Making a map from "thematically multi-sourced data": the potential of making inter-layers spatial relations explicit

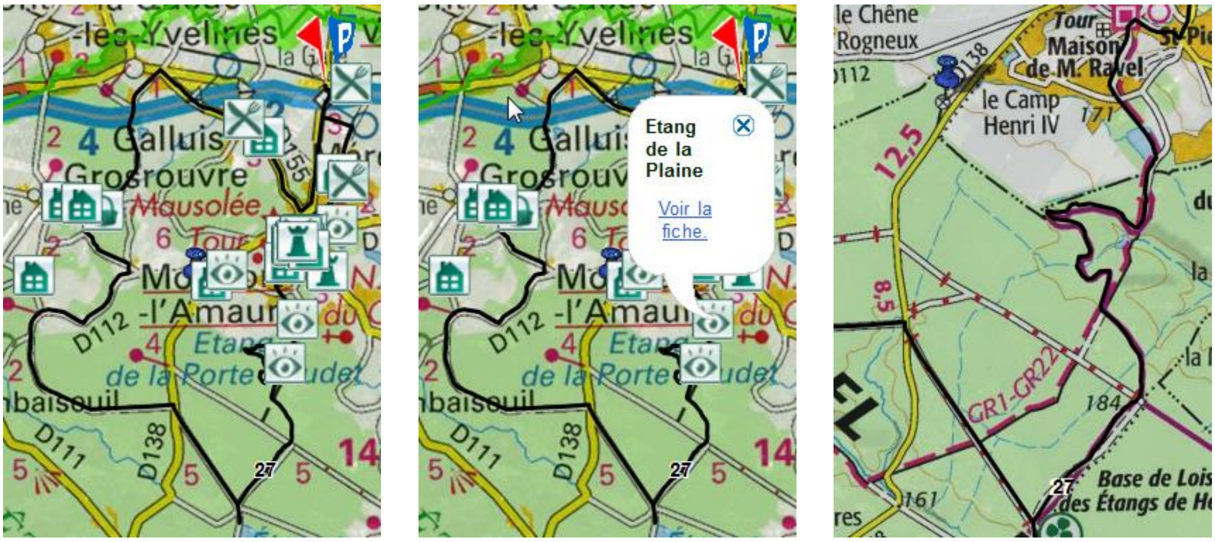
C. Duchêne

IGN Laboratoire COGIT, Saint-Mandé, France
Université Paris-Est, Champs-sur-Marne, France

1 Introduction

One kind of map commonly produced from the always increasing volume of available data is made of "base" data acting as a backdrop map used for localisation purpose, on which data related to a particular theme, of interest for the user, are overlaid (Jaara et al. 2011; Moseme and van Elzakker 2012). In the rest of the paper, we call "data layer" a dataset stemming from one source (authoritative or VGI), here considered internally consistent – i.e. reliability and redundancy issues underlined by Sester et al. (2014) in the case of VGI data have been handled before. We call maps composed of several data layers (each stemming from one source), *thematically multi-sourced maps*. Here we consider maps composed of two data layers: one base layer and one thematic layer. Most thematic maps have been based on this two layers principle for decades (and were multi-sourced), but they have long been made mainly by cartographic experts (Das et al. 2012). Recently, making and publishing such maps has become much easier. Their potential is huge for the users, however, the risk is huge to produce maps that are hardly legible or interpretable and can lead to bad decision making (Jaara et al. 2011; Das et al 2012; Gaffuri 2011; Balley et al. 2014; Sester et al. 2014). This is often due to a bad management of relations between the thematic and base layer. For instance, Figure 1 shows an interactive, multi-scale web map presenting a pedestrian route suggested by the authority in charge of a natural park in the neighbourhood of Paris. At default scale when first entering the web site (a), too many pictograms showing points of interest along the route are cluttered and hide a part of the route. Some of them actually correspond to topographical objects present in the base layer that are visible from the route, e.g. the pound highlighted in (b), which might therefore have been emphasised in the base map instead of being hidden by an eye-shaped pictogram. Picture (c) shows an extract of the map at the maximum zoom level that still enables to see the whole route. Because the geometry of the route is not distorted to take into account the symbolisation of the underlying transport network, it is hard to figure out e.g. what ratio of the route is on a footpath vs on a bigger road, or what portions follow a pedestrian route equipped with signs (magenta line in the base map) – two pieces of information that could be expected at this scale when choosing for which walk to go.

