A Practical Experience on Road Network Generalisation for Production Device

Jérémy Renard*, Silvio Rousic**

* COGIT laboratory, IGN-F, Université Paris Est
** CETE Méditerranée, DCEDI/AGIL, MEDDE

Abstract. This paper is based on collaboration between two institutes with the goal to test research processes on road network generalisation in a real production context. The first tested algorithms were only based on the data geometry as it is usually described in literature, with unsatisfying results. Then major improvements were brought by taking into consideration semantic information, with much better final results. The conclusion asks the question of the duality between genericity and data schema dependence.

Keywords: cartographic generalisation, road network, data schema, production workflow

1. Introduction

Cartographic generalisation aims at adapting the content of geographic data depending on their scale, to ensure the legibility of a map while preserving its main characteristics. The main goals of this cartographic process are to reduce the number of geographic objects in the map to avoid overcrowding and to schematise these objects in order to improve their legibility. The automation of generalisation is a wide research topic that has been studied for years, resulting in many implemented methods, single algorithms, or entire generalisation processes.

In case of a whole road network, reducing the scale of a map implies many generalisation operators to make the network sensible and readable at the final scale. One could cite detection of particular road structures like roundabouts or interchanges, collapsing of these structures to simplify their layout, road selection to keep only the main road axes and eliminate non significant ones, schematization and displacement to avoid road overlapping especially in case of close roads or mountain bends, etc. Most of these operators on road network have already been described in literature and implemented in generalisation software and platforms, and they seem to be efficient in a research context. Nevertheless, it could be interesting to test them in a context of production device, with different constraints like particular data schemas or intensive processing on a large amount of data.

The goal of this paper is to illustrate the collaboration between two French institutes, CETE Méditerranée and COGIT laboratory (IGN-F), to adapt research work on road network generalisation in a real production context. Section 2 details the background and objectives of this collaborative work. Section 3 focuses on the methods used to generalise road network data, then Section 4 gives an overview of first results and conclusions. Section 5 tries to go further to show how the results can be improved by taking into account the data schema. Finally, Section 6 highlights all conclusions that can be deduced from this work, asking higher level questions on how generalisation algorithms should be described to improve their efficiency without losing genericity.

2. Background and objectives

2.1. A context of collaboration between two institutes

CETE Méditerranée is one of the technical study centers of the French ministry for equipment and sustainable development (MEDDE). It provides engineering services in different topics within the scope of the ministry. Among these various topics, CETE Méditerranée is in charge of questions related to transport planning, and it maintains a road network database entitled RGC – for Routes à Grande Circulation (High Circulation Roads) – which lists all the roads that are the most used by network users.

The RGC database used to be maintained at a low level of detail, corresponding to maps at national scales (250K or 500K for instance). But regarding the new standards of French geographic data, it needs to be derived to a higher level of detail by using RGE – for Référentiel à Grande Echelle (High Scale Reference) – corresponding to local maps (50K or 100K for instance). Thus, this update of the database raises two issues:

- Data matching to transfer the attributes of the old RGC data to the RGE
- Cartographic generalisation to still be able to provide national or regional maps at lower scales from the RGE

To solve these two issues, CETE Méditerranée turned to IGN-F to take benefit of its expertise in the field. Matching two similar databases through data integration has been tackled by Mustière & Devogele (2008) or Olteanu-Raimond (2008), resulting in efficient processes used in production lines at IGN-F, so this first issue has been solved by working in collaboration with IGN-F development department. The second issue on cartographic generalisation is the background for a collaboration between COGIT laboratory (IGN-F) and CETE Méditerranée, which is the topic of this paper.

2.2. Objectives for both parts

Generalising a whole road network involves different methods and algorithms that have already been described in literature – Section 3 gives an overview of them. Most of them have been implemented in CartAGen, the generalisation platform of COGIT (Renard et al. 2010, Renard et al. 2011) and successfully tested on data samples. So CETE Méditerranée and COGIT decided to work together towards the final goal of using COGIT algorithms to generalise CETE data.

Through this collaboration, there are objectives for both parts. The goal for CETE Méditerranée is to test and to master tools to be able to perform a complete well-suited generalisation of RGC road network. To reach this aim, the open-source part of CartAGen platform is used as a library where a user can plug its own data, then road network algorithms can be processed on these data. The final goal is to adapt the algorithms to finally have a turnkey solution for automated RGC network generalisation.

