ICGC MRDB for topographic data: first steps in the implementation

Blanca Baella*, Dan Lee**, Anna Lleopart*, Maria Pla*

* Institut Cartogràfic i Geològic de Catalunya (ICGC), **Esri
Email: blanca.baella@icgc.cat; dlee@esri.com; anna.lleopart@icgc.cat; maria.pla@icgc.cat

1. Introduction

The Institut Cartogràfic i Geològic de Catalunya (ICGC) is a public law entity of the Generalitat de Catalunya, the Autonomous Government of Catalonia, devoted to the tasks related to Geodesy, Cartography, Geology and Geophysics in Catalonia as well as in promoting the studies to improve the production related workflows among others. During the past few years the ICGC has been analyzing the implementation of a MultiResolution DataBase (MRDB) for integrating the ICGC topographic databases. In 2012 a report [Baella et al, 2012] was presented in the 15th ICA Generalization Workshop in Istanbul, describing the situation of the ICGC topographic databases and the first approach to the design of this ICGC MRDB data model.

The report also included the expected steps for generating and populating this MRDB database. At the beginning it should contain data from the current Topographic Database at 1:5000 scale (BT-5M), 1m accuracy and 3D data, and the Topographic Database at 1:25.000 (BT-25M), 2,5 m accuracy and also 3D data. Data from databases at other scales, such as the Topographic Database at 1:50.000 scale or the Topographic Database at 1:250.000 scale, would be included later.

The implementation of the ICGC MRDB database includes as main tasks the setting up of the data model according to the conceptual model, the migration of the existing data to the new data model, the migration of the photogrammetric data collection from a CAD system to a GIS system based on a Data Base Management System (DBMS), the integration of generalization tools into the new system and the development of the tools for the management of the MRDB relationships. Because the BT-5M and the BT-25M are separated topographic databases, both with full coverage of the country, the first link establishment between the feature instances at different resolutions will be done using feature matching techniques. For later updating processes on the higher accuracy data, the lower resolution data and the links will be managed in the updating or generalization process.

During the last year the implementation of the MRDB data model has been started, the migration to a GIS photogrammetric production system has been completed and fully in production. Some proofs have been done to help the first establishment of the MRDB links between the BT-5M and the BT-25M data; commercial conflation tools from Esri are being tested for performing the matching between features in the two different scales.

This paper will explain the experiences on the data model implementation and the migration to a GIS production system. It will then describe in depth the tools and the results of the conflation tests using Esri conflation software. And finally, it will give some details about further steps.

2. Data model implementation

The ICGC MRDB data model is based on one single schema with linked data, where one feature belongs to one single resolution and has a link to one or many features of the other resolution (Figure 1).

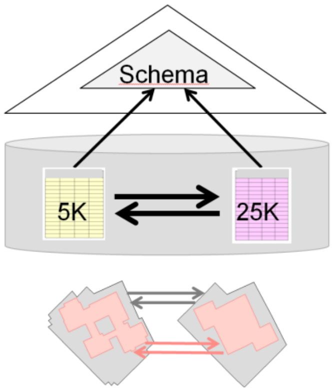

Figure 1. ICGC MRDB data model.

The data model is based on *simple features*, which are defined by a part of a real world phenomenon with common attribute values (Figure 2).

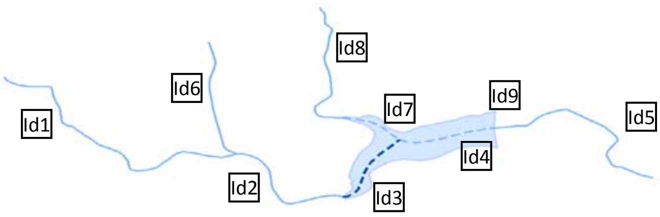

Figure 2. Each stretch of the river network with common attributes between two intersections corresponds to a feature instance.

In this version of the data model complex features, for example the entire river, have not been considered, not because of technical issues for their implementation but the required information to identify the simple features that compose a complex feature has not become fully available.

Each feature instance holds an identifier that is unique and persistent along its whole life-cycle. Moreover, each feature is characterized by spatial, temporal, metadata, and thematic attributes. The spatial attribute carries the geometric representation of the feature. The life-cycle information describing temporal characteristics makes feature versioning possible, that is, the propagation of changes across scales and the propagation of changes to related databases. The metadata attributes contain the information related to lineage and quality. Finally, features have a set of thematic attributes to characterize them.