We think that explicitly reasoning on spatial relations between foreground and background layers is a necessity to produce "good" thematically multi-sourced maps. This paper is a position paper inspired, between others, by this challenging sentence from Ormeling (2011): "*Despite all advances in digital generalisation, no overall generalisation theory has been worked out, nor are there convincing solutions for digital generalisation of all relationships between cartographic objects*". It aims at three things: (1) examine the role of relations in the particular case of thematically multi-sourced maps, and what theoretical knowledge we have about their evolution through map scales (Section 2); analyse what kinds of spatial relations should be considered, when and why, and how they have been used and modelled in previous research (Section 3); propose a first list of open research issues related to the use of explicit spatial relations to improve thematically multi-sourced map making (section 4).



(a) (b) (c)

Figure 1. Bad legibility in a map showing a suggested pedestrian route[1]

2 Relations and their generalisation in thematically multi-sourced maps

2.1 Role relations in thematically multi-sourced maps

Maps in general are used to support decision making or analysis. Generalisation aims at two things: abstracting the content of the map so that it fits to the target scale, and graphically represent this content in a legible way (Mustière et al. 1999). It has regularly been claimed that on a map, relations between features (horizontal relations in the framework of Bobzien et al. 2008) are as important as the features themselves or even more important, and must be carefully taken into account during generalisation (e.g. Papadias and Theodoridis 1997; Ruas and Mackaness 1997; Mackaness and Edwardes 2002; Touya et al. 2012).

Now, topographic maps are generalist and supposed to represent the ground with a neutral point of view, therefore on such maps, spatial relations between objects are preserved at best, and caricatured (e.g. through displacement) depending essentially on legibility threholds. In contrast, thematic maps are focused on a particular theme, which should thus obviously be emphasised, i.e. made particularly legible with respect to the backdrop layer. It is therefore needed to generalise the backdrop data more than for topographic data alone at the same scale – Das et al. (2012), who studied conflicts in web maps built from VGI data, claim that the density is generally too high for the intended level of detail (LOD) in such maps, and that it even increases when a foreground thematic layer is added. Moreover, not only the thematic

1 http://www.parc-naturel-chevreuse.fr/index.php?id=593&plan=27

layer itself is of interest for the user, but also the relations perceived with the backdrop data. For instance, a particular car accident happened on a particular road, close to a particular junction (Jaara et al. 2013; Gould and Cheng 2013). Such relations are spatial relations (the point representing the accident is topologically included in the line representing the road or very close to it, and close to the junction), and they are associated with a semantic by the user depending on its application (the accident happened on that road, possibly because of the proximity of the junction – or an accident has just happened on that road, which is therefore likely to be congested in the neighbourhood of that junction). In other words, the backdrop data provide a spatial context to the thematic data in order to make sense of them (Sester et al., 2014). Indeed, the Open Geospatial Consortium defines “base maps, data or layers” as “Spatial data sets that provide the background upon which more specific thematic data is overlaid and analyzed” (OGC, 2014).

2.2 What knowledge do we have on the evolution of relations through map scales ?

If *no overall theory of generalisation has been worked out* (Ormeling, 2011), some theoretical statements have been made about the processing of spatial relations during generalisation. Bobzien and Morgenstern (2003) propose to express model generalisation as a mathematical function, and identified a number of spatial relationships that should be invariants of this fonctions, i.e. that should be preserved during model generalisation: topological relationships of adjacency and inclusion, reachability (of a place by a network) and relative position (e.g. of point objects with respect to a linear object that is simplified). Duchene et al. (2012) formalise the rule, well known to traditional cartographers, that automated cartographic generalisation should detect and caricature spatial relations that are “almost present” but not completely and therefore introduce visual noise in the map – e.g., when a building is almost parallel to a road, after generalisation it should either be completely parallel to it, or clearly not parallel to it. Regarding thematic maps, Jaara et al. (2013) state that when their scale decreases, not only the content of the background and foreground layers, but also the background-foreground relations should often be, (1) abstracted to fit the level of detail that can be expected from the target map (if the difference of LOD is high enough), and (2) caricatured to improve their legibility and intelligibility (Jaara et al. 2013) – in both cases, in a way that takes into account the aim of the map. They take the example of a car accident happening close to a junction, in an accidentology map: at smaller scale, the accident would be represented at the junction to clearly indicate that it is considered a “junction accident”.

