Evolving from automating existing map production systems to producing maps on demand automatically

Nicolas Regnauld
Ordnance Survey Research Labs
Nicolas.regnauld@ordnancesurvey.co.uk

Abstract

Despite the potential benefits, automatic generalisation capabilities are seldom used in today's map production systems. One of the main reasons is the difficulty to reproduce existing styles achieved manually, necessary for gradually introducing automation in current production systems without spoiling the product. Another reason is the large investment and high risk associated with generating new products from scratch. We propose a strategy where we will gradually introduce automation into current systems, while ensuring that these automatic components will be reusable in the future for generating products on demand. We cannot cut the development costs, so we want to focus on maximising the benefits from it. We propose an architecture for a system that would derive on-demand products automatically. By formalising all the components of this system we can develop our short term generalisation solution in accordance to these formalisms. This should gradually build-up our generic derivation system.

Keywords: generalisation on demand, ontologies, Web services

Introduction

Automated generalisation has been an active research field for more than twenty years. Lots of topics have been explored, like the development of geometric algorithms, database modelling for storing multiple representations (at different resolutions) of geographic data, knowledge modelling for the use of systems capable of selecting the appropriate transformation where and when it is required. However, automated generalisation is still rarely used in today's map production. Some examples exist (Lemarie 2003), but they do not provide full automation and the route to setup these systems is complex. In this paper, we want to discuss the potential benefits that automated generalisation can bring to Ordnance Survey, but also what makes it difficult for automated techniques to find their way in the production systems. Based on this, we have defined a strategy to gradually improve the level of automation in our production lines. We have then adjusted our research plans to support this strategy.The first section of this paper discusses three different benefits that automated generalisation can bring to a map producer like Ordnance Survey. Section 2 presents different scenario for introducing automated generalisation in map production systems. Section 3 proposes a strategy to gradually introduce automation in our production systems, which leads to the review in section 4 of the research that exists or is required to support it.

1. Benefits automatic generalisation

The obvious main benefit of automating the generalisation process for a mapping company is the reduction of its production costs. This indirectly generates other potential benefits. Lower production cost can mean opportunity to create products targeting a narrower market. From cost reduction to increased flexibility, we see three main stages in the evolution of generalisation capabilities, each one bringing opportunities for new revenue streams.

- Stage 1: automating the production of known products, from known data. This is the natural first step for a company making maps manually: automate the process creating maps to reduce production costs. For this sole purpose, there is no need to model any cartographic rules that would ensure that the design of the target specifications is satisfactory. The design has been done manually when the map was designed, the automatic process is only trying to replicate it. It requires the appropriate algorithms for generating an adequate representation of the data. It also requires some cartographic knowledge that triggers these algorithms where and when required, with the appropriate parameters.

- Stage 2: automating the derivation of new products from known source data. The difference with the stage 1 is that the system is not built to derive a particular product, but it is built to respond to specifications and derive what is requested. This requires additional modelling to describe the product targeted (formalisation of the specifications of the product), and to represent the knowledge that will allow the system to trigger the generalisation algorithms when they are required: The data will be generalised when the system detects that they do not meet the specification, as opposed to the stage 1 where hard coded sequences of operations can be setup for a particular product.

- Stage 3: automating the generalisation of data coming from different sources. In this context, the difficulty is that the generalisation system can no longer exploit the schema of the source data, it can be anything. In addition, the map producer is no longer the body deciding of the creation of the map. The customer wants to see some data that he holds in a context that the data held by the map producer can provide. The purpose of the map dictates what contextual information is relevant. This requires:

- Modelling of the meaning of the data
- Formal description of the algorithms, to specify on what concept they work, and in which context.
- Modelling the user requirements and their automatic translation into some map specifications (that the generalisation system is able to understand)

These three levels have already been researched, extensively for the two first one, probably less so for the third one.

