An Enterprise System for Generalisation

N. Regnauld[1], S. Lessware[2], P. Martin[2]

[1]Ordnance Survey, Adanac Drive, SOUTHAMPTON, United Kingdom, SO16 0AS
Email: Nicolas.Regnauld@ordnancesurvey.co.uk

[2]1Spatial Group Limited; Tennyson House; Cambridge Business Park; Cambridge; CB4 0WZ; United Kingdom
Email: Seb.Lessware@1spatial.com
Paul. Martin@1spatial.com

1. Introduction

Ordnance Survey has been investigating the derivation of products from a single master database for a number of years. After Research projects successfully demonstrated the viability of automated solutions, an investment programme to build the capabilities to efficiently derive products from the large scale base data has been initiated. This paper describes the key elements of this programme, the technology used to develop the automated generalisation capabilities, and the migration process to move from a Research prototype to a maintainable, efficient production system.

Section 2 of this paper describes the requirements and the architecture chosen for the system. A competitive tender was let for the supply of generalisation tools for the Ordnance Survey's Multi-Resolution Data Programme. Section 3 describes these platforms, and how recent work with Ordnance Survey has resulted in development of a hybrid system in which Radius Clarity's Agent technology has been integrated into Radius Studio. This means that Radius Clarity's approach to automated cartographic generalisation is now available for deployment within enterprise service-based architectures and Radius Studio can therefore now be used to perform end-to-end automated map generalisation.

Section 4 uses the experience of OS VectorMap District [Revell et al 2011] to discuss the challenges involved with integrating research prototypes into a fully-fledged production environment. Section 5 discusses the longer term vision.

2- The Multi-Resolution Data Programme (MRDP)

Ordnance Survey has recently started a large programme of work to change the way products are created. This was triggered by a clear change in what Ordnance Survey's customers want. Feedback received from customers shows that the current Ordnance Survey products no longer fully meet their needs. This is highlighted by the fact that customers are far more informed so can, and will, now articulate their map data requirements, as opposed to choosing from a fixed set list of products which was the case in the past.

Ordnance Survey needs the flexibility to adapt its products to provide the best solution for each customer's unique set of requirements. It is difficult to do this using the current derived products (defined as 1:10,000 scale to 1:1,000,000 scale), so flexibility needs to be built into any future product portfolio. The creation and management of map data content that can be used to create product variations, individually styled to meet market needs, is a first step towards meeting this need. Achieving this and providing the capability to produce new products is the aim of MRDP. Efficiency is also very important, as customers and stakeholders of Ordnance Survey always expect more value for less money. Lastly, MRDP is also expected to deliver better consistency between products.

2.1 System Requirements

In order to deliver this increase in flexibility, efficiency and consistency, the following high level requirements have been set for the system:

- Provide reusable data components

Data conflation and data enhancement required to support the generalisation of a specific product could be useful for other products. We therefore want to store these so that they can be reused. This will bring flexibility in designing new products, by allowing the designer to select the appropriate data components and style them. Reuse of data components will also bring efficiency and contribute to improving consistency between products.

- Build a rich generic toolbox for generalising the data

Generalisation algorithms can often operate for a number of different scale reductions, and should therefore be designed so that they are easy to tune for exploitation across the range of their capabilities. This will allow the reuse of algorithms for different requirements, again providing flexibility, efficiency and consistency benefits.

- Support incremental updating

Once the first version of a product has been released, it needs to be maintained. For high quality products (like the paper map series), the product is manually edited to correct problems left after the automatic processes. This manual process is very slow and costly, so when the source data is updated, we only want to update in the product what is directly affected by the change, affecting as little as possible the other features.

2.2 Architecture chosen

The architecture designed for this system is based on a multi-resolution database. We call it multi-resolution instead of multi-representation, as the term representation may be confused with cartographic representation, while we want this database to stay as independent as possible to the representation used in specific products. However, the concept is the same, and the multi-representation database has been widely studied (Balley et al 2004). By keeping the multi-resolution database non product specific, we will have an architecture that follows the DLM/DCM (Digital Landscape/Cartographic Models) principles, first presented in (Grunreich 1985) and now widely adopted (Trevisan 2004, Bobzien et al 2007).