3 Explicitly managing spatial relations between base and foreground data: what relations, when, why and how?

Making a thematic map (single or multi-scale) from a thematic and a base layer, is a form of on-demand mapping (Balley et al., 2012) and consists in several steps among which, in short, at least the following two ones are influenced by the spatial relations between foreground and background objects: (1) integrate the thematic layer and one base layer (if several available) that best fit to it; (2) from this initial (base + foreground layers) combination, generate one or more combination(s) that fit the expected level(s) of detail and a given goal, either by generalisation or by “migrating” (as defined by Jaara et al. 2013) the thematic layer on base layers of lower LOD.

3.1 What kinds of relations?

The spatial relations considered at different stages of the process can be described from different perspectives :

- Existing (this route is locally very close to the road) VS expected (wished) relations (this route should be locally equal to the road)
- In terms of genericity, relations holding at the level of an instance (this accident is close to that junction) VS at a level of a population of instances (a feature type) (all accidents are on roads)
- In terms of symmetry, « hosting » VS « peer-to-peer » relation. Jaara et al. (2013) defined a thematic object “hosted by” a background object as a thematic object that is topologically included in the base object, both also having a strong semantic relation – e.g. accidents are hosted by roads, they exist on roads. This definition would deserve to be refined, since the topologic inclusion might not be a necessity. Cycle routes “aligned with roads”, considered in the use case of Balley et al. (2012), can be considered a hosting relation – the road network hosts the cycle routes. We suggest the term “peer-to-peer” relation for cases where the relation is more symmetric, like when a route passes north to a particular point.

3.2 When, why and how to use foreground-background relations?

Regarding the integration stage, (Sester et al., 2014) state that the integration of thematic and background layers should ensure that relations between both layers make sense, i.e. that (1) thematic and backdrop data have been integrated, and (2) the semantic of the relations is known. This first supposes that the semantic of the thematic layer itself is known. It can be described with the standards of the semantic web, by annotating the thematic layer (feature type) with an ontology, as in the prototype of expert geoportal set up by Toomanian et al. (2013). In this study, the annotation of the thematic layer was done manually, but Klien (2008) proposes a method to do it semi-automatically by analysing the existing spatial relations between thematic and backdrop data, with the support of an ontology describing expected relations for given types of thematic features.

To support the conflation process that is part of integration, both spatial relations existing at instances level, and expected relations defined at types level, are used. Most data matching processes analyse existing spatial relations between instances. In the framework by Toomanian et al. (2013), some expected (prohibited, in their case) relations are defined at type level (acting as integrity constraints), as well as rules stating the expected relations between instances depending on the ones existing in data before integration. Feliachi et al. (2014) integrate data from the semantic web with topographic data based on the semantic of both data and hypotheses on expected spatial relations.

Once foreground and background data have been integrated, it is possible to generalise the combination. Making the relations existing at instances level between the thematic and backdrop features explicit is a form of data enrichment that can help guide generalisation.

More precisely, with a constrained-based approach (as defined by Harrie and Weibel (2007)), constraints can be defined that enable to identify expected relations at instances level depending on initial ones. Such constraints could encourage to keep a particular feature rather than to prune it so that the relation does not disappear, of to modify the thematic feature after generalising the backdrop feature so that the final expected relation is respected. Previous studies have specifically concentrated on the representation and processing of relations at instance level during generalisation in constraint-based approaches, to preserve them at best (Gaffuri et al. 2008) or even to exaggerate them (Duchêne et al. 2012). Edwardes (2007, p.169) deals with the graphic generalisation of foreground point data while preserving their topological relations with background data, while using a deformation grid constrained by meshes of the road network.

