova Carta d’Italia” took more than 30 years to be completed, with data being manually acquired at the scale 1:50000 (1:25000 for densely populated areas) and then generalized to the target scale. Nowadays the 1:100000 scale is used in the “Serie 100” maps, comprising 278 sheets, each covering 30’ by 20’ (longitude-latitude), derived manually from the maps at the 1:25000 scale.

IGMI produces topographic maps at the scales 1:25000, 1:50000, 1:100000, 1:250000 and 1:1000000. Maps are published in sheets, with one 1:50000 map covering exactly four 1:25000 maps, and one 1:25000 covering 25 CTR maps.

The 1:50000 scale is used for the “Serie 50” maps, comprising 636 sheets, each covering 20’ by 12’ (longitude-latitude). Each sheet is derived manually from the maps at the 1:25000 scale. The “Serie 50”, started in 1966, is still under completion and at the moment covers 80% of Italy.

The 1:25000 is the base scale for the IGMI topographic maps. The first collection, the “Serie 25”, was started making the “Nuova Carta d’Italia” and was completed, using direct surveying or photogrammetric techniques, in 1965. The present serie follows a 1986 revision and counts 2298 elements covering 6’ by 10’ (longitude-latitude). Its production relies also on manual derivation from larger scale data (e.g. the CTRs). In 2000 the IGMI developed a single data warehouse where to store and retrieve all the data needed to produce their topographic maps: the data model DB25 is the core for the production of the new “Serie 25DB”. Despite the use of derived data, the production of the new “Serie 25DB” is progressing very slowly and, at the time being, only a very small part of Italy is covered with the updated maps.

The DM10/11/2011 is going to have a deep influence inside IGMI: with large scale data being available for all the Italian territory in a homogeneous model IGMI decided that the 1:25000 data will be produced only by generalization. In the plans of IGMI, data should be automatically derived, thus allowing it to be up-to-date and “synchronized” with the national data.

With data quality and accuracy being guaranteed by the RMAs, IGMI role would be to integrate data coming from different sources (e.g. IGMI's own database of toponyms) and to “dress” the maps; with geometric generalization being the hardest hurdle to tackle, IGMI will develop a new legend that should ease the job of symbolizing the 1:25000 data.

IGMI's aim is that little or almost no efforts should be placed in the production of the new 1:25000 maps, with IGMI resources being then able to focus on the production of the 1:50000 maps, that will be still produced by hand thus retaining the traditional visual and quality standards of IGMI.

With the DM10/11/2011 being approved just recently, IGMI has not set yet a precise road map for this “revolution” that will change one of its most known and traditional products, but is already moving in this direction. At present, the semantic generalization process to populate the DB25 classes from the DBT ones has been already developed and a working group to study the geometric generalization process has been nominated. IGMI is fully aware of the complexity to implement an automatic generalization process, and it is ready to find a tradeoff between the available resources and the quality of the results; the outcomes of this tradeoff will also depend on the capability and possibility of IGMI to exploit the results and the know-how gathered so far by the research community: a shared wish is that the new route set by the approval of the DM10/11/2011 will foster the research, bringing researchers and industry closer together for the better of both and the users as well.

Automatic generalization

Since 2006 the Department of Information Engineering of the University of Padova is working on a research project on cartographic generalization. This project, called CARGEN, is funded by the Regione Veneto and enjoys the cooperation with the IGMI. CARGEN is one of the projects on this topic started in recent years in Italy, showing an actual interest of the Italian mapping agencies in it; among these projects, CARGEN is the most long-standing and the most advanced, with results acknowledged by the international research community.

Most of the research on cartographic generalization in Italy is focused on deriving the IGMI 1:25000 and 1:50000 scale maps, that are commonly used as back drop maps in many different activities as regional planning and forestal or geological surveys; with the IGMI struggling to keep these maps up to date, it is of great interest to derive them from the regional data.