For COGIT laboratory, the contributions lies in the possibility to test algorithms on real production data. So the principles of the algorithms can be discussed, fitted to the data, and improved to be really applied in a context of production device. In this way, some validation could be provided to generalisation methods and algorithms.

3. Methods for RGC road network generalisation

3.1. Data description

RGC road network database is divided according to the layouts of French departments. Figure 1 shows an example of all RGC roads in a whole department (Bouches-du-Rhône), the network is not so dense as only high circulation roads are registered, but some particular areas still have a high density of roads. Anyway, as we talk about road axes that are intensively used for transport, there is obviously no small sinuous roads (like mountain roads) so it won’t be necessary to use generalisation methods for such roads like GALBE or AGENT (Mustière & Duchêne 2001). Apart from that, some characteristics of the RGC roads are very significant: for instance there is no little dead end that should be contextually removed, or there is no real difference between urban areas and rural areas. These characteristics allow avoiding generalisation methods which could have been used to generalise the network.

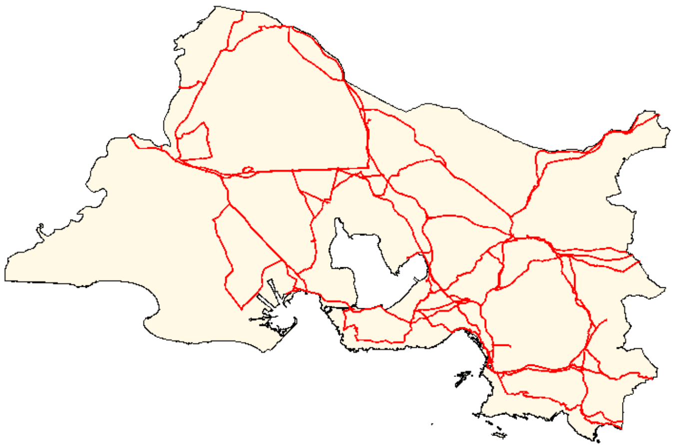

Figure 1. An overview in QGIS of all RGC roads in a whole French department (Source: CETE Méditerranée, ©RGC)

RGC roads have a high quantity of semantic information. More than forty attributes are attached to the roads: nature, number, classification ... In a first stage, all these attributes are not used for generalisation, taking into account only the geometry of roads as it is a common habit in literature. Indeed, by avoiding considerations on semantics, generalisation algorithms can be applied on any type of data without regarding their schema. That ensures some kind of genericity of the method.

The objective is to build a complete process to generalise the whole road network. Such a process has already been proposed by Touya (2010), but it needs to be simplified to fit to the characteristics of the current data. In our case, the process is based on several steps: first detection and collapsing of particular structures like roundabouts or dual carriageways, then road selection, and finally more complex issues like road displacement or interchanges collapsing.

3.2. Detection and collapsing of roundabouts

Roundabouts are particular structures of the network that need to be collapsed into a single crossroad when generalising at lower scales. Sheeren et al. (2004) proposed to detect roundabouts by using the faces of the topological graph induced by the network. A roundabout is a face whose shape is almost circular, which can be traduced by using Miller's index of compactness ($C=4\pi.area/perimeter^2$). According to Sheeren's experience, a threshold of 0.97 on this compactness can discriminate whether a face is a roundabout or not. Then roads connected to the roundabout face are extracted, and another step of detection is applied to know if adjacent faces of the roundabout are branchings: a branching is a small triangular face.

Finally, the roundabout is described with a central face (in red in figure 2), possibly branchings (in green in figure 2), internal roads delineating all these faces, and external roads connected to these faces.

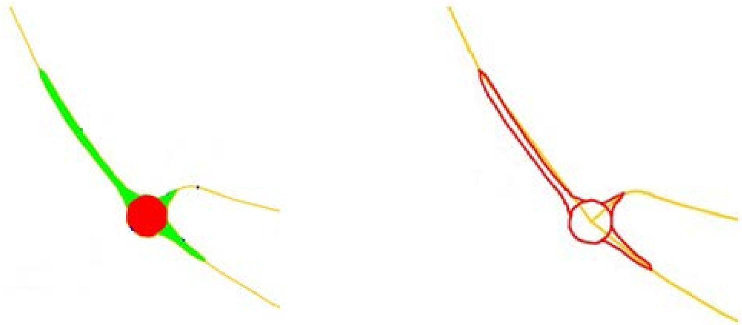

Figure 2. Detection (left) and collapsing (right) of a roundabout. The original roundabout is in red, the generalised roads are in orange.