Instead of generalising the integrated (foreground + background) combination, generating a lower LOD can be done while replacing the backdrop data with a lower LOD backdrop dataset and re-locating the foreground thematic data on it. This is named thematic data migration by Jaara et al. (2013), who propose a method based on data matching between the backdrop datasets, and a multi-criteria decision approach to relocate the thematic data on the matched features of the target dataset. Existing relations are made explicit at instances level in the larger LOD combination, and conditional transformation rules enable to express expected relations at lower LOD at instances level as well. Stern and Sester (2013) study the migration of polygonal data from a large scale to a small scale map, guided by statistical rules defining expected relations at the data type level (rather than on features). These rules are inferred from the relations identified at instances level at large scale. Edwardes (2007, p. 169-170) proposes a process to migrate thematic data using an elastic grid: a grid constrained by the backdrop objects (road partitions) is created at initial and target LODs, and the foreground objects are migrated by a Laplacian transformation. Here only the relative position between each foreground object and the border of the mesh it belongs to are considered.

Finally, Gaffuri (2012) defends the idea of computing and storing explicit relations between (possibly multiple) thematic and backdrop layers in the case of web mashup maps, so that maps presenting only features matching a given spatial query can be quickly compiled – e.g. a maps of pizzerias that are close to metro stations.

3.3 How are relations between foreground and background instances modelled?

To our knowledge, only a few studies have specifically dealt with the modelling of relations between thematic and backdrop data at instances level. Touya et al. (2012, 2014) proposed an ontology-like modelling of spatial relations relevant for on-demand mapping and generalisation and, based on a survey of previous studies, a taxonomy of commonly defined constraints on those relations. This model has been extended by Jaara et al. (2013) for the case of foreground-background relations, and used for thematic data migration of points hosted by a road network. The study by Maudet et al. (2013) explored the adaptation to generalisation of a framework from the multi-agent domain that models the case where some objects “host” other objects and influence their behaviours. It was first experienced with topographic data only, but using it for backdrop data hosting thematic data was identified as a perspective and is currently under study. In the context of 3D city models, Bucher et al. (2012) proposed an extension of the CityGML model with a framework to describe relations between features, both at geometric level and application level (closer to the user's understanding).

4 Improving multi-sourced DLMs and DCMs using explicitly modelled relations: a first list of open research issues

From the above discussions, we identify a first list of open research issues related to the use of explicit base-foreground relations in thematically multi-sourced maps (and, actually, DLMs), which we hope can be discussed and enriched during the workshop :

- Is it possible to set up a taxonomy of thematic-background relations commonly found in maps, expressed not only at geometric level but also with some semantic attached (e.g. points along a network can be events that can be aggregated, or landmarks for which aggregation has no meaning)?
- Can we exhibit computational methods to identify existing foreground/background relations?
- Is it possible to formalise generic knowledge describing how a given relation at a given LOD is "allowed" to be transformed at a less detailed LOD? Does it vary a lot with the use case (i.e. data types * intended use)?
- Is it possible to associate typical relations with typical LODs, in the same way as typical representations of features are associated to CityGML LODs ? How much does it depend on the considered used case?
- Indeed, is it possible to exhibit typical use cases?
- How much can we help a user annotate his own thematic data with metadata describing their semantics, based on thematic-backdrop relations automatically detected in the data?
- How should the presence of thematic data influence the generalisation of the backdrop data? In particular in the case of selection, how does it modify the importance of the features?
- Can we set up SDIs, based on an MRDB DLM/DCM structure for backdrop data and an ontology of thematic-backdrop relations, that would be able to integrate external thematic data at the relevant LOD, and derive meaningful maps at lower LODs? Could these kinds of processing be compatible with on-the-fly mapping? How to combine data migration and generalisation to achieve this?
- If the base data are pruned at any LOD according to the thematic of the foreground layer, how much would the base data of a given LOD vary from one thematic to another one? Is it possible to exhibit "universal" base data?
- Managing an MRDB means having explicit vertical links between different representations of the same features in the different LODs. What could we still do if we do not have such vertical links, but still different base datasets at different LODs?
- How much are the integration of thematic data at a given LOD and the generalisation of the map different, depending whether "hosting" relations with the base data can be identified, or only "peer-to-peer" relations?