The schema in Figure 1 shows a simplified view of the architecture, focusing on how the database is organised. The staging database is a mirror of the maintenance database. It is kept separate to prevent the processes involved with deriving products to interfere with those involved with updating the main database. This database is then used as a source for the data enhancement processes, to make implicit information explicit (like creating urban extents, deriving networks from topographic features, etc). The results of enhancements are stored in the detail content database. This is then generalised to populate the lower resolution databases. Each content database is then used to populate specific product databases.

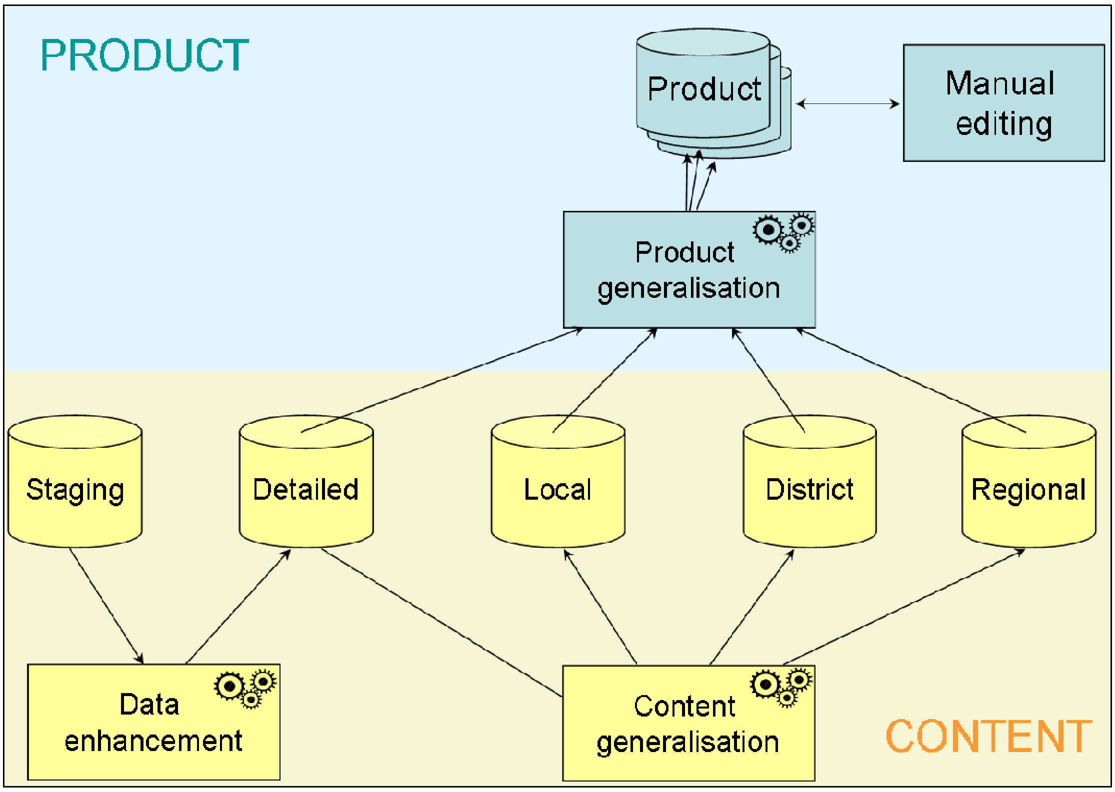

Figure 1: Simplified architecture used by MRDP

Content databases at each resolution (detailed, local, district, regional) are used as a single source for different products at similar resolution. The intention is to reuse the same content data for efficiency and consistency purposes.