2. Scenarios for introducing automatic generalisation in production systems

At ordnance Survey, we have in the past years concentrated our efforts into creating a seamless database that stores our most detailed data (base data) centrally. The idea is that only this database is maintained, and all the products are derived from it, as automatically as possible. We are at a transition time, where our production lines do not use efficiently our base data. They are either raster-based, or rely on separate vector databases not connected to our base data. In order to improve efficiency and cross-product consistency, we are looking at the different options to migrate our

production systems so that they eventually all use data sourced from our central high resolution database. We are only looking at vector based systems, as our base data's format is vector. This will make the process of updating the products easier to automate. In addition, vector data offer much more flexibility to create variations of the product (different look and feel, by adjusting the representation of each class of feature). Let's review the options that we have for achieving this transition:
1- Automate the production of current products. The problem here is that we have very little flexibility for changing the specifications of the existing products. This means that there will be very specific constraints on the generalisation algorithms (in terms of what they need to achieve). This requires a high investment for developing the system, and we have a low confidence associated with it, as we do not know how close to the existing manual work our future automatic processes can come. In addition, manual interventions to correct the side effects of the algorithms will mean new tasks for the cartographers, which could require new software or at least customisation of the existing one to handle it efficiently. So adding some automatic generalisation in the production line will have an impact on the software used for cartographic edition, which can prove very costly if new software is required.
2-An alternative to this solution is to vectorise the current raster product, match the vectors to the base data, and then gradually automate the propagation of the updates from the base data to the vectors representing this data in the product. This strategy has often been studied by mapping bodies holding data at different resolutions, see for example (Badard 1999) or (Gösseln and Sester 2004). The advantage of this is that the investment happens gradually. The updating can be done manually at the beginning, then once a particular manual generalisation task is identified as very costly, development can be ordered to automate that particular task. It is not trouble free tough: the automatic process needs to produce a result that fits with data which have been drawn manually, and also reproduce their style. The initial vectorisation is also very costly, and the matching with the base data complex.
3- Automate the derivation of new products (maps on demand), to generate new revenues from our base data. In this scenario, we have much more freedom in terms of the specifications of the target product, but it requires more complexity in the system to make sure that the automatic system generates a product which is of adequate quality (follows the rules of cartography).

3. Strategy to introduce automatic generalisation at Ordnance Survey

Developing automated generalisation processes is a lengthy and costly process, often associated with high risks. The high risk is due to the nature of geographic data which represent geographic phenomena taking various forms. The more variety in these forms, the more difficult it is for an automatic process to cope well with them all. Processes can be designed and implemented to represent these phenomena at lower level of details, but they will generally only produce satisfactory results for a certain percentage of the cases. This percentage is very difficult to estimate before hand, and when it turns out to be lower than expected, the additional costs incurred can be huge (manual process for dealing with non satisfactory cases or implementation of new processes to deal better with some particular cases to increase this percentage). Investing into this technology is therefore not a simple matter. For Ordnance Survey the main driver for investing in this technology is to reduce production costs, which are very high. Plans for deriving new products with a high risk associated with the

estimate of the development costs, plus the risk associated with how well the new product will sell, is unlikely to get approved and funded.
Therefore we want to focus on how we could build up automatic generalisation capabilities into the production systems of our current products, while making sure that all the components developed for this purposed will be ready for reuse in a different context in the future. If we can achieve this, then the more we automate parts of the current processes, the better we will be for building news systems for new products. Cost estimates based on reusing available tools will have a much lower associated risk and will eventually become affordable.
Therefore the best way forward for us seems to be the solution 2. Being at the start of this process, we have the opportunity to ensure that any development made in this context can be reused in the future in the context of the option 3.

Today's challenge is: *What do we need to do to ensure that what is developed for automating the incremental update of our current products can be reused by a more advanced system that will be developed in the future to derive products on demand?*
We have sketched out an architecture for such a system, summarised in Figure 1. At this stage, the aim is to define all the components, how they need to me modelled and described so that the whole system can work. Then we can use this framework to develop the components necessary for our more pragmatic and urgent tasks, with the prospect of reusing them later in an on-demand production scenario.
This architecture is presented in more details in [Regnauld 2007], we only summarise it here. The core of the system (the generic generalisation system in the figure), at a high level, is putting together an application that can turn the data available into the product specified by the user. To do this it needs to understand the input data, understand the user requirements, and have access to the knowledge and tools for deducing the flow of processes required to generate the product. This architecture is very modular, as such a system will take several years to put in place. We want to give it all the chances of evolving with other technologies, to avoid having a system that relies on obsolete technologies when or even before it is completed. Also if it gets implemented, it will have to carry on evolving to provide a long term stable platform for producing maps on demand.