The DM10/11/2011 will affect the generalization of large scale data in Italy, as it will be discussed next in this chapter. The most recent advances of the research on generalization, brought by the on-going CARGEN project, will be illustrated in the last two sections of this chapter.

Impact of the DBT

So far, due to the fragmented situation characterizing the Italian cartography, the interest on generalization translated to “local” projects, usually customized to a specific data model and thus confined within the boundaries of a Regione: till now there has been neither the chance nor the will to start a national project investigating this research field. Now things may change, as the newly adopted bill creates the opportunity to develop a generalization process that could be applied to data coming from any Regione.

As it was shown in the previous chapter, IGMI is already moving in this direction for the generalization of the 1:25000 scale; similarly, in the last months the CARGEN project made some tests, trying to apply the generalization process originally developed for the data of the Regione Veneto on data coming from other RMAs. These tests were intended to verify data interoperability: the good results show in general a wide compatibility among all the data. Some minor issues were present, as naming conventions and not matching attributes values, but they can be justified by the fact that the data used in the test had been produced before the approval of the bill and thus referred to slightly different drafts of the specifications and had no mandatory schema to abide to.

A more formal test will need to be done once the first data following the new specifications are made; surely though these early tests show that semantic generalization could be considered not any more a hurdle in the generalization process of Italian data.

With semantic generalization being set once for all, more resources could investigate the issues of geometric generalization: the development of a complete generalization process for all the Italian territory will require to investigate and design solutions for the different geographic features that "locally" characterize the country (e.g. canals in the lagoon of Venice, terraces on the reliefs of Liguria or Roman ruins in Rome).

Developments of the CARGEN project

Generalization of the 1:50000 scale

In the last three years, the scope of the CARGEN project has been extended, including the generalization to the 1:50000 scale. This new branch of research runs in parallel with the research on the 1:25000 as in fact these two researches overlap and the algorithms developed for the larger scale in most cases can be applied, with different thresholds, to the smaller one.

If the implementation of the generalization process to the 1:50000 scale benefits from the results achieved previously at the 1:25000 scale, designing this process still required a good deal of work: with the aim to generalize the IGMI 1:50000 scale maps, it was necessary to study the specifications, to design a data model for a database storing the generalized data and to find the mapping schema between the data at 1:50000 scale and those at larger scale, finally implementing the model generalization process.

The generalization process required also to develop some new techniques: with the research still on going, the first topics investigated were the generalization of buildings and roads, with the aim to improve the capabilities and the results of the algorithms developed for the 1:25000 scale.

The building generalization algorithm, performing a context analysis, detects urban and not-urban areas and then applies two distinct generalization strategies: in densely urbanized areas, it acts trying to balance the black and white ratio of the building blocks, merging buildings and working on the voids among them, in sparse settlements, it works on groups of buildings through typification (see Regnault, 1996) or on single isolated buildings performing selection or enlargement.

The road generalization algorithm works both on the geometric and semantic data. The algorithm, working mainly on the geometries of the road network, removes short dangling roads, snaps together t-junctions that are too close, collapses dual carriageways roads and detects and re-draws narrow road bends. The algorithm computes also a set of metrics (see Jiang et al., 2002) that allow to perform a re-classification of the data; this results in the harmonization of the road type and to a more accurate detection of the most important roads that is used to prune the network.

Figure 1: generalization of urban areas with different densities

Toward the creation of a framework

The generalization process developed originally in CARGEN was a console-run program relying on a set of algorithms, each dealing with a specific generalization task. The code relied heavily on the DBMS and specifically was dependent to Oracle Spatial. Working on the project, it was found that this architecture did not provide a good platform to develop and run the software: execution times were slowed by the continuous transactions with the DBMS server, and mixing Java and SQL made development more error prone and debugging more cumbersome.

