Digital transformation in topographic databases

Gema Martín-Asín López [a], Lorenzo Camón Soteres [a], Gonzalo Moreno Vergara [a], Andrés Arístegui Cortijo [a]

[a] *National Geographic Institute of Spain - General Division for Cartography and Land Monitoring, gmartinasin@mitma.es, lcamon@mitma.es, gmvergara@mitma.es, aaristegui@mitma.es*

Abstract: The increasingly widespread implementation of databases with geographical component, as well as the impregnation of geolocation culture, is driving a transformation in the storage, management and exploitation of geospatial information. Real-world elements go from being modeled as mere geometric representations, with just cartographic purposes, to be features with their own entity. Unique identifiers and lifecycle management are assigned to these features, allowing interactions between feature instances from different databases, that is, facilitating digital transformation and, therefore, increasing exponentially the exploitation possibilities.

In this regard, the National Geographic Institute of Spain (IGN, by its Spanish acronym) have implemented several processes in its National Topographic Database, such as the connection with the cadastral information, in order to take advantage of its updates and give feedback to improve cadastral data; or the link with the information, in addresses form, provided from different public administration, that is processed to geolocate features in the topographic database. Likewise, work is being done in order to implement new processes that allow linking with other data sets.

These processes, in addition to reusing information produced by different public administrations, constitute an advance towards the objective of geospatial information databases continuous updating.

Keywords: topographic databases, linking data, changes detection, reuse of information

1. Introduction

A topographic database constitute an abstraction and digital modeling of a country, or an area, that collects a variety of topographic features such as elevations, communications, hydrography, populated sites, etc., through vector data structures (sequence of geographic coordinates) that store two or three dimensional geometries using points, lines and polygons. Topographic databases are specially designed for the management and analysis of geographic information through the use of Geographic Information Systems (GIS), but they also allow the generation of a variety of cartographic products, as well as the transformation to other data formats, such as the ones used in Computer Aided Design (CAD).

Thus, the National Topographic Database of Spain (BTN, by its Spanish acronym) constitutes the highest resolution, three-dimensional Geographic Information System, which covers consistently and homogeneously the whole country with topographic information of general purpose. At first, it is the result of evolutions applied to the 1:25,000 National Topographic Map (MTN25, by its Spanish acronym) digital information, which have two main milestones:

- 1:25,000 Numerical Cartographic Database, that takes MTN25 geometries and encodes its elements for their exploitation through computerized processes. It lacked a third dimension and inherited the cartographic criteria applied to the MTN25, so that part of the elements were not in the true position and, therefore, it was very difficult to use it for a different purpose than cartography production itself.

- BTN creation, in 2005, based on the photogrammetric restitution files (data capture from stereoscopic pairs), with approximate scale 1:12,000, on which cartographic drafting processes have not been applied, what allows to obtain a database of greater resolution and with three-dimensional information. In addition, the topological and semantic continuity (analytical continuity in coordinates and spatial relationships, as well as their attributes) of features throughout their entire extension is ensured. Increasingly information and identifiers are provided to enable smarter and more versatile exploitation.

Figure 1. Perspective view of BTN information.

The production process of the National Topographic Map is thus reversed, which is automatically generated thanks to the versatility of being able to

exploit the spatial relationships between elements to make explicit (and automate) the criteria for selection, aggregation, simplification, displacement, etc., which the cartographer has traditionally applied in a tacit way to achieve a concrete and expressive representation of reality.
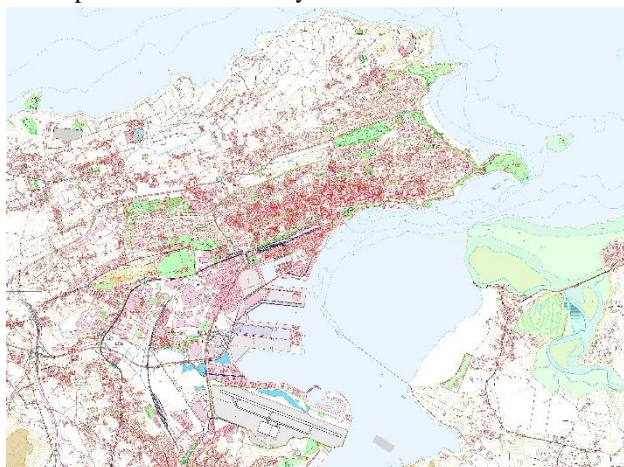

