2003
Cambridge Conference
www.cambridgeconference2003.com
Ordnance Survey
Romsey Road
SOUTHAMPTON
SO16 4GU
United Kingdom

Efficient Map Production By Re-Engineering and Generalising Your Data Assets

P Hardy
Laser-Scan Ltd
United Kingdom

M Meyer
National Survey and Cadastre
Denmark

Paper 4C.2

Ordnance Survey

Efficient Map Production By Re-Engineering and Generalising Your Data Assets

P Hardy
Laser-Scan Ltd
United Kingdom

M Meyer
National Survey and Cadastre
Denmark

Abstract

National Mapping Organisations are under ever stronger pressure to be self-funding, and hence to cut costs. At the same time, demands are increasing from users for up-to-date mapping at all scales, and for intelligent spatial data, that can underpin the new location-based services and e-government initiatives.

In the past, producing new data products, or small-scale mapping required enormous effort in manpower. To create the required structured data and appropriate mapping using the established GIS and digital mapping software tools has been a major challenge, which has not often been economically viable. However, there are now appearing new software products based on intelligent spatial objects and software agents, which for the first time provide the necessary tools and infrastructure to make these tasks economically possible.

This paper overviews the new generalisation and data re-engineering capabilities of a spatial data manipulation product. It shows how they are being used by NMOs such as Ordnance Survey Great Britain for large scale spatial data re-engineering to create new framework datasets, but also by NMOs in Denmark, France and elsewhere to derive automatically generalised mapping for smaller scale products. It provides a taxonomy of the various stages of spatial data re-engineering operations available. It then summarises the situation of KMS, Denmark as an example of an NMO who are currently developing a workflow for the generalisation of their existing topographic database from 1:10,000 to 50,000. It also covers the advantages and gains of NMOs banding together to influence the development of commercial products, as exemplified by the MAGNET consortium on generalisation.

Introduction

Rationale

Over the past few years, many national and regional mapping organisations have built a basic scale dataset covering their territory. The scale of such datasets varies from country to country. Britain is unusual in having a consistent dataset at scales of 1:1250 (urban) and 1:2500 (rural). Countries such as Denmark, Netherlands, France have 1:10,000 as their basic scale dataset. In other countries like Germany, the basic scale landscape model (Base DLM) is at about 1:10,000, but is produced separately by the regional mapping organisations.

This basic data was usually captured to automate the production of basic scale paper mapping. Other uses such as analysis in customers' GIS systems were secondary. It has emerged that such GIS applications require cleaner and more structured data than that for cartography, necessitating extensive data re-engineering such as that done by the Ordnance Survey to produce OS MasterMap [Hardy 2001] and [Curtis & Painter 2002]. Using the basic scale data as a starting point for automated derivation of other scales of mapping, and for GIS data products has always been on the agenda for these organisations. However, it has been pushed back while the urgent task of getting complete basic scale coverage was completed, and an updating service established.

While this has been going on, there have been increasing requirements for rapid creation of multiple products at different scales and specifications, combined with great pressures on organisations to cut costs and reduce staffing. An example is KMS Denmark, which has had to produce up-to-date 1:50K mapping for its military, at the same time as reducing staff and restructuring the organisation. These conflicting requirements have given a real incentive to automate the processes of re-engineering and generalisation.

Generalisation

Early attempts at automatic generalisation involved a GIS or digital mapping software application applying geometric algorithms, one at a time, to simplify, displace, exaggerate, aggregate, collapse or otherwise modify an individual map feature. The limitation of this approach is that the algorithms operate in isolation - a shortcoming that entailed heavy manual intervention by expensive cartographers to complete the process.

A major step forward was the introduction of object-oriented data models and associated object-oriented spatial toolkits. This enabled algorithms to operate in the context of the feature, so that it could use information about its neighbourhood to modify the effect of the algorithm.

The third generation of generalisation involves software agents [Lamy et al 1999]. In an agent-based system, the individual map features are activated as self-aware software entities, which strive to improve their 'happiness'. Instead of trying to specify which algorithms should be applied to which features; the user defines constraints and associated measures, and lets the agent decide for itself.

The real advances in automated generalisation came with the realisation that good generalisation requires:

- Contextual analysis - you can't generalise one map feature at a time in isolation. You have to consider groups of features as a whole.
- Adaptive processing - you can't apply a single algorithm to all features (even of a single class). You have to choose appropriate algorithms according to the circumstances of that feature.
- Backtracking - you can't get it right first time every time. You have to be prepared to assess whether an operation has made things better or not, and be prepared to undo it and try something else.