The data model is being implemented in an Oracle Spatial Database. After the schema has been created, the data coming from the BT-5M and BT-25M databases has been fully migrated to the new data model and loaded into the schema. The structure for storing the links will be implemented in the near future.

3. Migration to a GIS photogrammetric data capture system

The new ICGC GIS photogrammetric data capture system is based on Oracle as DBMS and on a commercial GIS in combination with a photogrammetric software. After testing different combinations of GIS software and photogrammetric data capture tools, GeoMedia and ImageStation Stereo for GeoMedia (ISSG), from Intergraph, were selected for their good performance, excellent ergonomics and easy integration with the ICGC workflows.

The architecture of the system is based on a central database on Oracle Enterprise, where data is consolidated, and several disconnected photogrammetric data capture working places using Oracle Express.

The ICGC long experience in implementing production workflows has shown that it is necessary to customize the commercial software to improve productivity. The ICGC application extends the functionalities of GeoMedia and ISSG, solves the 3D shortcomings of GeoMedia and Oracle Spatial and reduces the vector data visualization problems of ISSG. Moreover, the system optimizes the data capture and management, improving different aspects related to the feature catalogue management, providing tools for database management and ensuring the data quality by application design (Figure 3).

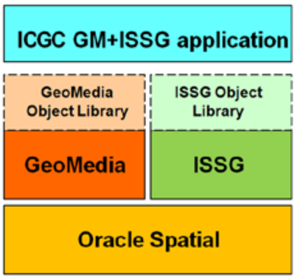

Figure 3. Components of the ICGC system.

The system is fully in production and the training of the team is completed. The cartographers show a high level of satisfaction with the new environment and the achieved productivity is so far similar to that obtained with the old CAD system. Considering that the data model is richer than the old one and that it will allow a wide range of exploitations, the system migration is a successful experience.

4. MRDB link establishment: using Esri conflation tools and geoprocessing tools

The BT-5M and BT-25M data overlap (covering the same geographic area) and contain common and uncommon features. It is necessary to match the common features and establish links between corresponding features in the two resolutions for future updating and other purposes. Knowing that feature matching is fundamental to the conflation tools, such as the ones developed by Esri (Environmental Systems Research Institute, Inc.), an effort has been made to test the potential use of the conflation tools and other geoprocessing tools in workflows to help establish links for the MRDB features. The idea is straightforward: find feature correspondences and assign a unique ID to each pair or group of matched features. The goal is to analyze the level of automation of the first link establishment process. A brief introduction of the conflation tools and some test details are given below.

4.1 Conflation tools and feature matching

Conflation is a process to reconcile multi-source datasets and improve data quality and usability. It involves identifying corresponding features (known as feature matching), making spatial adjustment and attribute transfer, and ultimately unifying datasets with the optimal spatial accuracy, completeness, consistency, and integrity. In ArcGIS Desktop 10.2.1 a Conflation toolset containing five tools was added to the Editing toolbox, along with another closely related tool, Detect Feature Changes, in the toolbox structure shown in Figure 4. Most of these tools currently support linear features. Case studies on using these tools in workflows to accomplish typical conflation tasks were presented in a recent paper (Lee, Yang, Ahmed, 2014).

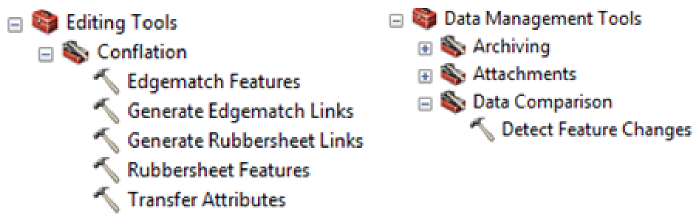

Figure 4. Conflation tools in ArcGIS Desktop 10.2.1.

The feature matching (FM) technique used for overlapping datasets is based on the analysis of proximity, topology, pattern, and similarity. The technical details and more discussions are beyond the scope of this paper and can be found in another recent paper (Yang, Lee, and Ahmed, 2014).