All databases will be updated incrementally. The change data will arrive in the staging content database, and be generalised to update all the lower resolution content databases. The regime of update for the products is product dependant. The change will be processed in chunks that we call clusters. A cluster is a set of features affected by change that need to be processed together. These are dynamically constructed, sometimes as an aggregation of smaller static clusters (like partitions formed by road lines). This aggregation ensures that the change will be encapsulated inside a cluster and not across its boundaries. There will be several types of cluster used to propagate different types of change.

This architecture shows a lot of redundancy of data storage. This is not a big issue for data that does not need to be manually edited. However, if several products at the same resolution need manual finishing, we try to minimise the data redundancy to minimise the manual effort required by, for example, doing the manual editing on a common product database before styling them differently in separate child product databases. This is only possible if the different styles have certain similarities, such as the size/width of symbols.

2.3 Technology used

A number of technologies are used by the MRDP programme to build the production systems. Some have been chosen because they are already available and widely used in Ordnance Survey. This is the case for Oracle being chosen as the database provider, and ArcGIS as the editing platform. However, we needed to integrate into the system a way to develop bespoke generalisation algorithms, as no existing platform provided the generalisation tools required to meet all the requirements. Ordnance Survey therefore started a competitive tender process at the end of 2009 to choose a platform for developing its own automatic derivation tools. The contract was awarded to 1Spatial in January 2011, allowing Ordnance Survey developers, in partnership with 1Spatial, to develop processes based on Radius Studio and Radius Clarity. These two platforms are presented in the next section. They are used to perform the data enhancement, content generalisation and product generalisation tasks (see architecture in Figure 1).

3- Platforms provided by 1Spatial

3.1- About 1Spatial

1Spatial Group Ltd is a geospatial systems and services provider that has been delivering operational data management solutions for 40 years. 1Spatial's focus and expertise lies in the processing, transformation, management and maintenance of geospatial data with an emphasis on data integrity, consistency and generalisation. 1Spatial provides automated generalisation solutions, primarily to regional and national mapping agencies in Europe such as KMS Denmark, IGN Belgium, IGN France and Ordnance Survey of Great Britain

3.2- Radius Clarity

Radius Clarity is a stand-alone application that uses 'Agent' technology to make map objects 'self and context aware'. This allows objects to 'co-operate' to achieve acceptable automatic cartographic (product) generalised results by trying a number of strategies to achieve local feature goals and selecting the set of results that provide the 'happiest' result across the entire dataset. This approach has been shown to provide production quality results for automatic cartographic generalisation (Duchêne, 2003).

Radius Clarity provides a graphical user interface that allows the process to be configured in detail and then tested and debugged. As well as configuring the processes, additional algorithms and capabilities can be written in Java and plugged into the framework. The processing is invoked via a user interface or from the command line.

3.3- Radius Studio

Radius Studio is 1Spatial's rules engine that allows data quality rules and data processing actions to be defined and applied to spatial data. Rules are defined as a tree of spatial or non-spatial first-order logic predicates and processing actions are defined as a sequence of predicates and geometric processing operations, making the platform well suited to content (model) generalisation tasks. Radius Studio can be operated as a standalone application or can be deployed as a service within an enterprise architecture (e.g. for automated production systems).

The advantage of Radius Studio over Radius Clarity is that the processes are deployed as J2EE components over a grid of servers to provide high availability and linear scalability. The grid technology is implemented using the OpenSource (GPL) GridGain framework, allowing many processing engine nodes to be deployed across a network where each processing job can be handled in parallel by a different node. It is currently in operational use by Ordnance Survey as a key component of their 'GeoSpatial Data Management System' (GDMS). The GDMS data set contains over

500 million individual features and the system needed to be capable of being published continuously by enforcing automated data quality checks and running data publication processes.

Datasets are read from an enterprise database such as Oracle Spatial and, like at OSGB, can optionally be managed using Oracle Workspace Manger to provide long transactions during data maintenance.