To implement these requirements necessitates a holistic rather than procedural generalisation environment [Hardy 2000]. In addition, generalisation is not just applicable to visual cartography (paper maps, or web or mobile screen). Increasingly, mapping organisations are being driven to derive lower resolution GIS data products from their framework data.

A current example of non-visual generalisation is in Germany, where the German Länder have captured a Base DLM in the ATKIS data model, but find that many GIS users find this too large data volume. Hence their current project to derive a 1:50K landscape model by 'model generalisation', which needs a subset of the techniques needed for cartographic generalisation. In particular, because the target dataset has no defined cartographic representation, there is no requirement to exaggerate thin lines, so no need for displacement. [Birth, 2003]

Beyond the visual map and the GIS dataset, generalisation will become increasingly important to the new generation of 'Location-Based Services' (LBS) [Hardy & Haire 2000]. These services, provided to the new media like mobile phones, PDAs and tablets, will need to extract and present only the information that is relevant at the time to a mobile customer.

Agents and the AGENT project

The AGENT (Automated GEneralisation New Technology) research project [Ruas 2000] was initiated as an EC funded project (Esprit 24939) in the late 1990s. Laser-Scan joined the multi-national consortium as the software supplier. Other members included IGN (the French national mapping organisation in Paris) in the lead role, and three universities: Zurich and Edinburgh (for their expertise in geography and cartography), and Grenoble (for its work in artificial intelligence). Three years of intensive research and development resulted in a prototype agent-based generalisation system, based on Laser-Scan's Gothic LAMPS2 object-oriented software.

The AGENT system incorporates a strategy which aims to resolve conflicts not by describing in detail how certain conflicts should be resolved, but by describing the desired final characteristics of the feature. For example, the system might incorporate the desired outcome that no building should be smaller than a certain size, that it should not be closer than a certain distance to another building and that it must not be moved by more than a certain distance from its original starting position. The outcome of any generalisation operation or set of operations is compared against this set of guidelines. This comparison is used to decide whether further operations are required, or whether the result should be discarded and a different operation applied. In this way, individual map features can be generalised in a way that is sensitive to their particular situation, with similar features potentially having quite different operations applied to them.

The AGENT prototype was made available in the commercial LAMPS2 Generaliser product and has been extended for production use at several NMOs including KMS Denmark [Sheehan 2001] and IGN France. For more on the Agent project and the resultant technologies, see [Lamy et al 1999] and [Ruas 2000].

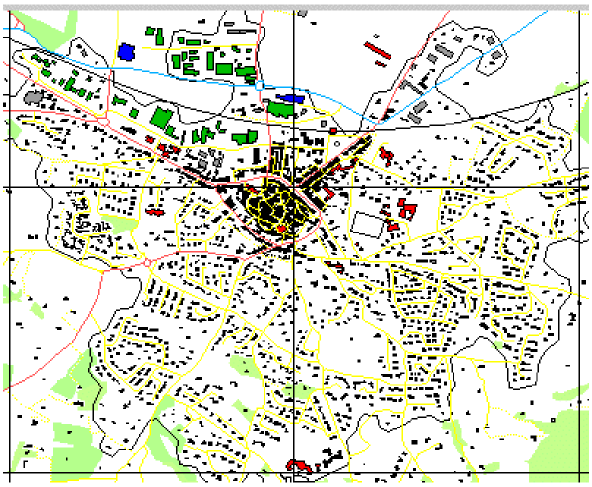


Fig. 1 - Framework data before generalisation

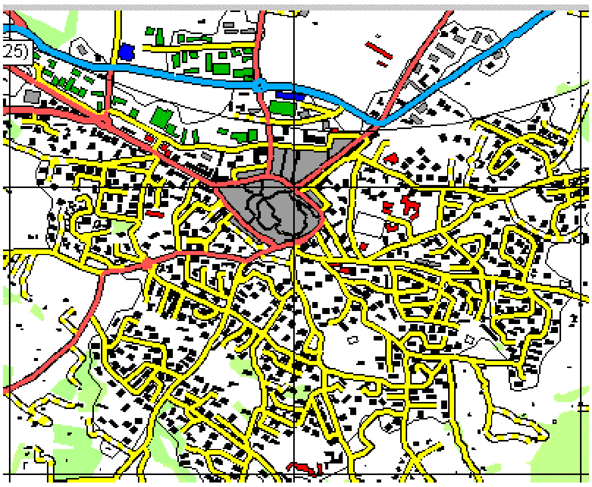


Fig. 2 - After Agent-based generalisation

Data copyright IGN France, presentation courtesy of the Agent project

Relationship of Generalisation to Data re-engineering

Early experience in attempting to automate generalisation has shown conclusively that you need a better starting point than spaghetti graphical data. Good generalisation requires clean data in a good data model, with the following properties:

1) Features with different characteristics must be in separate feature classes. It's no use just having a class 'roads', if you want motorways to behave very differently to urban streets during generalisation.