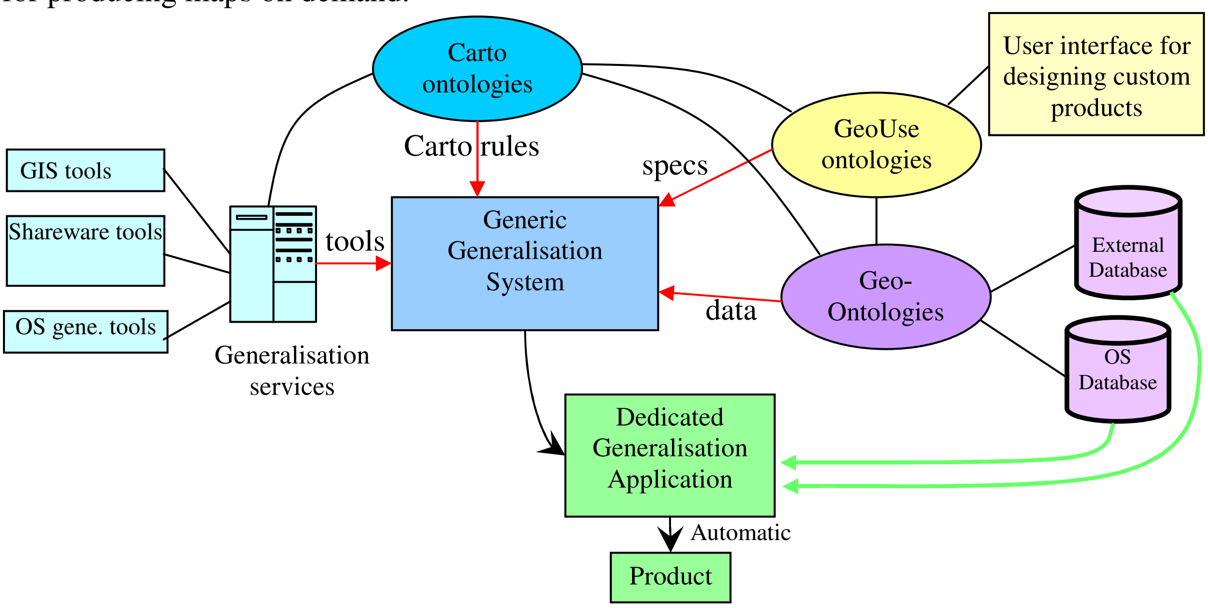

Figure 1: System architecture for tailoring on-demand products

4. Research to support this on-demand derivation system

In order to build the system proposed above, we need to research different areas. Fortunately, in all the areas mentioned here significant research has already been done and can be reused. This section tries to summarise for each component what existing research can be reused, and what needs to be done to adapt it or complement it for our purpose.

1. GeoOntologies: formalising the input data

Most of the generalisation systems or prototypes that we know today, work on a specific data model. If we want to allow a system to generalise data from multiple sources, we need to get away from this. We want to explore the use of ontologies to achieve this. (Dutton and Edwardes 2006) have recently shown the potential of Ontologies for conceptualising the generalisation process. Ontology has been an active research topic at Ordnance Survey for a few years. This topic has been investigated to respond to one of the challenges for Ordnance Survey in the near future, which is to enable third parties to integrate their data with the topographic data provided by Ordnance Survey (Mizen et al. 2005). The same technology has another advantage for us: it offers a way of breaking the dependency between the schema of the source data and the system that generalise it. The idea, as explained in (Goodwin 2005) would be to define a domain ontology where all the geographic concepts are described. These geographic concepts can describe the geography at different levels. The concepts described there would vary from very local ones, like buildings or road sections, to much more complex ones, like hydrology network, a town, or a country. The generalisation system would reason on these concepts. Another ontology, called a data ontology would provide the link between a particular data model and the concepts of the domain ontology. Then when the system requires some data, it would query the data ontology using references to concepts of the domain ontology. The data ontology would then translate the query into the schema used in the database, and instantiate the requested concepts as a result to the query. Any third party data, for being useable by the system, would need to have a data ontology associated with it to link it to the domain ontology. This required that the domain ontology evolves into an accepted standard. Even without considering the integration of third party data, modelling these concept using ontologies will achieve the independence between the system and the data. If the data model changes for some reason, then only the associated data ontology requires updating. The generalisation system will not be affected.