Figure 2. General view of Santander city and surroundings in BTN

1.1 BTN updating mode

Since the first BTN version was consolidated, photogrammetric restitution capture is abandoned. It allows lower production costs, since it does not require any specialized hardware or particularly qualified operators. In addition, it is easier to carry out the work solely using GIS software, which makes it possible to update directly on a database model and introduce a set of topological-semantic rules to formalize the so-called "tacit knowledge" into "explicit knowledge", similar to other international projects such as MGCP (Multinational Geospatial Cooperative Program) of NATO. Indeed, many quality controls that were carried out manually are now done in terms of spatial analysis, for example, a road cannot pass over a building, unless it is on a bridge or in a tunnel. Ultimately, this made it possible to automatically ensure BTN integrity and spatial coherence.
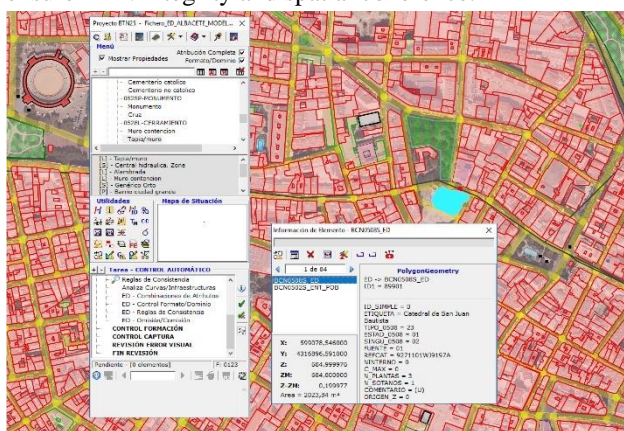

Figure 3. BTN controlled work environment.

1.1.1 *Main data sources*

Once the stereoscopic restitution is discarded, the geometric support and completeness of topographic elements are aerial orthophotographs and digital elevation models generated through the National Aerial Orthophotography Plan (PNOA, by its Spanish acronym), key products to carry out the updates, which guarantee a metric quality more than enough for BTN purposes.

In turn, BTN uses many other data sources to provide its features with the required semantic information (attributes) and to guarantee their completeness. Among them, it is worth mentioning the basic geographical gazetteer of Spain; itineraries that are compiled in the Nature, Culture and Leisure project, also from IGN; as well as sources provided by different public administrations such as Protected Natural Areas in Spain; the inventory of dams and reservoirs; wastewater treatment plants; the beach guide; the administrative registry of electrical energy production facilities; the assets of cultural interest registry; the Town Halls registry; the educational centers registry; the hospitals national catalog; buildings and constructions of national cadastres, and a long etcetera.

Thus, in its work protocol, the BTN makes the entire series of tasks explicit, indicating the sources of information to be consulted and the processes to be carried out to ensure the correct execution of the work.

1.1.2 *Updating strategy*

The strategy used from the beginning to address the updates has been working by territory blocks (for example, by provinces) until obtaining complete coverage of whole country. This methodology has proven to be very productive for massive tasks. However, once a product with a high degree of updating is achieve, working by blocks is no longer so advantageous, since important changes may have occurred in one area while you are working on another. Here arises the need to move to an update by changes mode strategy.

2. Towards update by changes

Thus, another approach must be used that takes advantage of the entire existing digital ecosystem to deduce changes over the country in the most continuous way possible without limitations on the opening and closing times of other projects. One may say that the objective is that "the database clock is as close as possible to the territory clock".
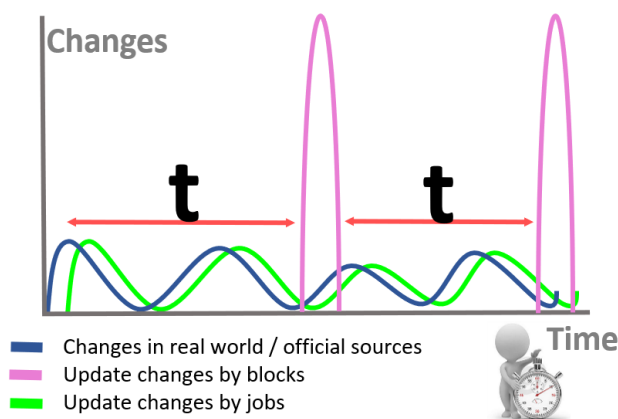
Figure 4. Updating by changes vs by blocks

To this end, the update by changes program was created in 2017, divided into three projects:

- Changes Detection. The first thing to know is what changes are taking place in the country. Here a wide range of possibilities opens up and the so-called changes generating engines come into play.

- Work orders management. When upgrading across large blocks of territory, work units are relatively easy to manage. However, changes in reality happen more or less randomly and affect variable areas, so it is necessary to package the updates in work orders, as well as manage the assignment, execution, reception and validation of these work packages in an agile and controlled way.