2) Feature coding must be consistent. If a line between a road and a house is sometimes in class 'fence or wall', sometimes in class 'building', and sometimes in class 'road edge', then good generalisation will be difficult.

3) Data must be geometrically clean. Generalisation is hard enough, without having to guess whether two lines should join or not, or where an area with two seed points should really be split.

4) Data must be continuous. Trying to generalise data a sheet or tile at a time and then join it afterwards introduces immense problems.

5) Structure must be explicit stored, for fast traversal. Generalisation requires repeated analysis of feature relationships, particularly neighbourhood relationships. Such analysis can be done using repeated spatial index searches followed by repeated geometric analysis, but the time taken is likely to be unacceptable. Much better is to build explicit structure before generalisation.

The above properties are needed for generalisation, and for many GIS data products, but not necessarily all needed for long-term storage. So it is usual to precede the generalisation processing stage with a 'data re-engineering' stage to get the data cleaned, structured, and into a suitable object data model [Hardy 2001]. Data re-engineering also has internal and external benefits beyond generalisation, as described in the following section.

Data Re-Engineering

Scope of data re-engineering

Many national and regional mapping organisations are only just becoming aware of the need for data re-engineering, while a few have already led the way with substantial projects. A few other NMOs like KMS Denmark do not need to re-engineer legacy data, as they have recently captured fresh data from photogrammetry.

Many of the examples below are taken from the flowline set up by Ordnance Survey GB in 2000 under their DNF (Digital National Framework) project to re-engineer the LandLine dataset of Britain. This flowline created a new continuous object dataset called OS MasterMap [Curtis & Painter 2002]. The flowline was built using Laser-Scan's LAMPS2 object-oriented geoprocessing tool. It took the 230,000 separate map sheets of 1:1250 and 1:2500 scale linework, and within one year generated a continuous dataset of 400,000,000 objects. See Figure 4 below.

Spatial data re-engineering as involved in DNF covers a wide range of processes, and these are described below: Not all steps are necessarily needed, dependent on the quality and modernity of the starting-point framework data. Also multiple steps may be achieved by a single set of processes. However, if multiple steps are to be applied, they are listed below in the logical order.

Taxonomy of data re-engineering

The following subsections describe the different stages of spatial data re-engineering now available, and highlight the benefits to the organisation and to the customer base.

Continuous Dataset

The first action is to merge map sheets and data tiles into a continuous dataset with no artificial breaks. This avoids the typical scenario where anything of interest happens at the intersection of four map sheets! The benefits are both external and internal:

- *External*: Satisfy customers by providing mapping on-demand, centred on their location. Enable new products and services such as Location-Based Services (LBS).
- *Internal*: Easier and lower cost maintenance (no edge match, no problems of time discrepancy between adjacent sheets).

Geometry Cleaning

The next step is to detect and correct mistakes and inconsistencies in the geometry of features. This removes overshoots, undershoots, slivers, kinks, and spikes, created previously by inaccuracies and errors in digitising and edit. The benefit is that the resultant data is much more appropriate for customer use and further analysis, so has added value. Note that modern non-graphical applications like routing or cadastral analysis are much less tolerant of geometric errors than is a human map-reader.

Coding Consistency

Next step is to detect and correct coding mistakes and inconsistencies accumulated over years in the data capture process. When OSGB did the processing for OS MasterMap, they reportedly found and fixed 60 million errors in the data. Some of these were geometric, but many were incorrect feature coding. Again, the benefits are in quality of data and ability to sell it for GIS analysis.

Meaningful Objects

A very important step is to create meaningful objects in the data, which correspond to the entities that the user wants to deal with. In the OS MasterMap case, this meant creating objects for houses, gardens and roads, rather than the walls, fences and kerbs of the previous LandLine product.