The two FM based tools relevant to this paper are Transfer Attributes (TA) and Detect Feature Changes (DFC). Depending on quality, complexity, and similarity of input features, FM can be highly accurate. Certain pre-processing (not discussed in this paper) can be especially helpful to obtaining great FM accuracy, for example validating geometry and topology (avoiding redundant vertices, unnecessary gaps, overlaps, and crossing lines). Post-processing is generally expected after using these tools.

Each FM based tool optionally produces a match table (Figure 5), storing match information including source and target feature IDs (SRC_FID and TGT_FID), unique match group IDs (FM_GRP), match relationships (FM_MN), and matching confidence level (FM_CONF). Conveniently the FM_GRP values can serve as links between corresponding features in BT-5M and BT-25M.

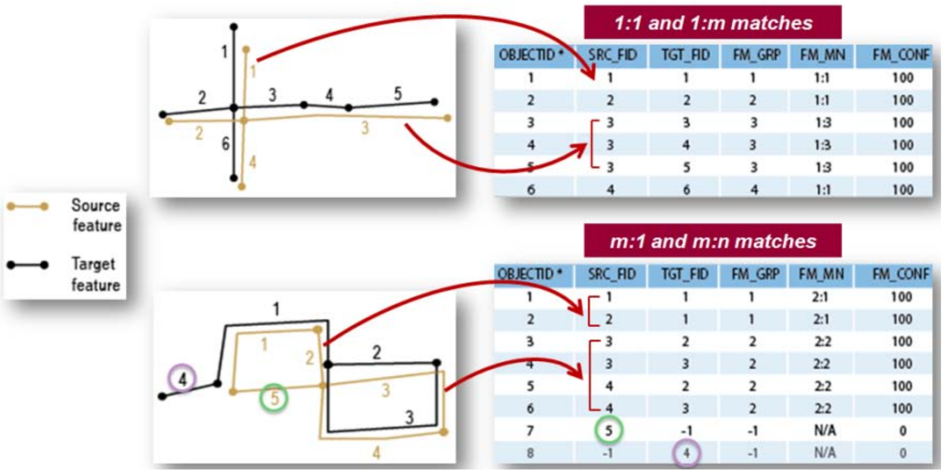

Figure 5: Information in match table.

4.2 Test scenarios

The initial testing focused on three scenarios; each required establishing links (unique IDs) between corresponding features of the two resolutions.

Scenario A: Establish links between road centerlines of BT-5M and BT-25M and transfer classification attributes from BT-25M to BT-5M road centrelines

The road centreline test datasets were BT25_VIA1 (4013 features) and BT5_VIA1 (3877 features); see Figure 6. The main goal was to obtain unique IDs for corresponding centrelines in the two datasets. Meanwhile, a set of road classification attributes were also transferred from BT25_VIA1 to BT5_VIA1.

Figure 6: Road centrelines – data area and close-ups.

The workflow consisted of two steps: conflation and evaluation, and post-processing. They are presented below.

a. **Conflation and evaluation (automatic)**
   A geoprocessing (GP) model (see Figure 7) was built to establish links between road centrelines of BT-5M and BT-25M and to transfer classification attributes from BT-25M to BT-5M road centrelines. It also evaluated the results and produced information that could facilitate post-processing.

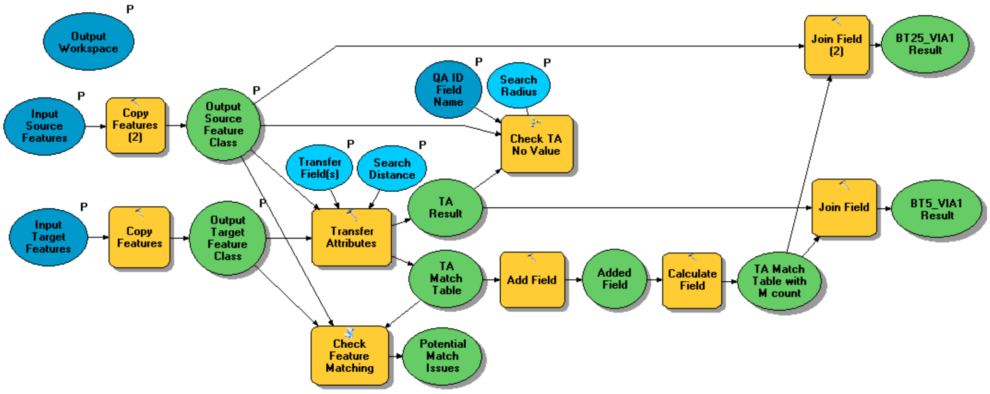

Figure 7: GP model for link establishment, attribute transfer, and evaluation on VIA1s.

