Automated Derived Map Production
The Journey So Far

DEREK HOWLAND SAMUEL ANDREWS  OSGB
July 2018
Background

1791 Ordnance Survey GB begins mapping Britain.

2003 – 2009 Research team formed to investigate automated generalisation. Built up generalisation toolset and worked on projects investigating creating existing products automatically.

2013 – 2016 Multi-Resolution Data Programme (MRDP) commences.
Vision:
“*To improve the currency and consistency of current products and allow simpler creation of new derived products that meet changing business and customer needs.*”

2016 Generalisation Information Engine (GenIE) system in production, using a combination of COTS and bespoke (research) software. Generates 5 ‘Open Data’ national products, however some manual intervention is still required plus a minimal amount of manual editing.
• Ordnance Survey first launched OpenData in 2010.

• In March 2016 Ordnance Survey launched three new Open Data products, created automatically by the GenIE system.
Background

1791 Ordnance Survey GB begins mapping Britain.

2003 – 2009 Research team formed to investigate automated generalisation. Built up generalisation toolset and worked on projects investigating creating existing products automatically.

2013 – 2016 Multi-Resolution Data Programme (MRDP) commences. Vision:

“To improve the currency and consistency of current products and allow simpler creation of new derived products that meet changing business and customer needs.”

2016 Generalisation Information Engine (GenIE) system in production, using a combination of COTS and bespoke (research) software. Generates 5 ‘Open Data’ national products, however some manual intervention is still required plus a minimal amount of manual editing.

2016 – 2017 Generalisation Proof of Concept (POC) using pure COTS to build a solution to derive ‘Premium Products’ with 100% automation in a single end-to-end production process.

2018 Deliver Premium product Vector Map Local (VML) replacing manually maintained product
Proof of Concept
Automated Derived Product Creation
Core Principles

- **Trust the source data**
  - Improve source data if necessary
  - No manual editing of product

- **COTS tools only**
  - No bespoke development or workarounds
  - Minimize maintenance costs
  - Maintain Cartographic excellence

- **Full end-to-end automation**
  - No manual input throughout process from data load to publication
  - Business efficiency

- **Data quality ‘good enough’**
  - Understand the end-user’s needs – meet them but no more
  - Negotiate out constraining specifications
  - Acknowledge VML is a Premium product

© Ordnance Survey 2018 | Confidential
Development Principles

- Mixed team – IS Developers and Cartography Developers
- All development is on Azure VMs in the cloud
- Aim to use Commercial off the Shelf (COTS) software
ArcGIS geoprocessing tools linked together to create powerful workflows.

Parameters can be quickly modified and model re-run from correct tool.

Interactive Exploration – easy to retrieve data at any point of the workflow.
Distributed Processing

Architecture

ArcGIS Parent Node
- Python Script
- Partitioning Model

ArcGIS Child Node
- Python Script
- Initialise Model
- Run Model
- Finalise Model

Workflow Manager

Source Azure SQL

Target Azure SQL

FME

Topographic Data

Tactical Data

OS

© Ordnance Survey 2018 | Confidential
Rapid Prototyping – using OS 1/1250 core data

© Ordnance Survey 2018 | Confidential
Rapid Prototyping – 1/10,000

Detailed source data

Rapid prototyping from detailed source data

© Ordnance Survey 2018 | Confidential
Rapid Prototyping – 1/25,000

Current 1/25,000 product

Rapid prototyping from detailed source data

© Ordnance Survey 2018 | Confidential
Rapid Prototyping – 1/50,000

Current 1/50,000 product

Rapid prototyping from detailed source data
Land Parcel Polygons

- An inferred property polygon based on address seed points
- Created by iteratively merging topographic areas to building seeds
Land Parcel Polygon Statistics (National Product)

Number of VMs in Azure: **10**
Number of Input Features: **150 million**
Number of Output Features: **40 million**

One button press, one workflow: **14 hours**
Development Project
Build a fully automated solution to deliver the 1/10,000 Vector Map Local product

Project objectives
• Replace the existing 20 year old production process with fully automated system
  • Current product manually maintained by team of 20 cartographers
• Ensure ongoing revenue with significant efficiency savings
• Improve product currency (reduce from average 3 year currency to 3 months)
• Deliver a ‘framework’ system that can be re-used for other products to improve consistency
• Reduce hardware / support costs through distributed processing in the cloud
Challenges

Trust the source data

❖ Improve source data if necessary
❖ No manual editing of product

OSGB has a large-scale database of high quality from which derived products are generated, however it is not 100% perfect. During product development we found a number of different types of source data error, including:

- Duplicate features
- Attribution errors
- Currency issues
- Geometry errors
- Extent errors
- Spelling mistakes

Previously such errors would have been resolved during product maintenance by manual editing by Cartographers. But with an automated flowline source data errors impacting derived output MUST be corrected at source.
Challenges

COTS tools only (1)

- No bespoke development or workarounds
- Minimize maintenance costs
- Maintain Cartographic excellence

Our previous auto-generalisation solution used a mixture of COTS (Commercial Off The Shelf) software and bespoke software, leading to issues with:

- Integration
- Maintenance
- Re-usability

Working closely with Esri Inc in Redlands, Esri UK we have been able to quickly resolve bugs and progress enhancement requests, allowing OSGB to deliver a fully automated solution.