Unique Ids

Closely bound with the creation of meaningful objects is assigning them unique IDs. In the OS MasterMap case, this is a TOID (Topographic Object IDentifier) - a 16-digit number that identifies this particular house, garden or road throughout its life. IDs are very important to getting the benefits of re-engineered data, as they act as the key by which other operations can refer to this object. In particular they allow:

- Incremental update - once a continuous dataset has been built, updates are no longer restricted to shipping replacement tiles or sheets. Users can request updates whenever they want, for whatever area they want, so that they always have up-to-date data and mapping. Unique IDs are used to say which objects have been added, modified or deleted since the previous update.
- Value-added user data - IDs are the way that users link their business information to your framework map data, be it addresses, ownership, planning information, environmental details, or whatever.
- Collections of objects - Often users want to deal with bigger entities than your framework objects. Use of IDs allows them to handle cleanly complex objects such as a hospital made up of several buildings.

Topology & Structure

Topology is the mathematical concept of spatial structure. It expresses explicit geometric relationships, such as “connects to, touches, adjacent to, within”. A topological model consists of nodes, links, and faces, with references linking them to form a continuous structure. The real-world objects then refer to these primitives, to identify relationships including:

- Shared edges between land polygons.
- Junctions between streets in the road network.
- Colinearity of administrative boundaries with roads and streams.
- Adjacency of buildings to roads.

A good study of the use and management of topology is “The Balance between Geometry and Topology” [van Oosterom et al 2002]. An example of Dutch data (courtesy Topografische Dienst) is displayed as Figure 3 below, showing the shared edges between polygons and the node points where edges meet. The OS MasterMap data, shown as Figure 4 below (Crown copyright) is also fully topologically structured.

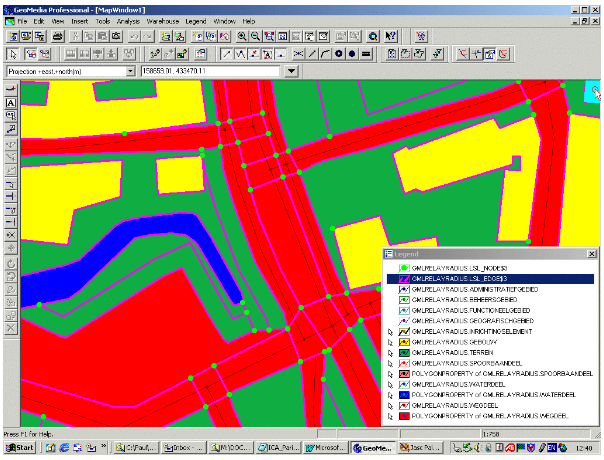


Fig. 3 - Topology - shared edges and joining nodes

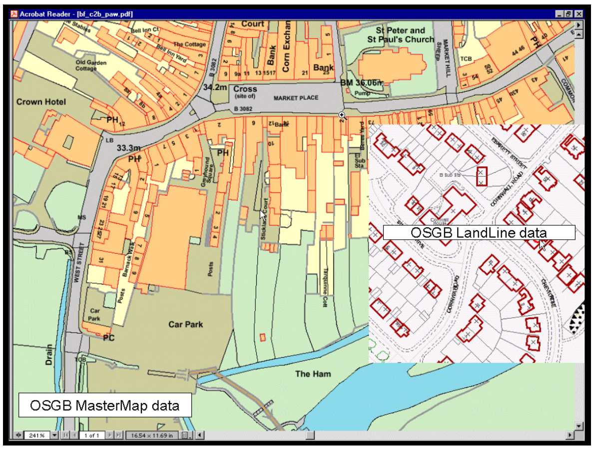


Fig. 4 - OS MasterMap data re-engineered from LandLine

Re-engineering your spaghetti data into explicit topology has substantial benefits, both internal and external:

- Topology enables reliable use of the framework data for new applications like routing.
- The explicit structure can greatly speed up further analysis. Operations like “find all the urban areas that are adjacent to river sections that are downstream from this point” are important for flood disaster planning. They can be hundreds of times faster when applied as a topological query rather than a repeated spatial index search.
- It helps detect inconsistencies in the data - parcels with two seed points, or impossible pairs of feature codes sharing a common link.

Accuracy Improvement

Once the previous steps have been completed, you have a re-engineered dataset, suitable for producing better maps, or digital data products. However, it can still be improved. The framework data of many mapping organisations is based on digitising of existing mapping, which in turn is based on datums and surveys that date back a hundred years or more. Increasingly modern applications and the new generation of Location-Based Services (LBS) make use of accurate positioning information, available from GPS or satellite imagery. When GPS positions are combined with old mapping, then major discrepancies are highlighted.

The solution is to re-engineer the framework data again, to relate it to modern datums such as WGS84, and to make it match GPS positions. To do this requires picking up all the features, moving them a few metres or tens of metres, and putting them back down again. This is much easier to do if you have previously re-engineered your data into a continuum of meaningful objects, rather than worrying about the effects of sheet edges. The information for how much to move things can come form a range of sources - modern aerial photography is a favoured principal source, particularly when combined with accurate GPS-based georegistration.