The TA tool played dual roles:

- To generate the match table containing the FM_GRP values. In the end matched source and target features received their common FM_GRP values as links; see Figure 8.
- To transfer the desired attributes, including the feature unique identifier, from lines in BT25_VIA1 (as source features) to corresponding lines in BT5_VIA1 (as target features); see Figure 8.

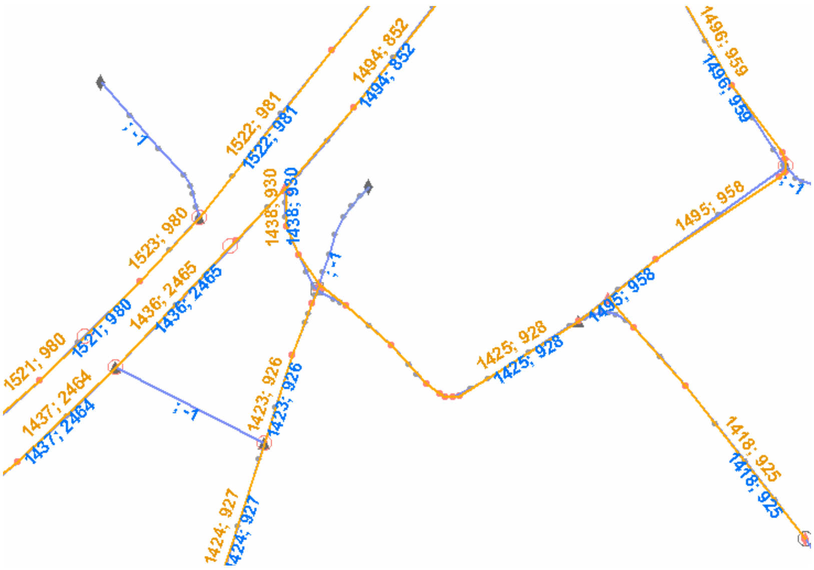

Figure 8: Resulting VIAs of BT25 (orange) and BT5 (blue), labelled in the order of "feature unique identifier; FM_GRP" values.

The evaluations of the results were done using conflation supplemental tools and other geoprocessing tools, focusing on three aspects:

- <u>Potential FM issues</u>: matched features were analysed against a few logics and questionable cases were flagged for review.
- <u>Potential missed matches</u>: the unmatched BT5 features (which didn't get transferred attributes) were checked against BT25 features; potential missed matches were flagged for review.
- <u>Attribute transfer in M:N relationships</u>: The TA tool simply picks any one of the M source features for the transfer; therefore, it might be important to know from which of the M source features the attributes were transferred. Features involved in M:N matches were flagged for review.

b. Post-processing (interactive)

It was necessary to review the evaluation results described above. The flagged issues were loaded into a queue for interactive inspections; review notes were added to a field to keep track the status.

- Review of potential FM issues: From 2788 matches, 69 cases were flagged including 36 false-alarms and 33 needed actions. For example, match errors (green lines) occurred in the complex roundabout area with multiple ramps from single or dual centrelines, as shown in Figure 9; the FM_GRP values (links) and the transferred attributes for the involved features would need to be corrected.

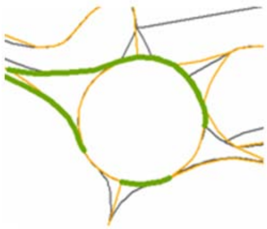

Figure 9: Example of flagged match issues.

- Review of potential missed matches: Overall 267 issues were flagged; 140 of them needed decisions or corrections. For example, one of the dual centrelines was not matched, therefore didn't receive the feature unique identifier value; see Figure 10.

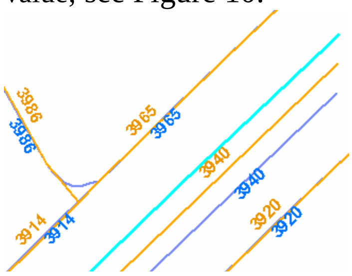

Figure 10: The highlighted line was missed in matching process.