Once the principles of these ontologies have been defined, they need to be gradually populated. Research studies focusing on deriving high level geographic entities from topographic data, such as (Chaudhry and Mackaness 2005) or (Boffet 2000), can be reused to populate some concepts of these ontologies. The knowledge that they have explicated only needs to be translated into the languages chosen to describe the ontology (like OWL).

It is important to stress that ontologies for describing geographic information is in its infancy, and not ready for full scale exploitation yet. As mentioned in (Dolbear and Hart 2006), “Spatial and semantic technologies today are largely incompatible, and semantic technologies are poor at handling very large spatial databases”. However, we believe that the idea behind it can bring a much higher flexibility for using geographic data. Today’s limitations only mean that more research is required. Such research is currently ongoing, so these limitations will surely be overcome in the future. In the meantime we intend we have the possibility to emulate the functionality of these ontologies using a multi-representation database, which is a much better understood technology (Bédard et al. 2002, Vangenot 2004). We would use the conceptual level of the ontology to describe the concepts and their relationships, and instead of expressing it into OWL (which has too many limitations for the time being), we would populate these concepts in a multi representation database, by deriving them from the data we hold using traditional programming languages.

3. Geo-Use Ontology: formalising the user requirements

This ontology will describe different typical usages of geographic information or the tasks that the user want to perform with it. For each usage or task, there will be indication of what geographic entities are relevant and how important they are. The geographic entities will map those from the GeoOntology. For example, for a pedestrian navigation usage, networks of roads and paths are important, as are landmarks, which will be of different types in urban and rural contexts. All the concepts mentioned need to be defined and mapped to the GeoOntology. The context of use will also be important (on which device, in which conditions of light, at which distance is the map to be read, all this can influence the content of the map). This ontology need to make the link between high level requirements specified by a user and machine readable specification to drive the generalisation system.

Once the Geo-Use ontology has been designed, the user interface will need to be designed. This should allow the user to define what he wants either in broad terms (using the high level concepts of the geo-Use ontology) or in detail (describing his requirements using many more lower level concepts (scale, thresholds, relative importance between themes, etc.). Designing a map is not a simple process, as explained in (Forrest and Pearson 1990). User will rarely have enough cartographic knowledge to design a good map. So an approach like the one proposed by (Hubert and Ruas 2003), to guide the user through the design will be required. In their approach the system proposes sample maps to the user, who chooses the one closest to what he has in mind. The system, based on the user's choice and feedback, generates new samples to propose to the user. This iterative process converges towards the representation that the user has in mind.

4. Carto-Ontology: formalising the cartographic knowledge

All the cartographic rules or constraints which define the characteristics of a good map need to be stored in a machine readable format. The cartographic knowledge should also include a formal description of all the generalisation operators that can be necessary during a generalisation process. Numerous pieces of research (for example (Robinson et al. 1978, Brassel and Weibel 1988, Shea and McMaster 1989) have been conducted to define a set of generalisation operators (simplification, elimination displacement, etc.) or measures (minimum distance, sinuosity, granularity, etc) that can be used to classify the multitude of algorithm that can be designed. We would need to choose or refine one that can evolve towards a standard, so that new algorithms can be formally described (see below). The carto-ontology would also express the relation between the cartographic rules (or constraints), the measures used to assess them and the operators that can help to enforce (or resolve) them.

5. Generalisation services: formalising the description of the algorithms

In order for a machine to automatically trigger some generalisation algorithms, measures or other tools (to perform spatial analysis for example) sensibly, it needs to know what they do and when they are relevant. At least the following information needs to be recorded against each algorithm:

- Geographic concept on which it operates. This should directly relate to the geographic concepts of the domain ontology mentioned above.
- Type of operator or measure. Each algorithm implements an operator or a measure, some times a sequence. This information is critical as the carto-ontology only reason on operators and measures, it does not know about algorithms.
- Context of use. The context of use is also an essential piece of description for an algorithm. It should describe things like the scale range, the level of abstraction required for the input data and the output, information about the importance of the geographic concepts being generalised for the map (depending on the purpose of the map). All the concepts used here must map to those defined in the Geo-Use ontology.