Accuracy improvement does have its negative factors also to consider. What about all your users who have digitised their own information on top of your base mapping? When your features move, what happens to theirs? Displacement vectors are a useful starting point, but topology snapping is also a useful tool in re-establishing spatial relationships between users data and your framework data.

Conflation

So, finally you have a cleaned, structured, continuous dataset of meaningful objects, which has been positionally improved. Is the task now finished? It is probably the case that you have other datasets that contain useful information, which could add value to the base data if combined. An example is the OSGB project called ITN (Integrated Transportation Network). This involves taking a road centrelines dataset called OSCAR, which had poor positional accuracy and was in places incomplete or out of date. It did however have good attribution (road numbers, one-way streets). The ITN project combined it with OS MasterMap, which had better positional accuracy and was more up-to-date.

The task of combining two datasets to create something that is of greater use than either of the originals is known as conflation. It is a hard task, because it is rare that there is a simple one-to-one relationship between objects in the two datasets. The task of matching each object from one dataset with one or more objects in the other requires excellent spatial intelligence as well as fuzzy analysis of attributes. The toolset needed for this task is related to that for single dataset re-engineering, but tailored to meet demanding requirements. OSGB used Laser-Scan LAMPS2 as the engine on which to develop their solution for the ITN project.

Cross-Enterprise storage

A global transition is under way in that location data (including mapping) is becoming centralised, and being handled by mainstream information technology rather than specialised GIS software. Increasingly, spatial data is being held in large databases (typically Oracle 9i), which act as a central warehouse to the enterprise. However, deriving multiple products at different scales and specifications by generalisation has intensive requirements for spatial access to data. These requirements put great demands on the pure relational model, which although good for generic queries and for short transactions, is not optimised for the local lookups and reactive consequences needed for good generalisation.

Therefore, the preferred flowline approach is to extract from the warehouse (or the set of separate files if no warehouse has been built yet) the subset of data needed for a particular product, and load it directly into an object subsystem that is optimised for localised access and rapid modification. At the end of the data re-engineering or generalisation processing, the results are transferred back to the warehouse.

Centralising storage in this way avoids having many different datastores, each with their own idiosyncrasies and costs. Access is coordinated, control is much more enforceable, and system management such as archiving is easier and more reliable. If you have a major fire, then it is much more likely that you have a backup of your main Oracle database off-site, than you have up-to-date copies of all the individual files of the legacy production systems.

Relational datastores are advancing in capabilities, both by the manufacturers such as Oracle with their forthcoming 10i release, and by third-parties. Laser-Scan has its Radius family of products, typified by the Radius Topology product, which provides explicit topology intelligence directly in an Oracle 9i database.

Maintenance

One vital point to note, is the importance of a coherent maintenance strategy for data that has been re-engineered. It is no use investing in re-engineering your master data into a continuous, clean, structured object form, and then letting it degrade back to ragged dirty spaghetti because your legacy editing tools cannot handle the demands of the new form.

Another important aspect of maintenance is to establish and enforce a rigorous lifecycle of the meaningful objects and their Unique IDs. The Ordnance Survey has defined such a lifecycle that lays down the rules for when a feature remains the same object, and when it goes away to be replaced by new objects. An example might be that, if an area is split in two, then the larger half retains the original object ID. If it is split in three, and no part is more than 50% of the original size, then all the parts have new IDs.

Managing the transition

As with all large projects, there are pros and cons to the big bang approach (as fast as possible) or the piecewise (many careful steps). Experience indicates that it should be done in as few steps as possible, in as short a time as possible, commensurate with economic pressures, and the need to keep production going through the transition period. This is because the costs and dangers of keeping parallel and diverging sets of data in use for a long period outweighs the risks of the fast approach.

Generalisation

Generalisation can be considered to be a special kind of data re-engineering, where the target product is at a smaller scale or reduced resolution. It is considered further in the next two sections.

KMS Denmark

External Pressures and Resultant Strategy

KMS, like many NMOs, has pressures from its Government and from its users. The Government wants the organisation to be leaner and production to be more efficient. KMS are not to produce paper maps for the consumer market any more, and the only paper maps produced at present are for the military. Focus has been shifted towards digital products with a wide range of applications. The users - partners, private businesses and public organisations - have a need for new digital products, in this case small scale topographic maps that are consistent with the existing digital topographic database (TOP10DK).