In 2010 it was decided to develop a new architecture for the software: the new code should be runnable also in a GUI environment and should not be DBMS dependent. In the following two years, the present version of the software took form. A GIS was chosen, to be the GUI environment for the software, with OpenJUMP being finally selected due to its simplicity and integrability with the selected IDE Eclipse, and the library originally handling the connection with Oracle Spatial was transformed into an abstraction layer in order to extend the capabilities of the software to deal with different DBMS. Custom libraries were then developed, enabling this layer to work with Oracle Spatial, PostGIS, and a virtual DBMS implemented in RAM using OpenJUMP, JTS classes and custom developed code.

All the latest algorithms developed in CARGEN can now be run as plug-in of OpenJUMP, thus easing debugging and the visualization of data; thanks to the libraries developed, the code can be run on one of the DBMS supported, or on any vector data loaded in OpenJUMP.

While the conversion of the old algorithms is still on-going, the new architecture allowed us to develop some new tools, that are the first bricks in the construction of a generalization framework. As a future perspective, the generalization process, that now is simply a not-yet-fully-completed porting of the old code into this new architecture, will benefit further more of the framework that is being built, and will finally migrate from research to an actual tool that could be useful in real production.

Visualization plug-in

One of the first plug-in developed to took advantage of the new GUI environment, was a tool able to “dress” the data with a symbolization that resembled that used on the IGMI maps, that are used as reference for the generalization at the 1:25000 and 1:50000 scales. This plug-in extends the capabilities of OpenJUMP to display only simple styles, allowing to draw complex symbols. Styles are read in the official IGMI documentation and coded by hand in a csv file following a specific syntax, that is based on the SVG standard, but it allows to define styles as composed by multiple styles. The plug-in can apply one specific style to a single layer in OpenJUMP, or automatically apply the correct IGMI style to one or multiple layers, using a specific attribute and the layer name to decide which style to apply.

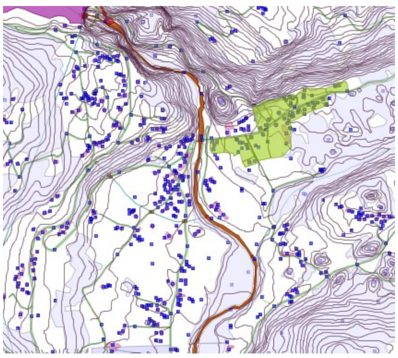
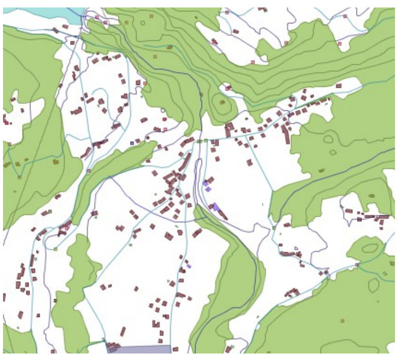
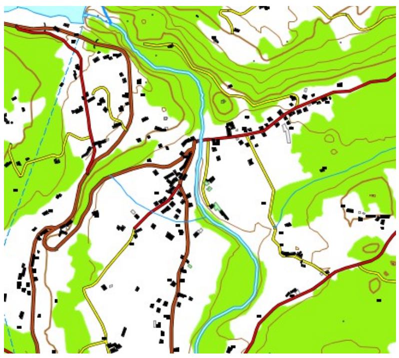

Figure 2: sample of original data, generalized data and generalized data styled

Logging and backtracking plugin

Logging is a very important activity when dealing with complex projects, either during development and debugging, and while reporting the results of the run. This is even more true working with generalization, when the results are the outcome of a long sequence of algorithms. The logging library developed is based on the basic Java Logger classes, but has been customized to handle special messages that the developer can insert in its code to signal the transformations applied to an object during its generalization. Inserting this functionality in the logger, the developer can, with only one line of code, inform the user about the progress of the operations and gather information that can be used, once the process is completed, to create a detailed final report of what has been done and eventually to backtrack all the steps of the process.