If all the algorithms are described following the same standard formalism, the set of algorithms can grow without the carto-Ontology (generalisation knowledge) needing updating. The system should be able to pick-up the new algorithms when they are relevant (when they implement the operation required and work in the present context).

Formalising the description of these algorithms is particularly interesting if the set of algorithms is large. Until now, all the generalisation platforms proposing automatic generalisation capabilities have a limited number of algorithms available. Recent research done at the University of Zurich do propose a way forward to make the number of generalisation algorithms available grow. They have developed a solution using Web Services to allow developers to publish their algorithms as generalisation services (Neun and Burghardt 2005). Once published, these generalisation services can be called from within any development environment. The initial research as proved the concept by developing and publishing as web services algorithms developed in standard Java. These can then be called remotely from a different environment. At Ordnance Survey Research Labs, we have experimented with these web services to check whether it would be possible to publish as generalisation services some processes which run in specific proprietary environment (like a commercial GIS). We have tried successfully to publish and call a process that runs under Clarity (Laser-Scan 2003). This proves that this technology can be used to make commercial platforms interoperate. This is very interesting, as all platforms have their strengths and weaknesses. Instead of committing to one, take advantage of its strength and suffer from its weaknesses, we have now the possibility of developing each part of the system on the platform that provides the best support. This is feasible technically, there is now of course the licensing issue that needs to be examined. WebGen does provide a standard way of publishing platform independent algorithms, and also has the potential for allowing to publish platform dependant algorithms in the same way, offering in this way a mean to achieve platform interoperability. It would be a good idea to extend this technology with the formal description that we have discussed above. This would allow these services to be selected automatically by an automatic generalisation system.

6. Generic generalisation system

The core of this on-demand generalisation system uses all the other parts to put together a dedicated automatic generalisation process. It performs the following tasks:

- Data extraction

The core of the system will not know about the data at the beginning of the process. What it will know is the specifications of the product, and the types of source data that have been selected directly or indirectly by the user. With this it should be able to query the database through the GeoOntology, to retrieve the data.

- Reasoning

The rules or whatever form of knowledge that the system will have to work with are stored in the CartoOntology. The specifications of the product will drive the selection of the relevant ones. Then the core of the system will require a mechanism that can use the rules to choose and orders the operators that need to be applied, and make sure they converge towards a solution that fulfils or approaches the specifications.

- Selection of services

The result of the reasoning will determine a high level operation that needs to happen on a particular geographic phenomenon. The system will have to select the appropriate service that can perform this operation on this type of phenomenon, in the context of the product specifications.

- Report

At the end of the process, the system should be able to produce a result (a map or a dataset), and some reports about local problems which have not been solved satisfactorily. Unsolved problems can either be due to the system lacking adequate tools, or the knowledge base to be incomplete, or the target specified to be unachievable.

The core of the system could be a complex piece of software. However, little in these requirements is new. The AGENT project (Lamy et al. 1999) already implements most of this. It uses meso and micro agents to represent different levels of geographic concepts. Its reasoning capabilities are based on constraints, as proposed in (Ruas 1999), and is able to converge towards a solution and report on the quality of the result obtained. The main update that we want to use here is that the constraints would not propose direct algorithms, but operators. Then an additional module is required to select the appropriate algorithm that implements this operator, based on the context of use. This way of decomposing the problem of choosing an algorithm is similar to what was proposed in (Mustière 2005) . The constraints will then be much more generic, allowing the system to reuse them in different contexts, which is essential if we want the system to derive products on demand.

Conclusion

In this paper, we have briefly discussed the difficulties of introducing automatic generalisation processes in existing map production units, particularly for Ordnance Survey, despite the benefits that it can bring, both in terms of reduction of production costs and increased flexibility to generate new revenues from new derived products.

Based on this evidence, we have set-up a strategy to gradually introduce automatic generalisation capabilities in our production lines. We have sketched the architecture of a system designed for deriving automatically cartographic products from multi source geographic data. The components of this system can all take advantage of existing research. While we don not want to develop this system straight away, we want to understand all its component so that any development made for automating parts of the current production process, are described and packaged in a way that they can be reused by the system later. For this to happen, all the components of the generic system need to be formalised. We propose to look at the research in semantic modelling, and ontologies in particular to achieve this.