Until 2000, the production of topographic maps at KMS was in the form of paper maps at scales of 1:25,000, 1:50,000, 1:100,000, 1:200,000 and 1:500,000. The production was based on photogrammetric surveys and manual updates of each scale individually.

Alongside these paper products, KMS had established TOP10DK, a digital topographic database at the scale of 1:10,000, which is widely used in local government and state organisations for various planning purposes. Designed for GIS purposes, it is a consistent, continuous, attributed, and topological ‘clean’ dataset. Unlike the datasets of many NMOs that were captured from existing paper mapping, TOP10DK was captured directly from photogrammetry, and hence does not require the kind of re-engineering described in the previous section.

In 2000, KMS formed a new strategy concerning its topographic products, aimed at deriving generalised databases from TOP10DK. The aim of the strategy was:

- To base future topographic production on TOP10DK as the basic scale starting point.
- To derive generalised databases at a range of scales upon which KMS and others could establish product and services.
- To build the production mechanisms on a standardised database implementation with open interfaces, so that various platforms could use the data.

The reasons for the chosen strategy are many; data will be updated synchronously and from the same source instead of individual and asynchronous update. It is expected that the production will be considerably more effective and that the digital output of the derived database will have a much wider scope than the traditional paper maps. Finally, the data model will be consistent for all scales.

KMS's strategy aims at establishing derived databases at scales of 1:25,000, 1:50,000, 1:100,000, 1:200,000 and 1:500,000. These scales are considered to serve the current and future needs of KMS for production of conventional topographic products, for electronic distribution on KMS map servers, and for the production of further products by partners.

KMS are now in the process of making specifications and data models for these scales. In particular, at the scale of 1:50,000, KMS has started the production of a database, which is being used in the production of paper maps for the Ministry of Defence - a major customer of KMS.

Specific Requirements and Actions

In 1999, KMS put out a Request For Tender with paid benchmark, to assess what tools commercial suppliers could provide to help meet its emerging strategy. The shortlist included ESRI, Intergraph, and Laser-Scan. The benchmark used KMS data and a set of 60 generalisation scenarios for different feature classes. The results of the tender and benchmark showed that no current single commercial product could provide off the shelf what was needed by KMS.

KMS chose to work with Laser-Scan because its plans for use of the Agent approach was seen as important, as it had the embedded intelligence to handle the context of features. In addition, the active object processing tools were able to tackle many of the scenarios of the benchmark. KMS therefore took the Laser-Scan Agent project prototype for a pilot system trial. It funded some enhancements by Laser-Scan, notably new process methods, constraints, and algorithms, to handle special situations occurring in KMS data. In addition, KMS did some development itself.

Implementation of Generalisation Flowline

The flowline that has evolved for the production of 1:50,000 scale military paper maps is as follows:

- Data are extracted from the Oracle database.
- Bridges are located, slopes are collected and building heights are determined in Laser-Scan software.
- Traffic and areas are generalised, buildings are selected and line weeding are performed in Laser-Scan.
- First manual editing is performed in GeoMedia.
- Buildings are generalised in Laser-Scan (Agent).
- Second manual editing is performed in GeoMedia.
- Bridges are generalised, road ends, power lines and symbols are reclassified in Laser-Scan software.
- Various cartographic processes are carried out in an Intergraph environment.
- Rasterisation of data is carried out in Intergraph Iplot.
- Data are prepared for plot in Intergraph Map Publisher.

It is planned that data in future should be handled in Oracle through all processes. Currently data are converted from and to Laser-Scan's Gothic object environment using the Feature Manipulation Engine (FME).

The illustrations below show rural and urban data during generalisation (Copyright KMS). Figure 5 shows Agent-based generalisation of rural farms, involving substantial displacement of buildings from the exaggerated roads, while retaining the L-shaped inherent shape of the farm during aggregation.

Figure 6 shows the use of object intelligence to identify motorway slip roads through their topological connectivity. During generalisation, the motorways have to be greatly exaggerated in width, which would obliterate the slip roads (the on and off ramps), if they were kept in their true positions. Therefore, the non-motorway ends of the slip roads have to be displaced away from the junction, while retaining connectivity. However, the slip roads are not explicitly coded in the KMS data, so a true/false object value method (motorway slip road) has been defined. This uses the in-built topology information of Gothic to dynamically identify these particular motorway lane sections for special treatment.