5 References

Balley S., Baella B., Christophe S., Pla M., Regnauld N., Stöter J. 2014. Map specifications and user requirements. In Burghardt D., Duchêne C., Mackaness W. (eds). *Abstracting Geographic Information in a Data Rich World. Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography, Springer, Berlin, Germany, 2014, Chap. 2, p.17-52.

Balley S., Jaara K., Regnauld N. 2012. Towards a Prototype for Deriving Custom Maps from Multisource Data. 15th ICA Workshop on Generalisation and Multiple Representation - jointly organised with EuroSDR Commission on Data Specifications, Istanbul, Turkey, 13-14 September 2012, http://generalisation.icaci.org/images/files/workshop/submissions2012/genemr2012_session1_sub1.pdf

Bereuter P., Weibel R. 2013. Real-time generalization of point data in mobile and web mapping using quadtrees. In Cartography and Geographic Information Science, vol. 40 no. 4, p. 1-11.

Bobzien M., Burghardt D., Petzold I., Neun M., Weibel R. 2008. Multi-representation Databases with Explicitly Modeled Horizontal, Vertical, and Update Relations. Cartography and Geographic Information Science, 35(1), 2008, pp. 3-16.

Bobzien M., Morgenstern D. 2003. Abstracting model generalization: a comprehensive approach towards automated generalization. *Proceedings of the 21st International Cartographic Conference*, Durban, Afrique du Sud, 10-16 August 2003, 136-145.

Bucher b., Falquet G., Clementini E., Sester M. 2012. Towards a typology of spatial relations and properties for urban applications. In T. Leduc, G. Moreau and R. Billen (Eds.) *Proceedings of Usage, Usability, and Utility of 3D City Models – European COST Action TU0801*, Nantes, France, October 2012.

Das T., van Elzakker C., Kraak M. 2012. Conflicts in Neogeography Maps. In *Proceedings of AutoCarto'2012*, Columbus, Ohio, USA - September 16-18, 2012, 14p.

Duchêne C., Ruas A., Cambier C. 2012. The CARTACOM model: transforming cartographic features into communicating agents for cartographic generalisation. *International Journal of Geographical Information Science*, Volume 26, Issue 9, September 2012, pp. 1533-1562.

Edwardes A. 2007. Re-placing Location: Geographic Perspectives in Location Based Services. PhD Thesis, University of Zürich, 2007.

Feliachi A., Abadie N., Hamdi F. 2014. Intégration et visualisation de données liées thématiques sur un référentiel géographique, Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances, RNTI-E-26, 35–46.

Gaffuri J. 2011. Improving Web Mapping with Generalization. *Cartographica*, vol. 46, no.1, p.83-91.

Gaffuri J. 2012. Toward web mapping with vector data. In: Xiao N, Kwan M-P, Goodchild M, Shekhar S (eds) *GIScience 2012. 7th international conference on geographic information science*, Columbus, September 2012. Lecture Notes in Computer Science, vol 7478. Springer, Berlin, p 87.

Gaffuri J., Duchêne C., Ruas A. 2008. Object-field relationships modelling in an agent-based generalisation model. *11th ICA Workshop on progress in automated map generalisation*, Montpellier, France, 2008.

Gould N. & Cheng J. 2013. A prototype for ontology driven on-demand mapping of urban traffic accidents. *Proceedings of the 16th ICA Workshop on Generalisation and Multiple Representation, jointly organised with the ICA Commission on Map production and Geo-Business*, Dresden, August 23-24.

Harrie L. and Weibel R. 2007. Modelling the Overall Process of Generalisation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications*. W. Mackaness, A. Ruas, L.T. Sarjakoski (editors), 2007, p.67-87.