- Review of attribute transfer in M:N relationships: Among the 2961 matched features in BT25_VIA1_prep, 354 were in M:N relationships. If it matters from which source features the attributes should be transferred, then all these records would need to be checked. This part of the review was not done for this paper due to time limit.

For making corrections of the 33 match issues and the 140 missed transfers, total 173 cases, it would require interactive examination and editing to determine the right matches and to correct the links and the transferred attributes. This work was left out from this paper due to time limit.

The overall matching accuracy should take into account of matched accuracy and unmatched accuracy. Unless all cases can be checked and verified, the matching accuracy can only be estimated based on the results of evaluation and review. The matched accuracy was estimated as the percentage of the correct match count over the total match group count; and the unmatched accuracy was estimated as the percentage of the correctly unmatched count over the total unmatched count. The average of the two estimates became the overall accuracy estimate. Table 1 summarizes the estimated accuracy values for the data used in this scenario.

Table 1: Feature matching accuracy estimates

Matched group count	**Correct match count**	**Matched accuracy estimate (MA_est)**	**Unmatched count**	**Correctly unmatched count**	**Unmatched accuracy estimate (UnMA_est)**
2788	2755 (2788 – 33)	98.8%	1845	1705 (1845 – 140)	92.4%
Overall matching accuracy estimate (average of MA_est and UnMA_est) = **95.6%**					

The processing time estimates in each steps of the workflow are shown in Table 2. The processing times for reviewing attribute transfer in M:N relationships and for making corrections are yet to be evaluated as indicated above.

Table 2: Processing time estimates

Conflation and evaluation	**Review of potential FM issues** total 69 cases; 2-3 cases per min.	**Review of potential missed matches** total 267 cases; 2-3 cases per min.	**Total** (Not including the time for reviewing M:N relationships and for making corrections)
1min 3sec	23-35 min.	89-134 min.	113-170 min. (2 – 3 hours)

<u>Scenario B: Establish links for watercourse centrelines at BT-5M and at BT-25M</u>

The watercourse centreline test datasets were BT25_FLU1 (86 features) and BT5_FLU1 (224 features), as shown in Figure 11. The goal was to obtain unique IDs as links for corresponding features.

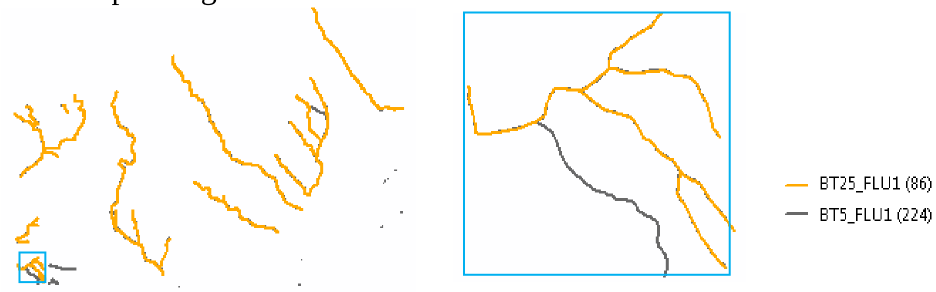

Figure 11: Watercourse centrelines – data area (left) and a close-up (right).

The datasets had clean topology and reasonable link breaks, therefore no pre-processing was needed. The workflow consisted of conflation and evaluation, and post-processing, as presented below.

a. Conflation and evaluation (automatic)

A GP model, shown in Figure 12, was built to establish links between watercourse centrelines of BT-5M and BT-25M. It also evaluated the results and produced information that could facilitate post-processing.

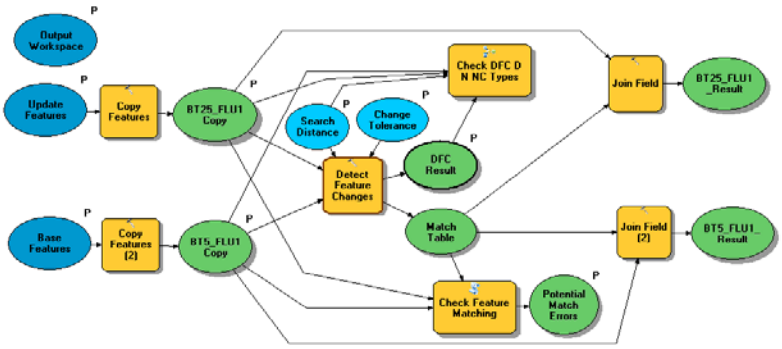

Figure 12: GP model for conflation and evaluation on FLU1s.