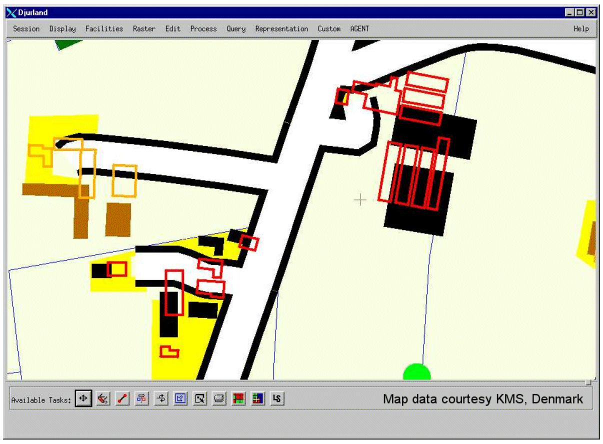
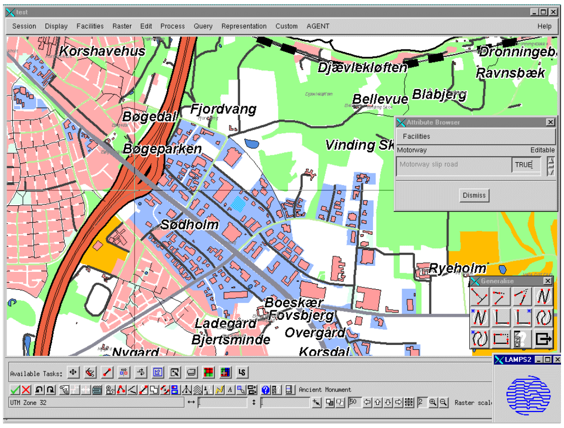


Fig. 5 - Agent-based generalisation of farms



Motorway slip road TRUE

Fig. 6 - Object views are used for intelligent selection

Achievements

So far, approximately 10 out of 109 sheets have been produced by automatic generalisation. The current production is still under development, and as the workflow shows, is not completely automatic yet. Most generalisation activity is based on scripts developed by KMS, and the development is constantly being improved, based on the experiences made in the production. KMS expect to exchange flowline scripts and example parameter sets with their partners in the Magnet consortium, to mutual benefit.

Currently, the gains in effectiveness have been modest, compared to the updating of the traditional paper maps. However, it must be remembered that the output of the new production flowline is a new topographic product, in contrast with the traditional production, which was merely updating an old product. The new output is a digital database with a variety of attributes, and hence a variety of uses. It is expected that the production will be made more effective as the development of generalisation tools and algorithms proceed.

Forward Plans

The plans of KMS are firstly to implement the new *Clarity* product (see section 5 below) in place of the Agent prototype, but also to improve the over all workflow in order to make generalisation and production fully automatic. It is the plan to generate generalised databases at the above-mentioned scales, synchronised with the update of TOP10DK (which is every fifth year) and to update selected themes (mainly buildings, roads) in a shorter cycle - possibly several times a year.

It is further the long-term plan to implement the developed methods in the production of the small-scale maps, at the above-mentioned scales.

Finally, it is an aim to strengthen co-operation with the NMOs involved in *Clarity,* and with other institutions that take part in the generalisation network.

Generalisation, MAGNET, and *Clarity*

Rationale

Although generalisation had been a research topic for many years, and various GIS products had introduced simple generalisation operators, the NMOs realised that the tools they needed for their future production generalisation flowlines were not commercially available. Laser-Scan, through its involvement in the Agent project had a promising framework for generalisation, and called together interested parties to find a way forward.

The MAGNET Project

A workshop for national and regional mapping organisations interested in generalisation was held at Laser-Scan, Cambridge, UK in May 2002. Current progress was reported and much valuable feedback presented. In addition, there was a well-informed discussion about future requirements, product direction and the means of achieving business objectives.

It was agreed that generalisation was a key interest of the user community and that the current technology and platform offered a very promising way ahead. However it is a relatively specialist area, not representing a mass market. Consequently, a partnership approach was decided upon, to spread the investment load and to provide a platform in which users could influence product directions and contribute expertise and research. Hence the emergence of the proposal for a Mapping Agencies Generalisation NETwork (MAGNET).

MAGNET & *Clarity*

MAGNET is made up of a commitment by Laser-Scan to a new generation of product for cartographic and geographic data generalisation, with matching commitments from consortium members. The new generation product is named *Clarity*. It uses a re-write of the current AGENT core, consolidating the experience gained from the AGENT project research and from the early adopters. The gains from this development include:

- A Java interface. The historic Laser-Scan Gothic O-O environment uses a proprietary scripting and database programming language. This has been replaced by Java, both at the user interface level and at the data modelling level. *Clarity* uses this, and so needs only widely available software engineering skills.
- A new operator interface for defining measures, parameters and process sequencing. This is supported by improved tools for tracking and tuning algorithms.
- An extensible framework in which users can readily add new algorithms as they become available.