- Integrated and controlled update environment. To incorporate all these changes in the topographic database, it is important to use a work environment coordinated with the work orders management, to guide the operator in the tasks to be carried out, and that allows a joint update of all thematic areas covered to ensure consistency of data. In addition, although certain themes present a greater dynamism of changes, such as transport and buildings, in most cases they affect objects from other areas. For example, the construction of a highway or an urbanization will surely affect transport routes, buildings, hydrography, elevations, etc. The strategy, then, is to take advantage of the greater dynamism and impact that certain geographic objects produce to trigger the update in the rest of the themes.

3. Change's generating engines

In the wide range of possibilities to detect changes, the following classification can be made:

Social media detection engine. It is clear that if, for example, a road is built, there is a minimum public information offered in an official channel, an announcement in an official bulletin, an interest group that may disagree or agree with the project, etc. This is

called "digital trace" and in this regard some small computer programs (so-called *bots*) are being implemented to track a multitude of sources on the Internet and then filter and classify what has changed, what is changing, what will change and where.

Image change detection engine. The detection of changes between successive images over time (for example, from different years photogrammetric flights or from satellite images) was a great challenge that could only be solved with guarantees using artificial intelligence. In a first test carried out in 2017, a reliability of more than 95% was achieved in changes detection in buildings and communications between images. Currently, work in this line continues to be developed and taking a step further, focusing on feature extraction and its comparison with the ones in the database, in order to detect errors or lack of updating or, even, to automatic update of certain elements in the future.

Change detection engine between existing geographic information. The huge availability of data, both public and private, makes it possible to detect changes with respect to the existing data in our topographic databases. This is a clear example of what the term digital transformation implies, especially when the information managed by the different organizations is properly organized in databases, uses unique identifiers for its feature instances, incorporates information about its life cycle and is accessible through standardized protocols, since it allows to relate feature instances of a database with those of another, and to determine when there are additions, deletes and updates.

However, despite the many possibilities, each of these approaches requires considerable research and development work until it is productively implemented.

4. BTN implemented processes

Important advances have been made in BTN to implement the third changes engine mentioned in the previous section, meaning, relying on other existing geographic information and linking to it. The key here is both knowledge of the official reference information existing in public administrations (previously identified, classified and hierarchical) and other useful unofficial data sources. Based on these sources of information, processes have been implemented that allow detection and, sometimes, automation in the update, some of which are shown below.

4.1 Linking with cadastral information

The objective is to connect topographic and cadastral information and reuse the efforts made in updating one data set to take advantage of them in the other. The process has been divided into two phases, a first (from 2016 to 2019) in which BTN information on buildings and constructions is prepared and the connection with Cadastre is established through the cadastral reference, and a second that takes advantage of the life cycle defined in the cadastral database features to act on the related elements in the BTN. It should be taken into

account that the buildings and constructions produced by the General Directorate of Cadastre (DGC, by its Spanish acronym) cover the entire country with the exception of Navarra and the Basque Country, which have their own systems for cadastral management, with which similar integration jobs have been approached.
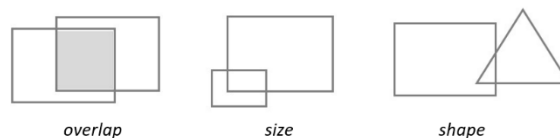
In the first stage, after the cadastral data analysis, a set of processes was designed using ETL (Extract, Transform and Load) tools that allow the identification of common geographic objects among the cadastral data and BTN. In this first integration, the two input sources are conceptually different: the elements from BTN are groups of buildings or blocks, while those from the Cadastre are divided by cadastral reference, forming individual buildings, although adjacent.
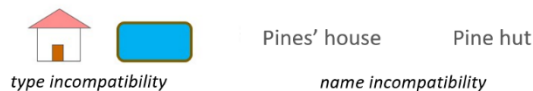

Figure 5. Conceptual differences between buildings in the first stage.

Thus, conflation rules (homologous features identification from different data sources) play with relationships based on overlapping surfaces, shapes and similarity of areas, multiplicity relationships (1:1, 1:n) and type compatibility, forming a decision parameters matrix. The process automatically assigns a geometry and a set of attributes. In the event that minimum values are not met in the decision parameter matrix, a conflict is returned and an operator must review using orthophoto. The percentage of success in the automatic assignment was in the order of 70%.

Figure 6. Conflation rules in the first stage.

As a complement to all this process, an exhaustive review of the territory was carried out with an orthophoto to correct omissions, commissions, classification problems and positional errors. All the conflicts found, of the order of three million, were reported to the Cadastre so that it could integrate them into its incident management system.

The result of this first project stage allowed to achieve the following strategic objectives:

- Semi-automatic updating of the BTN buildings, obtaining as a result the cadastral geometry together with the BTN semantics (attribution).
- Analytical linking of the objects defined in the two data sets, through the cadastral reference.
- Establishment of a framework for interoperability and cooperation between the IGN and the Cadastre.