Being a fully configurable COTS solution will allow the re-use for new products, increasing consistency and improving ‘time to market’.
A major challenge in creating a Cartographic product automatically is **Label Placement**.

We used the Esri Maplex Label Engine.

Replicating Cartographically selected and placed labels by an automated solution is NOT easy. Extensive configuration is required.

Working with the Maplex team in Cardiff UK we resolved a number of issues with a swift turn-around in ‘hotfixes’, and agreed a number of enhancements which will further improve label placement in future products.
Challenges

- Full end-to-end automation
  - No manual input throughout process from data load to publication
  - Business efficiency

Scaling the processing through Distributed Processing

- Released December 2016 through Esri Professional Services
- Uses Parent and Child processes to partition and process large spatial datasets
- Child processes can be distributed over many processing nodes (Azure VMs), run in parallel and monitored via central interface
Challenges

Data quality 'good enough' (1)

- Understand the end-user's needs – meet them but no more
- Negotiate out constraining specifications
- Acknowledge VML is a Premium product

Product Specification

The development team built close working relationships with the Product Management Team, and through them with the end Customer

Changes to the product were kept to a minimum but included

- Product Schema
- Product Specification
- Feature Depiction

All changes were agreed. These often led to a product improved in quality, usability and consistency in addition to the greatly improved currency that automation brings

© Ordnance Survey 2018 | Confidential
Quality assurance

Vector Map Local is a ‘Premium’ product

During development quality is assured by making changes to ‘master’ code through a Continuous Integration environment and by continuous testing.

In production Quality Certification is run automatically after each stage:
- Data Load
- Geo-Processing
- Edge-matching
- Publication

Checks are made for:
- Expected file sizes
- Expected feature counts in each feature class
- Invalid data (eg invalid geometries)
- Numbers of tiles generated

Data quality ‘good enough’ (2)

❖ Understand the end-user’s needs – meet them but no more
❖ Negotiate out constraining specifications
❖ Acknowledge VML is a Premium product

Challenges

❖ Understand the end-user’s needs – meet them but no more
❖ Negotiate out constraining specifications
❖ Acknowledge VML is a Premium product

© Ordnance Survey 2018 | Confidential
## Automated Derived Map Production

### System Orchestration:

<table>
<thead>
<tr>
<th><strong>DATA LOAD</strong></th>
<th><strong>DISTRIBUTED GEO-PROCESSING</strong></th>
<th><strong>EDGE MATCH</strong></th>
<th><strong>PUBLICATION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MAIA</strong></td>
<td><strong>10 Azure VMs with 4 processing nodes on each</strong></td>
<td><strong>150 million product features edge-matched in 1 day</strong></td>
<td><strong>10591 tiles of GML product in full supply and change only update</strong></td>
</tr>
<tr>
<td><strong>TACTICAL</strong></td>
<td><strong>Detailed content database</strong></td>
<td><strong>Generalised product database</strong></td>
<td><strong>10591 tiles of colour, backdrop and b/w raster product</strong></td>
</tr>
<tr>
<td><strong>SQL</strong></td>
<td><strong>Python</strong></td>
<td><strong>SQL</strong></td>
<td></td>
</tr>
</tbody>
</table>

- **450 million features loaded from MAIA and tactical data**
- **1838 partitions**
- **2 days**
- **2 ½ days**
- **1 day**
- **4 days**

© Ordnance Survey 2018 | Confidential
One button press will:

- Create Azure Virtual Machines & scale up Databases
- Deploy ESRI Workflow Manager config
- Extract source data
- Run parent job – run parent models, split UK into partitions and create child jobs
- Run child jobs – run child generalisation models and write features to product database
- Edge-match features across partitions
- Publish - UK national set of 3 raster products plus GML with Change-Only-Update
- Provide quality assurance throughout process
- Tear down VMs & scale down databases
In 10 Days the full product set is refreshed across Great Britain with current data.

**RASTER PRODUCT**
- 10,561 tiles Full Colour
- 10,561 tiles Backdrop Colour
- 10,561 tiles Black and white

**VECTOR PRODUCT**
- 10,561 tiles GML3 format
- Change Only Update

Fully edge-matched
Consistent depiction
Quality assured

© Ordnance Survey 2018 | Confidential
Product benefits - improved quality

- Increased currency of data: reduced production time and increased speed to market providing even more value to our customers
- Improved production and product quality through data processing: partitioning, national datasets, edge-matching and creating change-only updates.
- Increased consistency of output data making it more suitable for both contextual and analytical products. Increased content consistency with other products

Business benefits – efficiency savings

- Reduce hardware / support costs through use of COTS software and distributed processing in the cloud
- Remove manual input, freeing up production resource to work on other production flowlines
- Use of COTS software removes the cost of developing and maintaining an in-house bespoke capability
Derek Howland
Principal Production Consultant - Automated Generalisation
derek.howland@os.uk +44 (0) 2380 055225