Forward Plans

The development roadmap for *Clarity* also includes extending the range of data types that can be generalised. The current system concentrates on the data types that cause most work in manual flowlines, notably buildings and roads. It is anticipated that refinements of existing algorithms and new algorithms will extend the range to include rivers and drainage patterns, contours, areas and navigational charts. Extensions are also envisaged to other scale ranges.

Subsequent phases are proposed to address incremental generalisation and 3D data. An automated text placement capability (*ClearText*) will be integrated with the solution. In the medium term, migration of underlying Object-Oriented capabilities to an industry standard database platform (Oracle) is envisaged.

Conclusions

- Many NMOs have substantial assets in their existing digital cartographic data, which are under-used and under-valued. In many cases, this is because the form and structure of the data matches that of the paper products from which they were captured.
- Re-engineering and generalising this legacy data to provide new products is very beneficial, but existing feature-based GIS software has not been up to the task.
- This paper presents a taxonomy of spatial data re-engineering and of generalisation, with examples from major NMOs, showing the benefits to be gained from data re-engineering and generalisation.
- New spatial processing software has now appeared, based on intelligent spatial objects and software agents. At the same time, database technologies have matured to enable interoperability and best-of-breed flowline solutions. These for the first time provide the necessary tools and infrastructure to make data re-engineering and generalisation tasks economically possible.

Acknowledgements

Some tables and text used in this paper are updated from previous Laser-Scan publications, notably references [Woodsford 2003] and [Hardy et al 2003].

References

Birth, K, 2003, “ATKIS-Projekt Modell- und kartographische Generalisierung, und die Entwicklung geht weiter”, Kartographische Nachrichten, Magazine 3/2003, page 119-126, Kirschbaum Verlag GmbH, http://www.kirschbaum.de

Curtis E. & Painter I., 2002, “Meeting the master challenge”, Ordnance Survey GB Press Release, http://www.ordsvy.gov.uk/media/newsreleases/2001/september/master_challenge.htm

Hardy P.G. 2000, “Multi-Scale Database Generalisation For Topographic Mapping, Hydrography And Web-Mapping, Using Active Object Techniques”, IAPRS, Vol. XXXIII, ISPRS Amsterdam, Netherlands, July 2000, or online at http://www.laser-scan.com/papers/isprs2000pgh_1436.pdf

Hardy P.G., 2001, “Active Objects And Dynamic Topology For Spatial Data Re-Engineering And Rich Data Modelling”, Dagstuhl Seminar 01191 on Computational Cartography and Spatial Modelling, May 2001, or online at http://www.laser-scan.com/papers/dagstuhl2001pgh.pdf

Hardy P.G. and Haire, K., 2000, “Generalisation, Web Mapping and Data Delivery over the Internet”, ICA workshop, November 2000, Barcelona, Spain, or online at http://www.Laser-Scan.com/papers/barcelona2000pgh.pdf

Hardy, Hayles & Revell, 2003, “*Clarity* - A New Environment For Generalisation, Using Agents, Java, XML And Topology” ICA Workshop, Paris, http://www.geo.unizh.ch/ICA/docs/paris2003/papers/hardy_et_al_v1.pdf

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W.A., and Weibel, R., 1999, “The Application of Agents in Automated Map Generalisation”, 19th ICA/ACI Conference, Ottawa, 160-169.

Ormsby, D., and Mackaness, W. A., 1999, “The Development of Phenomenonological Generalisation within an Object Oriented Paradigm”: Cartography and Geographical Information Systems, v. 26, p. 70-80.

Ruas A., 2000, “Project AGENT: Overview and Results of a European R&D Project in Map Generalisation”, ICA Workshop, November 2000, Barcelona, Spain.

Sheehan, R, “Generalisation at the Danish Mapping Agency - A Solution Based on the AGENT Research Project and Laser-Scan Object-oriented Technology” GIM International July 2001, Volume 15, Number 7.

Van Oosterom, Stoter, Quak & Zlatanova, 2002, “The Balance Between Geometry and Topology”, Spatial Data Handling 2002, Ottawa, Canada, or online at http://www.radius.laser-scan.com/pdf/chpt16-147.pdf

Woodsford P.A., 2003, “MAGNET - Mapping Agencies Generalisation NETwork”, GeoInformatics magazine April/May 2003, http://www.geoinformatics.com/issueonline/issues/2003/aprilmei_2003/html/article3.html