Radius Studio rules and processing actions are defined using a web-based user interface and managed as XML in a rules repository. The list of processing functions can be extended with customer-specific operations that are written in Java and deployed into the processing engine.

3.4- Radius Clarity / Radius Studio integration

Although both use the same underlying object oriented technology, these two different platforms have very different strengths and capabilities: Radius Studio for content (model) generalisation and enterprise scalability and Radius Clarity for Agent-based cartographic (product) generalisation. They both make use of a flexible object model that can be used to implement most existing spatial domain models and a versioned object database with very low query latency supporting efficient spatial processing.

The integration of Radius Clarity's Agent technology with Radius Studio for the MRDP programme was achieved by building Radius Clarity's Agent libraries into the Radius Studio engine and providing new built-in functions for invoking these Agent processes to achieve cartographic generalisation. This integration means that Radius Clarity's approach to automated cartographic generalisation is now available for deployment within the scalable, enterprise service-based architectures and Radius Studio can therefore now be used to perform end-to-end automated map generalisation within National Mapping and Charing Agency production environments. For MRDP, Radius Studio's web service API is invoked automatically from a Business Process Execution Language (BPEL) workflow during the content and product stages of the architecture.

Options for extending this integration include the ability in future to not only invoke the Agent generalisation processes but to also define and debug processes within Radius Studio. Other extensions that are in development or on the roadmap include access via Esri's ArcGIS REST API to ArcGIS server or ArcGIS online data sources and using Safe Software's FME to extend the range of data formats and data sources that can be processed.

The ability to combine model and cartographic generalisation as a set of business rules within a single scalable environment opens significant possibilities for the future, not least the ability to ensure that spatial databases are maintained in a constantly product ready state for on demand mapping and digital product generation purposes.

4- OS VectorMap District – from a Research prototype to a MRDP production system

In the early stages of MRDP, a request from the government of Great Britain required Ordnance Survey to create a new product that would be made available for free to the public. As MRDP was not ready to deliver such product at the time, the new product, called OS VectorMap District, was derived almost entirely automatically from data held in Ordnance Survey databases, using processes developed in the Research department of Ordnance Survey. In order to maintain the process, it was decided that it should be integrated with MRDP, to comply with its architecture and standards.

4.1 The research prototype

The process has been described in detail in [Revell et al 2011]. Figure 2 and Figure 3 show the source data and the results obtained from the automatic process. All of the generalisation processes were developed using Radius Clarity, making heavy use of the explicit topology and also its Multi-Agent capabilities for generalising the buildings. The symbolisation was created using ArcGIS 9.3 and the automatic name placement is executed using Maplex.

Figure 2: Source data used by the automatic process, with a simple symbolisation

Figure 3: Results derived automatically and styled in a similar way to OS VectorMap District

The automatic generalisation of the whole of Great Britain for the Beta version of OS VectorMap District took 21 days of processing using three servers running in parallel on different 10km tiles. Each server had a processor Intel® Xeon® X5550 (2.67GHz), with 4GB of RAM and running 1Spatial Radius Clarity 2.7.

4.2- The migration process

In order to integrate the research processes with MRDP, the source code is currently being refactored, with the following objectives:

- Make the generalised features more reusable

Split the processes between content generalisation and product generalisation. This will provide reusable data components a various resolutions, before product dependant generalisation is applied.

- Make the tools more reusable

Many tools used for this process can be reused for deriving different products, by tweaking some parameters. These tools are often embedded into larger processes and need to be isolated in separate components, with the appropriate parameters exposed, ready for reuse.

- Make the process more maintainable

We are following an “Acceptance Test Driven Development” approach to refactor the code. It ensures that all components are behaving as planned. It will also be very valuable when tools start to be reused for other products. We will then be able to perform regression tests, to make sure that upgrades required on a tool for the needs of a particular product, do not adversely affect another product depending on the same tool.