Jaara K., Duchêne C., Ruas A. 2011. Toward the generalisation of cartographic mashups: Taking into account the dependency between the thematic data and the reference data throughout the process of automatic generalisation. *14th ICA Workshop on Generalisation and Multiple Representation - jointly organised with ISPRS Commission II/2 Working group on Multiscale Representation of Spatial Data*, June 2011, Paris (France), 7p., http://generalisation.icaci.org/images/files/workshop/workshop2011/genemr2011_Jaara.pdf.

Jaara K., Duchêne C., Ruas A. 2013. Preservation and Modification of Relations Between Thematic and Topographic Data Throughout Thematic Data Migration Process. *Cartography from Pole to Pole, Selected Contributions to the 26th International Conference of the ICA, 25-30 August 2013, Dresden*, Lecture Notes in Geoinformation and Cartography, M. Buchroithner, N. Prechtel, D. Burghardt (Eds.), 103-118.

Klien, E. Semantic Annotation of Geographic Information. PhD thesis. Institute for Geoinformatics, University of Muenster, 2008

Mackaness W. & Edwards G. 2002. The Importance of Modelling Pattern and Structure in Automated Map Generalisation. *Joint ISPRS/ICA workshop on "Multi-Scale Representations of Spatial Data"*, Ottawa, Canada, 2002.

Maudet A., Touya G., Duchêne C., Picault S. 2013. Improving multi-level interactions modelling in a multi-agent generalisation model: first experiments. *Proceedings of the 16th ICA Workshop on Generalisation and Multiple Representation, jointly organised with the ICA Commission on Map production and Geo-Business*, Dresden, August 23-24.

Moseme M., van Elzakker C. 2012. Neogeography Map Users and Uses. In *Proceedings of AutoCarto'2012*, Columbus, Ohio, USA - September 16-18, 2012, 13p.

Mustière S., Zucker J.-D., Saitta L. 1999. Cartographic generalization as a combination of representing and abstracting knowledge. In *Proceedings of the 7th International Symposium on Advances in Geographic Information Systems, ACMGIS*, Kansas City, USA, p. 162-164.

OGC 2014. Glossary of Terms, Open Geospatial Consortium. http://www.opengeospatial.org/ogc/glossary (accessed August 2014).

Papadias D. & Theodoridis Y. 1997. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographic Information Science*, vol.11, n°2, 1997, pp.111-138.

Ruas A. & Mackaness W. 1997. Strategies for urban map generalisation. In *Proceedings of the 18th International Cartographic Conference*, Stockholm, 1997, pp. 1387-1394.

Sester M., Jokar Arsanjani J., Klammer R., Burghardt D., Haunert J-H. 2014. Integrating and Generalising Volunteered Geographic Information. In Burghardt D., Duchêne C., Mackaness W. (eds). *Abstracting Geographic Information in a Data Rich World. Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography, Springer, Berlin, Germany, 2014, Chap. 5, p. 119-155.

Stern C., Sester M. 2013. Deriving constraints for the integration and generalization of detailed environmental spatial data in maps of small scales. *Proceedings of the 16th ICA Workshop on Generalisation and Multiple Representation, jointly organised with the ICA Commission on Map production and Geo-Business*, Dresden, August 23-24,

Toomanian A., Harrie L., Mansourian A., Pilesjö P. 2013. Automatic integration of spatial data in viewing services. Journal of Spatial information Science, n°6 (2013), pp. 43–58, doi:10.5311/JOSIS.2013.6.87.

Touya, G., Balley S., Duchêne C., Jaara K., Regnauld N. and Gould N. 2012. Towards an Ontology of Generalisation Constraints and Spatial Relations. *Proceedings of 15th ICA workshop on generalisation and multiple representation.*

Touya G., Bucher B., Falquet G., Jaara K., Steiniger S. 2014. Modelling Geographic Relationships in Automated Environments. In Burghardt D., Duchêne C., Mackaness W. (eds). *Abstracting Geographic Information in a Data Rich World. Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography, Springer, Berlin, Germany, 2014, Chap. 3, p. 53-82.