Collapsing a detected roundabout is then far simple and can be detailed in three steps:

- Calculating the center of the roundabout
- Removing internal roads (and so internal faces)
- Reconnecting external roads to the center of the roundabout

3.3. Detection and collapsing of dual carriageways

Dual carriageways are represented in the data with both of their carriageways, resulting in two parallel roads. These dual carriageways need to be collapsed in order to keep only one road which should ideally be in-between the two original carriageways. Thom (2005) gave some propositions to detect and collapse dual carriageways, but we use a different approach here.

As it is the case for roundabouts, the detection is based on the geometric properties of the faces of the topological graph induced by the network. The idea is to detect long narrow faces by means of elongation, convexity and compactness, because these faces represent separators of dual carriageways. It is closed to the method of Touya et al. (2010) to detect hedges in a vegetation layer. First the convexity of a face is calculated, by comparing its area with the area of its convex hull. Our experience highlights a threshold of 0.8 to determine if a road network face is convex or not. Then there are two possibilities:

- The face is convex enough (convexity>0.8), so elongation can be used as a significant descriptor of its shape. The elongation is calculated as the ratio length/width of the minimum bounding rectangle of the face. If it is higher than 5.0, the face is considered as a dual carriageway separator.
- The face is not convex enough (convexity<0.8), so elongation cannot be used. Instead of elongation, compactness is calculated through Miller's index. If compactness is lower than 0.2, the face is considered as a dual carriageways separator.

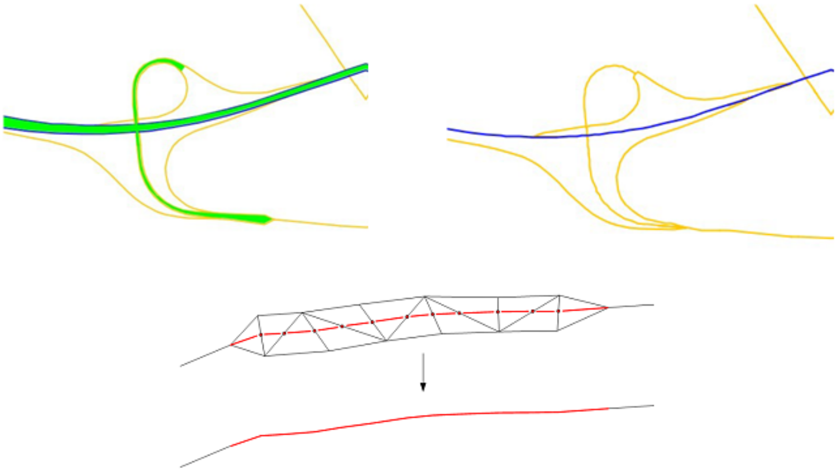

Figure 3. Detection (top left) and collapsing (top right) of a dual carriageway, by creating a central skeleton of the separator through triangulation (bottom)

Then, dual carriageways are detected as the roads delineating separators. To collapse them, the idea is to use the method of Regnauld & Mackaness (2006) to collapse polygonal rivers represented with their banks. A Delaunay triangulation of the face is computed, then the triangulation is used to build an internal skeleton of the face. This skeleton is filtered and smoothed, and it results in a central line representing the new generalised road, taking place in-between both dual carriageways. Finally, original carriageways are removed, and reconnection of the network to the new central road is ensured, especially for slip roads in interchanges.

3.4. Road selection

After detection and collapsing of road structures, another stage of generalisation is to eliminate some roads of the network to avoid overcrowding and keep only the most important roads. Two algorithms are used to perform road selection. The first one consists in building road strokes by means of good continuation of major road axes (Thomson & Richardson 1999). These strokes are considered as the most important roads and are always kept unchanged in the data.

The second algorithm that is used for road selection is computation of shortest paths with Dikjstra (1959) algorithm by means off a selection of attraction points in the network, following the method of Richardson and Thomson (1996). In our study case, there is no contextual information attached to the data (e.g. where are city centres, activity areas ...), so attraction points are randomly affected to create some kind of regular grid over the network. The idea is then to compute all shortest paths between all attraction points, to attribute higher values to roads that are the most travelled by shortest paths, and to eliminate the minor roads regarding their values.

Anyway, it is important to notice that the RGC network has already been widely selected to keep only high circulation roads, so it is not necessary to eliminate roads again except in high density areas.

3.5. Going further: road displacement, interchanges …

In addition to previous treatments, some additional algorithms could be applied on RGC network to fully generalise it.