- Make the process more efficient

The process also needs to be able to run in update mode. The initial prototype only allowed the refresh of full tiles ($10km^2$). The new system uses the notion of clusters, which are geographic regions, designed to minimise the edge matching issues when clusters are processed independently. These clusters are created by partitioning the space using the road network plus a few hidden links between the coastline and roads. When change occurs inside a cluster, the cluster is reprocessed. When change occurs at the cluster boundaries, a wider cluster is used.

- Make the process available in an enterprise system.

The last major change required by the architecture concerns the handling of the process. The research prototype used a crude system using batch files to sequence the processing of the different tiles. In the new system, processes developed in Java under Radius Clarity will be exposed in Radius Studio, which provides the appropriate interfaces for being called as a service by a BPEL workflow.

5- Conclusion and perspective

Based on the success achieved with research prototypes to derive maps automatically from large scale data, Ordnance Survey is now investing in building, in partnership with 1Spatial and other partners, an enterprise system to derive its future products.

As MRDP will deliver a library of map components (different feature types available at different levels of resolution), we have started to investigate how we can exploit this to provide products tailored for specific users. In particular we want to build a system that allows users to bring their own data, and build a map that integrates these data with the background features relevant to the context of use targeted for the map. On this research, we are collaborating with IGN France and Manchester Metropolitan University (Touya et al 2012).

In addition to reusing the data (map components) provided by MRDP, we are also looking into ways of reusing the tools. We want to do this by making them available to other researchers, through Webgen services (Neun et al 2005), more precisely the OGC WPS version of Webgen (Foersters et al. 2008). We have now successfully created a client and server that allow us to package and expose processes running on Radius Clarity as a Webgen WPS service, or call other services from the platform. While this server is not yet open to the outside world, we intend to look into the hardware and software licence implications of creating an online library of tools, for use in research contexts.

References

Balley, S.; Parent, Christine; Spaccapietra, Stefano, 2004: Modelling geographic data with multiple representations. *International Journal of Geographical Information Science*, vol. 18, num. 4, p. 329-354

Balley S and Regnauld N, 2011b, Models and standards for On-Demand mapping. In: *Proceedings of 25th International Cartographic Conference (ICC'11)*, Paris, France.

Bobzien, M., Petzold, I., Burghardt, D. 2007, Automated Derivation of a 1:300 000 topographic map from Swiss DLM VECTOR 200, in: Proceedings of 23rd International Cartographic Conference. Moscow, Russia, CD-ROM.

Duchêne, C, 2003: Automated map generalisation using communicating agents. Proceedings of 21st International Cartographic Conferenc (ICC 2003) in Durban, South Africa

Foerster T, Burghardt D, Neun M, Regnauld N, Swan J, Weibel R, 2008, Towards an interoperable web generalisation services framework – current work in progress. In: 11th ICA Workshop on Generalisation and Multiple Representation, Montpellier, France.

Grünreich, D., 1985, Computer assisted generalisation: Papers CERCO-Cartography Course. Frankfurt am Main, Institut für Angewandte Geodasie.

Neun M and Burghardt D, 2005, Web Services for an Open Generalisation Research Platform, in 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña, Spain.

Radius Studio (www.1spatial.com/software/sware.php?id=2)

Radius Clarity (www.1spatial.com/software/sware.php?id=8)

Revell, P., Regnauld, N., Bulbrooke,G., 2011, OS VectorMapTM District: Automated Generalisation, Text Placement and Conflation in Support of Making Pubic Data Public, in: Proceedings of 25rd International Cartographic Conference. Paris, France, http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/D3-Generalisation/CO-358.pdf

Touya, G., Balley, S., Duchêne, C., Jaara, K., Regnauld, N., Gould, N., 2012, Towards an Ontology of Generalisation Constraints and Spatial Relations, submitted to this workshop

Trevisan, J. 2004: From DLM to multi representation DCM - Modelling an application on buildings, ICA worshop on generalisation and multiple representations; 20-21 August 2004 – Leicester, UK.