The key point here is that being at the beginning of developing automatic capabilities for our production systems, and knowing the cost of developing them, we want to make sure that we can make the best use of them in the future. We want to avoid developing processes that get buried in a black box and die with it when the technology it relies on becomes obsolete.

Reference List

Badard, T., 1999, On the automatic retrieval of updates in geographic databases based on geographic data matching tools. *Proceedings of the 19th International Cartographic Conference*, 47-56.

Bédard, Y., Proulx M.J., Larrivée, S., and Bernier, E., 2002, Modeling Multiple Representations into Spatial Data warehouses: A UML-based Approach. *Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, and the ICA Commission on Map Generalization*, Ottawa, Canada, http://www.isprs.org/commission4/proceedings02/pdfpapers/361.pdf.

Boffet, A., 2000, Creating Urban Information for Cartographic Generalization. *Proceedings of The 9th International Symposium on Spatial Data Handling*.

Brassel, K. E. and Weibel, R., 1988, A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems*, **2**, No. 3, 229-244.

Chaudhry, O. and Mackaness, W. A., 2005, Visualisation of Settlements Over Large Changes In Scale. *ICA Workshop on Generalisation*, A Coruna, Spain, 7-8 July, http://ica.ign.fr/Acoruna/Papers/Chaudhry_Mackaness.pdf.

Dolbear, C. and Hart, G., 2006, Combining spatial and semantic queries into spatial databases. *poster presented In: International Semantic Web Conference*.

Dutton, G. and Edwardes, A., 2006, Ontological Modeling of GeographicalRelationship for Map Generalization. *ICA generalisation workshop*, Portland, USA, June 2006, http://ica.ign.fr/Portland/paper/ICA2006-Dutton-Edwardes.pdf.

Forrest, D. and Pearson, A., 1990, Map design. In *Information Sources in Cartography* (London:.Bowket-Saur), pp. 168-188.

Goodwin, J., 2005, What have Ontologies ever done for us - potential applications at a National Mapping Agency. *OWL: Experiences and Directions*, Galway, Ireland, 11-12 November, http://www.mindswap.org/OWLWorkshop/sub18.pdf.

Gösseln, G. V. and Sester, M., 2004, Integration of geoscientific data sets and the german digital map using a matching approach. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **35**.

Hubert, F. and Ruas, A., 2003, A method based on samples to capture user needs for generalisation. *Fifth Workshop on Progress in Automated Map Generalization*, Paris, 28-30 April, http://www.geo.unizh.ch/ICA/docs/paris2003/papers/hubert_et_al_v0.pdf.

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W. A., and Weibel, R., 1999, The Application of Agents in Automated Map Generalisation. *Proceedings of the 19th International Cartographic Conference*, 1225-1234.

Lemarie, C., 2003, Generalisation process for Top100: research in generalisation brought to fruition. http://www.geo.unizh.ch/ICA/docs/paris2003/papers/lemarie_v0.pdf.

Mizen, H., Dolbear, C., and Hart, G., 2005, Ontology Ontogeny: Understanding how an Ontology is created and developed. *Proceedings of the first International Conference on GeoSpatial Semantics*.

Mustière, S., 2005, Cartographic generalization of roads in a local and adaptive approach: A knowledge acquistion problem. *International Journal of Geographical Information Science*, **19**, No. 8-9, 937-955.

Neun, M. and Burghardt, D., 2005, Web Services for an Open Generalisation Research Platform. *8th ICA WORKSHOP on Generalisation and Multiple Representation*, A Corune, Spain, July 7-8th, http://ica.ign.fr/Acoruna/Papers/Neun_Burghardt.pdf.

Robinson, A. H., Sale, Randall D., and Morrison, Joel L., 1978, *Elements of cartography* (New York: Wiley).

Ruas, A., 1999, Modèle de généralisation de données géographiques à base de contraintes et d'autonomie, PhD thesis, Université de Marne la Vallée, France.

Shea, K. S. and McMaster, R. B., 1989, Cartographic generalization in a digital environment: When and how to generalize. *Proceedings of AUTO-CARTO 9*, 56-67.

Vangenot, C., 2004, Multi-representation in spatial databases using the MADS conceptual model. *ICA Workshop on "Generalisation and Multiple representation"*, Leicester, UK, 20-21 August 2004, http://ica.ign.fr/Leicester/paper/Vangenot-v2-ICAWorkshop.pdf.