The goal of the second stage is to establish the mechanisms for BTN buildings and constructions incremental updating. The general work flow is as follows:

- From the Cadastre`s Web Feature Service (WFS), buildings variations that have occurred since the last update are obtained (taking advantage of the life cycle defined by the Cadastre).
- These variations between two dates are subjected to a new conflation process with new criteria based on the translation, segregation and grouping of buildings (the geometries of the two entrance sets are already conceptually equal, after the first stage).
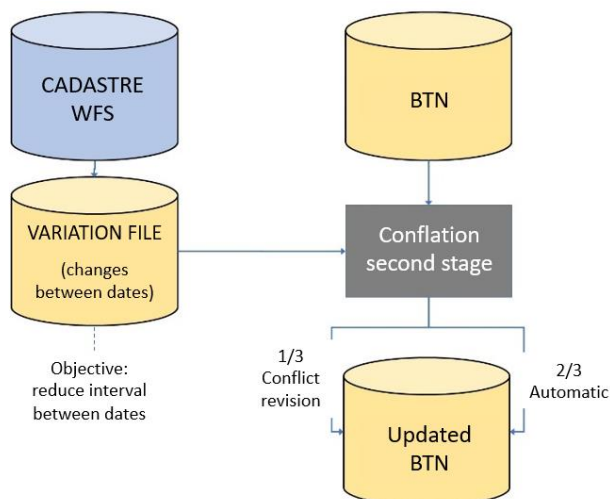
Figure 7. Second stage general scheme.

- Visual revision, with orthophoto, of the conflictive cases in which an automatic assignment is not achieved. Currently, the order of 1/3 of the variations must be reviewed, for which a specific functionality has been integrated into the work environment that facilitates conflicts analysis and resolution.

Figure 8. Specific functionality to solve conflicts.

The result of this second project stage allows to achieve the following strategic objectives:

- The reuse of official data of the Administration served through the Spatial Data Infrastructure of Spain standard services.
- Taking advantage of the high updating degree of cadastral data to continuously and incrementally improve the BTN buildings, minimizing human intervention.
- Feedback for Cadastre, providing a continuous source of verification of the cadastral geometry.

Thus, this project is a clear example of collaboration between public bodies that undoubtedly strengthens the necessary cooperation between administrations with the aim of serving the general interest.

4.2 Geolocation processes

The geolocation processes developed allow to extract the location of services or facilities from the address lists provided by official sources and assign them to BTN feature instances, providing them with the most significant attributes such as the name, the code (with the original identifier of the service or facility) and type.

Figure 9. Illustration of the geolocation process.

The process steps can be summarized as:

- Downloading the address lists from official sources, then composing the correct address and correcting spelling mistakes.
- Applying different geocoders (Cartociudad, WFS cadastral Addresses, etc.) in order to obtain points for each geocoder and address.
- Determination of the main use to which the buildings are dedicated. This main use is extracted from the alphanumeric information of Cadastre and is assigned to BTN buildings through the cadastral reference.
- Crossing of these points obtained from the different geocoders with BTN buildings.

In this crossing, some parameters and priorities are defined depending on the geocoder used, distance from the points to the buildings, concurrence of several points, main use of the building (this gives robustness to the process, for example, it prevents automatically assigning a town hall to a residential building), etc. An indicative value of safety is obtained from assigning the service or facility to the building and based on this value, an automatic assignment is made or sent to a checklist for an operator to resolve the conflict.
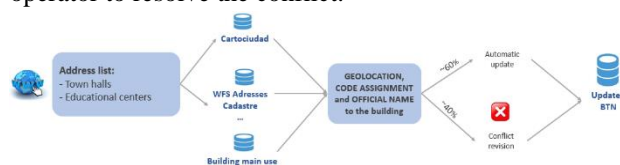

Figure 10. Geolocating process general workflow.

Currently, these processes are being applied to update town halls and educational centers of BTN, being able to extrapolate to any other topic for which addresses are available.

5. Conclusions

There are innumerable advantages of using databases to store, manage and exploit geographic information. This

fact is not new and it is evident with the examples that have been shown in this article, but even today, it should be noted because of the confusion that sometimes occurs between the terms cartography, databases with geographic information and data viewers.

One of the difficulties for the efficient exploitation of this entire information ecosystem is the existing disparity in the way of accessing the available data sources. For example, in some cases there are such databases with unique identifiers, life cycle management and access through standardized protocol. But in other cases what are provided are web viewers that do not allow automated access; alphanumeric lists of elements, or maps in image format that can only be exploited visually.

Therefore, progress must continue to be made in all public administrations and organizations that manage information related to the territory, both in distribution, as well as in the way of updating and linking with data from other organizations, to allow a more intelligent exploitation of information and, ultimately, continue to delve into digital transformation.