This time the DFC tool was used to generate the match table containing the FM_GRP values to serve as the links; the inputs for DFC were BT25_FLU1 as the Update Features and BT5_FLU1 as the Base Features. The advantage of using DFC is that it outputs features with the change types, as listed below, which reflect match situations; therefore, once symbolized, they can help visualize the matching result effectively; see Figure 13.

S (spatial change) – matched features with spatial difference.
NC (no change) – 1:1 match with no spatial difference.
N (new) – extra update feature with no match to base feature.
D (delete) – extra base feature with no match to update feature

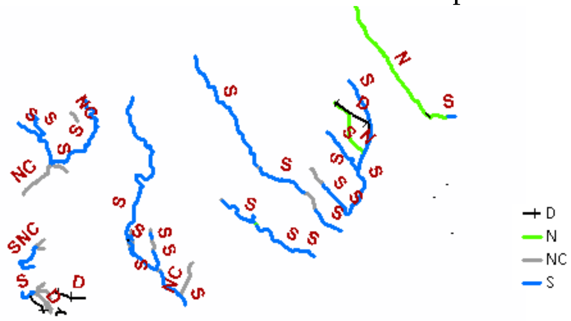

Figure 13: Change types of DFC result.

Similar to Scenario A, the matched features received their common FM_GRP values as links, as displayed in Figure 14.

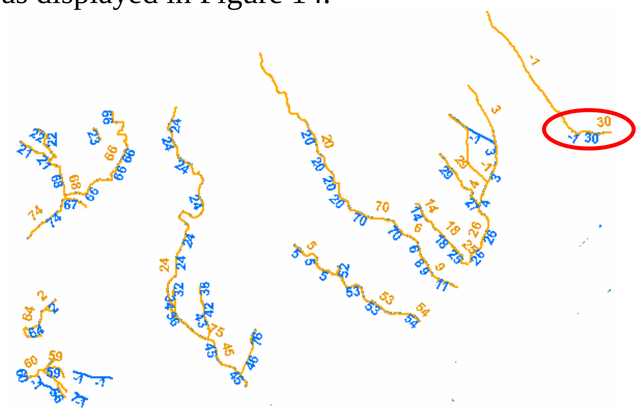

Figure 14: Resulting FLU1s of BT25 (orange) and BT5 (blue), labelled by FM_GRP values (the established links).

The evaluations were done similarly to Scenario A and focused on potential FM issues and potential missed matches, which were flagged for review.

b. Post-processing (interactive)

From 76 matches, 5 cases were flagged for potential match issues, but they were all false-alarms mostly caused by the uneven ending of lines.

From 29 unmatched cases, 5 were flagged for potentially missed matches, 3 of them needed decisions or corrections. These cases mostly occurred where the BT5 features contained many short lines, as shown in Figure 15 – an enlargement of the area inside the red oval in Figure 14. Some features were not matched; DFC produced unexpected N and D features. These features would need to be selected and assigned a desired FM_GRP value as their link.

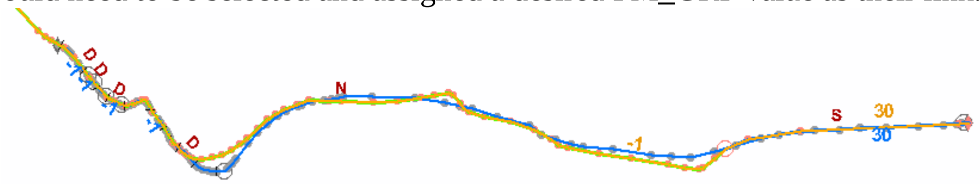

Figure 15: Missed matches occurred where multiple short lines were found in BT5 data.

The estimated match accuracy for matched groups was 100%; and estimated accuracy for unmatched cases was 89.2%. So, the estimated overall matching accuracy for the data of this scenario was 94.6%.

The total processing time was estimated less than 0.5 hour, including automatic processing 6 sec. and the remaining time for interactive correction of the 3 missed matches.