For instance, in case of general overcrowding due to a high density of roads in a small area, road displacement could be necessary to avoid overlapping between close roads. The best solution to perform this is to use the elastic beams of Bader (2001). Unfortunately, this process is not integrated in the platform CartAGen.

To go further, one of the important remaining issue on road generalisation concerns detection and collapsing of major interchanges. There is some interesting research work on the topic – Mackaness & Mackechnie (1999) or Dogru et al. (2009) could be cited as good examples – but they are not mature enough to be integrated as a fully automated solution in CartAGen.

4. Results and first conclusions

The first experiments have been carried out with the existing algorithms described in Section 4, which only take into account the geometry and the topology of the road network. The results were not good enough to be integrated in a production workflow. The lack of quality comes mainly from the structures detection process which is not comprehensive. Detection of roundabouts and dual carriageways is based on the faces of the topological graph induced by the whole network, and all roads are considered equally during the stage of detection. As a consequence, the existence of parasite roads which do not take part of a structure can disrupt the detection of this structure. Figure 4 shows examples of undetected roundabouts with two particular cases. Figure 5 shows problems that occur for dual carriageway detection when roads are crossing through interchanges.

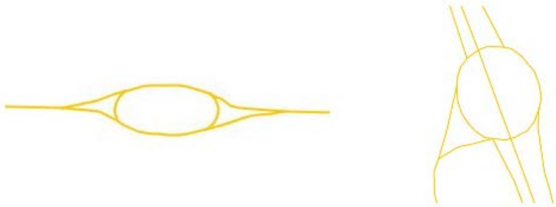

Figure 4. Undetected roundabouts because of an imperfect circular shape (case 1) or a parasite road running over (case 2)

In terms of evaluation, we can observe that around 80% of the roundabouts are well detected. The results are even worse regarding the dual carriageways, only 60% of them are well detected and many problems of reconnection and continuity appear, almost each time that an interchange is met.

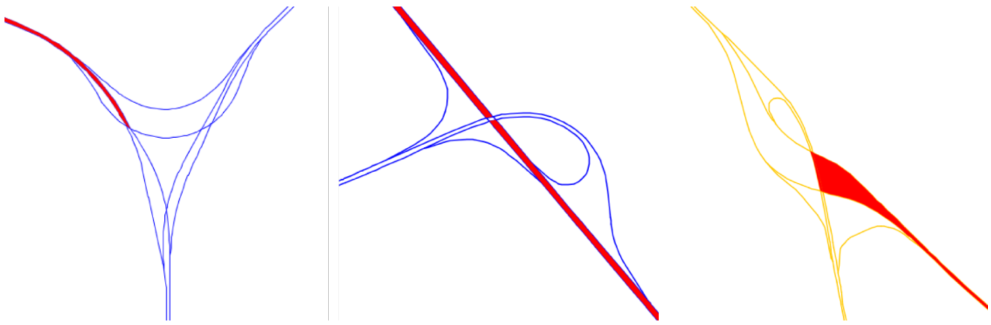

Figure 5. Undetected dual carriageway separators because of interconnections

Apart from detection problems, structures collapsing gives satisfying results as far as they are well detected. Only few improvements could be performed in properly reconnecting the extremities of generalised roads, and overall in generalising interchanges as global structures. But obviously, when structures are badly detected, they cannot be correctly collapsed. Road selection through strokes and shortest path computation seems to gives very good results, but the tests have not been carried out as far as possible.

These results lead us to a first conclusion: detection algorithms as they are described in literature are not robust enough to bear a production context with real data. The problem comes from their philosophy which consists in only taking into account geometrical considerations to perform generalisation. However, this idea is very sensible to ensure the genericity of the algorithms because they can be applied on any type of data regardless of the schema, but in a real production device the first results highlight the necessity to consider data schema and semantic information as means to improve the algorithms.

5. Improvements by adapting methods to the data schema

5.1. How to take benefit from data schema?

Most of the problems of road structures detection come from the creation of a topological graph induced by all roads of the network, with the consequence of having locally parasite roads over structures. The solution to face this question is to construct a topological graph based only on roads that are concerned in the structures we want to detect, and not all roads. For instance, considering the case 2 of undetected roundabouts in Figure 4, such a parasite road should be excluded of the topological graph. To reach this goal, two attributes are checked: the attribute NATURE must be different from "Motorway" or "Dual carriageway" and the attribute DIRECTION must be different from "Double". Indeed, roundabouts are composed of minor roads with a single driving direction. With such a query, all potential parasite roads are not considered while building the topological graph, and it allows detecting roundabouts of case 2. To deal with case 1, the solution consists in lower the threshold of Miller's index to 0.80 (instead of 0.97), so roundabouts that are not perfectly circular are detected without introducing any abusive detection.