The logger is also integrated with the GUI environment, allowing to do a visual backtracking: drawing on screen the previous steps of the process allows to investigate more effectively what happened during the generalization. The logger has been also provided with some special functions that can be used to interact with the GUI also when debugging step-by-step, thus tackling the problem of the GUI being by default unresponsive during this activity.

Partitioning plugin

While memory management seems a least problem with the increasing resources enjoyed by the present computers, partitioning a dataset to generalize is still a question to be answered, especially when data are not processed on an actual DBMS on a server, but on a virtual one in the RAM of the client.

The tool developed for such a task starts from the assumption that given a generalization algorithm it is possible to define a partition of the space that matches or reasonably approximates the results of the generalization on the unpartitioned dataset. This assumption is based on the fact that the actual implementation of any algorithm has a defined “horizon” or “range” beyond which it has no effect on the data. Depending on the algorithm, this “range” may or may be not an actual distance: in general it is a set of information that is necessary to the algorithm to compute an informed choice on how to generalize the data; we called it Minimum Necessary Information Unit (MNIU).

The partitioning plugin actually leaves to each algorithm to define its MNIU (and thus to the algorithm's developer), but acting both as a library and as an interface, it provides functions to partition the dataset before processing, to handle parallel and sequential processing of the tiles, to sew together the pieces afterward and to gather the MNIU. For this latter task there are two main approaches: one is based on actual distances, it differs “viewing” and “acting” ranges, and it is useful to partition datasets of discrete objects as buildings; the other is useful for “continuing” objects, as networks, and it is based on a look-ahead functionality.

Conclusions

The approval of the DM10/11/2011 set a giant step forward in the field of cartography in Italy. The new legislation is going to transform the way data is produced and exchanged and, with the working group behind these specifications still at work, it is hoped that it will be the first of a set of steps that will reform even deeper and further all the Italian SDI.

With no data yet produced following these specifications, it is still uncertain how long will it take for all these changes to be correctly adopted by the MA; although there are not any results yet witnessing how good these specifications are nor how they will actually effect the production and exchange of cartographic data, early tests are very positive and indicative of a drastic improvement over the previous situation.

At the same time, the research on cartographic generalization carried on in the CARGEN project is advancing: thanks to new tools, algorithms are more easy to develop and test, over time, the dimension of the code base is increasing and its quality improving, thus allowing to harness effectively more complex problems; a new process, for the generalization to the 1:50000 scale is under development, leading to the investigation of new solutions and to the improvement of old ones and the framework that is being built will act as a container, allowing to capitalize what has been achieved so far.

The definition of a single data model for large scale topographic data in Italy, the intent of the IGMI to derive the DB25 from this model and the improving results of the research are opening new perspectives on cartographic generalization in Italy, making these exciting times for the research in this field.

Bibliography

DM10/11/2011 http://www.digitpa.gov.it/fruibilita-del-dato/dati-territoriali

Belussi A., Liguori F., Marca J., Migliorini S., Negri M., Pelagatti G., Visentini P. (2011) Validation of Geographical Datasets against Spatial Constraints at Conceptual Level, Proceedings of The Urban Data Management Symposium, Delft, The Netherlands, Sept. 28-30, 2011.

Jiang B., Claramunt C. (2002) A Structural Approach to Model Generalisation of an Urban Street Network, 5th AGILE Conference on Geographic Information Science, April 25th-27th, Palma, Spain

Regnault N. (1996) Recognition Of Building Clusters For Generalization, Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, The Netherlands, 185-198

Renard J., Gaffuri J., Duchene C. (2010) Capitalisation problem in research - example of a new platform for generalisation: CartAGen 2010, 13th ICA Workshop on Generalisation and Multiple Representation, 12-13 September, Zurich

Savino S., Rumor M., Congiu S., (2011) Automated cartographic generalization in Italy. In E. M. Fendel, H. Ledoux, M. Rumor, and S. Zlatanova (Eds), The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38-4/C21, 109-114, doi:10.5194/isprsarchives-XXXVIII-4-C21-109-2011 ISSN 1682-1777