<u>Scenario C: Establish links between buildings at BT-5M and at BT-25M</u>

Unlike linear features, the buildings at BT-5M and at BT-25M, as shown in Figure 16, differ in the following ways:

EDI2_5M – very detailed polygon buildings at BT-5M (3889 features)
EDI2_25M – Generalized polygon buildings at BT-25M (2845 features)
EDI0_25M – Generalized point buildings at BT-25M (636 features)

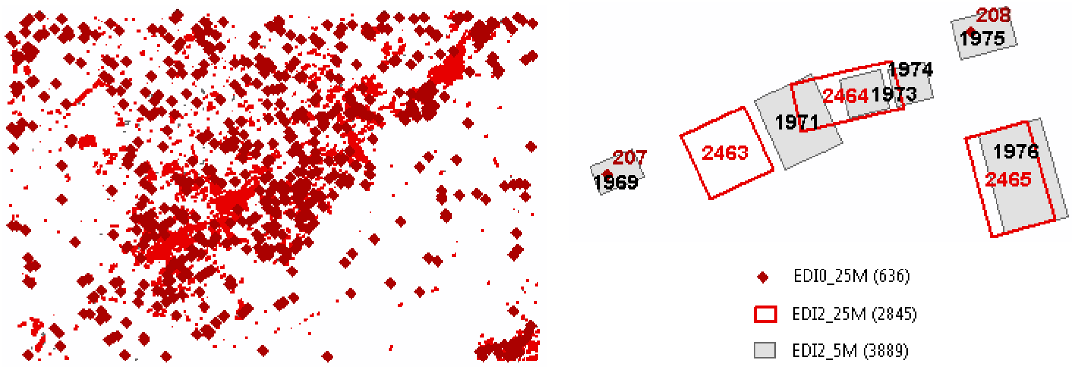

Figure 16: Buildings of BT-5M and BT-25M - test area and a close-up.

The goal was still to find corresponding buildings and assign unique IDs to them as links. The matching was done through spatial analysis, which relied on the knowledge of how the buildings have been generalized from BT-5M to BT-25M. For example in the close-up area of Figure 16, the polygon buildings 2463 and 2464 of EDI2_25M should correspond with 1971, 1973, and 1974 of EDI2_5M; it is a 2:3 correspondence.

The workflow consisted of establishing links and post-processing, as presented below.

a. Establishing links (automatic)

A GP model shown Figure 17 was built to find corresponding buildings and assign unique IDs for each group of matched features to serve as links between buildings of BT-5M and BT-25M. Figure 18 shows examples of the established links (unique IDs) on a variety of building groups.

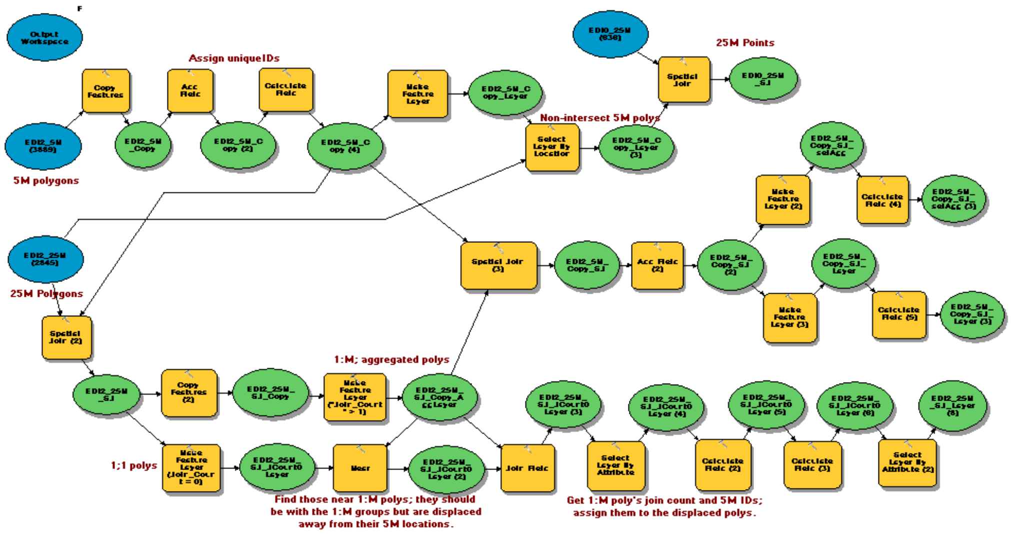

Figure 17: GP model for establishing links on EDIs.