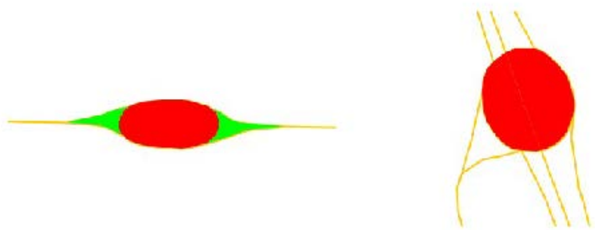

Figure 6. Improved detection of initially undetected roundabouts of Figure 4

Concerning the detection of dual carriageways, the main problems are due to the interchanges between several road axes, as shown in Figure 5. The ideal solution would be to treat each dual carriageway one by one to avoid reconnections. That is possible by using the attribute NUMBER: all roads of the same number are part of the same transport axis and are computed together to create a topological graph without considering other roads, then this topological graph is used to detect dual carriageways, and this operation is repeated for each different road number. In fact, road axes are treated one by one for detection and collapsing. This method solves almost all initial problems on dual carriageways.

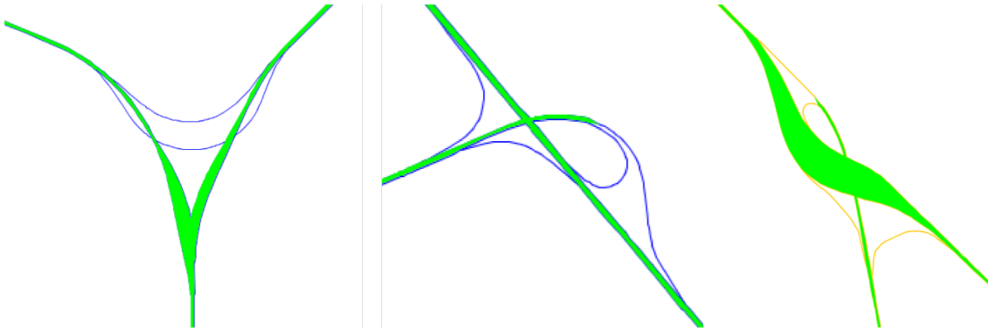

Figure 7. Improved detection of initially badly detected dual carriageways of Figure 5

The way semantics are used to improve initial algorithms is not so important in itself; what is essential is to admit the fact that considering data attributes as a mean to perform generalisation is a very significant idea to improve generalisation results.

5.2. Improved results

The improvements of existing algorithms by taking benefit of the data schema are very significant in terms of results. 95% of the roundabouts are now well detected and correctly collapsed. Almost 100% of the dual carriageways are correctly detected, and most of them are then correctly collapsed. If we lose genericity while introducing some dependence to the data schema, we gain far better results in the generalisation process. Indeed, we achieve to adapt the very theoretical and generic process of Touya (2010) to the particular case of a real production device with significant characteristics in the data, resulting in a global process that is partly described by Figure 8.

The remaining errors are mainly due to a lack of precision in the data: wrong attributes (a motorway whose nature is designed as "Slip road"), missing attributes (no road number), even missing roads that create holes in the network. Without these errors that lie in the data producer upstream, improved detection algorithms would give almost 100% of good results in all cases. The main effort that should be done concerns the collapsing of dual carriageways that still offers some imperfections in complex cases.

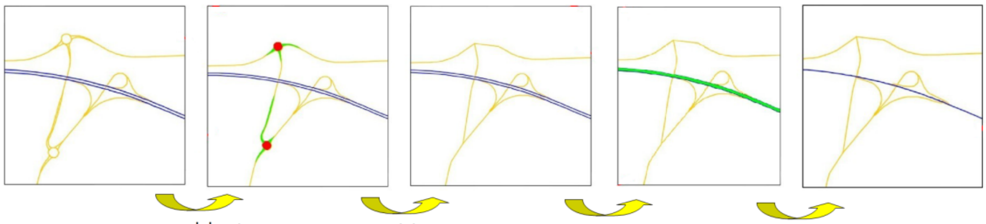

Figure 8. Complete process of detection and collapsing of road structures. Road selection and displacement should then be applied afterwards*

It is also important to notice that these improvements make the algorithms much faster, as topological graphs are created based on several roads rather than the whole network.