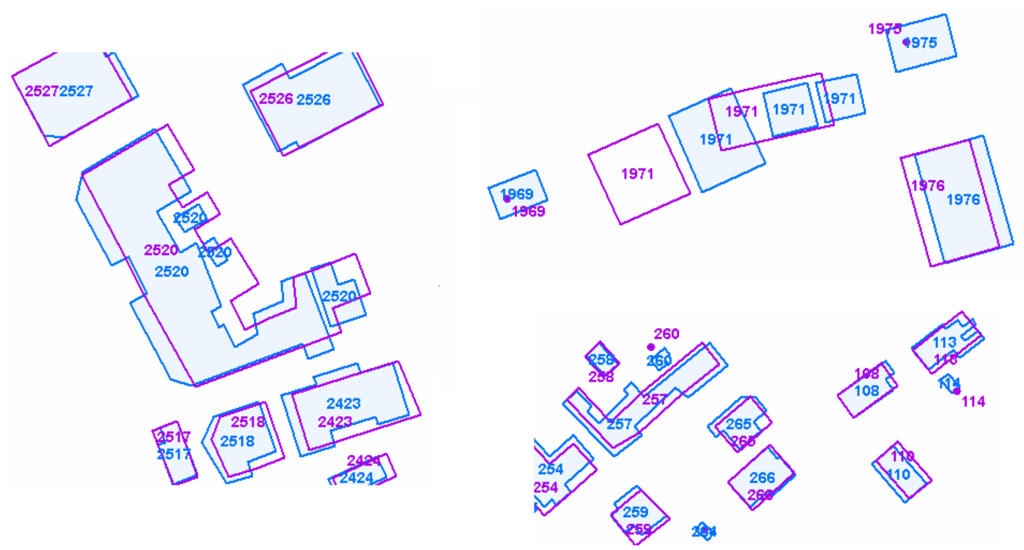

Figure 18: Unique IDs (links) on buildings at BT-25M (purple) and at BT-5M (blue).

b. Post-processing (interactive)

From the 636 point buildings at BT25M, 612 were correctly linked with the polygon buildings at BT5M. The other 24 were reviewed and confirmed to be correctly not linked.

From the 2845 polygon buildings at BT25M, 2780 were linked with the polygon buildings at BT5M (total 2753 groups). It was noticed that buildings with shared walls had a slight chance of being grouped incorrectly. The 27 shared walls were identified and reviewed; only 2 group ID issues were found among buildings with shared walls. The remaining 65 buildings were reviewed and confirmed to be correctly not linked.

The overall linking accuracy for the data of this scenario was estimated 96.7%.

The total processing time was estimated less than 0.5 hour, including automatic processing 6.9 sec. and the remaining time for interactive review and corrections.

5. Conclusions

The implementation of the ICGC MRDB is a long and complex process that introduces significant changes in the production workflows. To minimize the impact on the production environment, ICGC is doing the migration in progressive steps: design and implementation of the data model, migration of the production system and finally implementation of the MRDB links.

The conflation tools, along with other geoprocessing analysis tools, developed by Esri, have a great potential to facilitate the first link establishment for the MRDB and to harmonize attributes between multi-scale features. For the linear features, the Transfer Attributes tool and the Detect Feature Changes tool give the option to generate a match table, which establish the links between corresponding features and allow the attribute transfer. For polygons and points, the power of spatial analysis, especially the overlay tools, makes it easy to identify spatially related features.

The results of this initial test suggest that the first link establishment for the ICGC MRDB can be highly automated and the time for post-processing is expected to be reasonable and worth spending. More tests, using data covering different geographic areas and other feature types, should be performed to have more reliable information about the process.

References

Baella, B., Lleopart, A., Pla, M. (2012). ICC Topographic Databases: Design of a MRDB for data management optimization, *15th ICA Generalization Workshop, Istanbul, 2012.*

Lee, D., Yang, W., Ahmed, N. (2014). Conflation in Geoprocessing Framework - Case Studies, *GEOProcessing, 2014, Barcelona, Spain*.

Yang, W., Lee, D., Ahmed, N. (2014). Pattern Based Feature Matching for Geospatial Data Conflation, *GEOProcessing, 2014, Barcelona, Spain*.