6. Conclusions

In this work, we tried to apply existing algorithms in a real production device through a whole road network generalisation. First results prove that the algorithms as they are described in literature are not powerful enough to be considered as fully operational. Then improvements were applied on these algorithms by taking into consideration the data schema. Improved processes and operators that have been tested seem to be strong enough to support a production device, but still need to be improved to be able to perform a perfect result on 100% of the road network, especially for dual carriageways collapsing which is not totally satisfying.

Anyway, the most important conclusion of this work is probably that road network generalisation algorithms based on data geometry need to be adapted to the data structure, including semantics, to be really powerful and relevant. Similar conclusions could probably be drawn while testing other algorithms on different cartographic themes, as it is the case for Revell et al. (2006) or West-Nielsen & Meyer (2007) – other papers should probably be cited as examples. This observation should lead us to question how to conduct generalisation research.

Although generalisation algorithms and processes are usually designed to be as generic as possible by only considering geometric properties of data, the practical experience described in this paper underlines the necessary dependence to the data schema to perform well-suited generalisation solutions in a production workflow. The question underlying this conclusion is the following: what is the final goal of generalisation processes? To help research looking ahead to find new generic solutions or to propose fully automated solutions for practical cases especially in a production context? Our approach to how we conduct automated generalisation research depends on our answer to this question.

References

Bader M (2001), Cartographic Displacement in Generalization : Introducing Elastic Beams. 4th ICA Workshop on Generalisation and Multiple Representation, Beijing

Dogru O, Duchêne C, Van de Weghe N, Mustière S, Ulugtekin N (2009) Generalization Approaches for Car Navigation Systems. 24th International Cartographic Conference, Santiago

Mackaness W A, Mackechnie G (1999) Detection and Simplification of Road Junctions in Automated Map Generalization. GeoInformatica, 3: 185-200

Mustière S, Duchêne C (2001) Comparison of Different Approaches to Combine Road Generalization Algorithms: GALBE, AGENT and CartoLearn. 4th ICA Workshop on Generalisation and Multiple Representation, Beijing

Mustière S, Devogele T (2008) Matching Networks with different Levels of Detail. GeoInformatica 12(4): 435-453

Olteanu-Raimond A-M (2008) A Multi Criteria Fusion Approach for Geographical Data. Quality Aspects in Spatial Data Mining, Taylor and Francis: 45-56

Regnauld N, Mackaness W (2006) Creating a Hydrographic Network from its Cartographic Representation : A Case Study Using Ordnance Survey MasterMap Data. International Journal of Geographical Information Science 20(6): 611-631

Renard J, Gaffuri J, Duchêne C (2010) Capitalisation Problem in Research - Example of a New Platform for Generalisation: CartAGen. 12th ICA Workshop on Generalisation and Multiple Representation, Zürich

Renard J, Gaffuri J, Duchêne C, Touya G (2011) Automated Generalisation Results using the Agent-Based Platform CartAGen. 25th International Cartographic Conference (ICC'11), Paris

Revell P, Regnauld N, Thom S (2007) Generalising and Symbolising Ordnance Survey Base Scale Data to Create a Prototype 1:50 000 Scale Vector Map. The Cartographic Journal 44 (3): 251-257

Richardson D E, Thomson R C (1996) Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks. Cartographica 33: 75–83

Sheeren D, Mustiere S, Zucker J-D (2004) How to Integrate Heterogeneous Spatial Databases in a Constistent Way? Proceedings of Conference on Advanced Databases and Information Systems (ADBIS). Budapest, Springer: 364–78

Thom S (2005) A Strategy for Collapsing OS Integrated Transport Network Dual Carriageways. 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña

Thomson R C, Richardson D E (1999) The "Good Continuation" Principle of Perceptual Organization applied to the Generalization of Road Networks. Proceedings of 19th International Cartographic Conference, Ottawa: 1215-1223

Touya (2010) A Road Network Selection Process Based on Data Enrichment and Structure Dectection. Transactions in GIS 14(5): 595-614

Touya G, Duchêne C, Mustière S (2010) Généralisation et integration pour un fond vert commun entre l'IGN et l'IFN. Revue Internationale de Géomatique 20(1): 65-86

West-Nielsen P, Meyer M (2007) Automated Generalisation in a Map Production Environment – the KMS Experience. In Generalisation of Geographic Information: Cartographic Modelling and Applications, Elsevier: 301-313