Challenging traditional business processes in the approach to automated product creation
In the beginning...

1791
The Board of Ordnance began the task of mapping the entire British Isles

1801
First map of Kent to help defend the south coast from a Napoleonic invasion

1914
Developed sound ranging and aerial photography to identify enemy position

2012
Detailed mapping of the Olympics in London to protect from attack
Background

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.
Core Principles

Trust the source data

Improve source data – only correct if necessary

COTS tools only

No bespoke development or workarounds

Full end-to-end automation

No manual editing of product

Data quality ‘good enough’

Understand the end-user’s needs – meet them but no more. Negotiate out constraining specifications
Product Specification

- Has constrained previous automated generalisation. This has resulted in a level of:
  - Bespoke development
  - Manual editing

- Where an automated COTS output is different to current spec:
  - Challenge current specifications
  - Can ‘different’ be good?
  - What is ‘good enough’?
  - Where is the customer value?

- Analysis of past generalisation attempts:
  - Generalisation at resolutions 10K, 25K, 50K
  - Speeding up generalisation gaps
  - Use of tactical data
  - Excessive validation

- Use of tactical data
- Excessive validation
- Specifications and different issues:
  - What is ‘good enough’?
  - Where is the customer value?

Generalisation Tools

- Analysis of Esri toolset to create output at a number of resolutions: 10K, 25K, 50K
- Evaluation of usability / skillset required
- Test NFRs – Scalability, Performance

Source Data

Content/Product

Product Specification

Generalisation Tools

Product Triangle
Overview of Progress

Oct 2015
- OS Visit to Esri Inc Redlands
- Initial learning and joint prototyping
- Models built for data output comparison with current system

June 2016
- OS team starts work on Generalisation Proof of Concept
- Improving Esri models
- Address risk areas
- Initial models for zoom-stack (1:10K, 1:25K, 1:50K)

Sept 2016
- VectorMap Local (VML) prototype developed
- Rapid progress

Dec 2016
- Team diverted to Land Parcel Product

May 2017
- Land Parcel Product delivered
- WMX / Azure based distributed processing proved
- Provisional business case for VectorMap Local (VML) approved

Dec 2017
- Deliver fully automated end-to-end production system for 3 monthly refresh of VectorMap Local (VML) - premium product
POC

Proof of Concept
POC
Roads
Grade Separation

- Grade level attribute stored at start and end of links
Grade Separation

- Grade level attribute stored at start and end of links
Dual Carriageway Collapse
POC

Promontories
Foreshore
Manmade Surface
Tidal Boundary (blue line)
Fill Promontories
No Skin of the Earth
Source data
Marine Water
Foreshore
Foreshore

Tidal Boundary (blue line)

Marine Water
POC
Buildings
Objective is to create district-like data

**Black** = main COTS tool used
Buffer roads
Convert using Feature to Line tool
Random Points to generate the point feature class
Split Line at Point to get consistent, individual lines
Select by Location any buffer which intersects a building
Copy Features to continue with only selected buffers
Select by Location the split lines which are entirely within the buffers.
Buffer with dissolve and flat end
Buffer roads again
Use buffered road to **Erase** overlapping building polygons
Other process to bring-in buildings away from the roads

Eliminate Polygon Part
Aggregate
Dissolve
Vector Map Local

• Premium Product
• 1:10,000 Scale
• National Dataset
• Used for backdrop/analysis
Case Study

Land Parcel Polygons
Land Parcel Polygons

- An inferred property polygon based on address seed points
- Created by iteratively merging topographic areas to building seeds
Partitioning

Inputs:
- Road links
- Tidal water links
- Extent of realm

Size of partition based on feature density.
Coastal partitions
Road based partitions
Extent of the Realm
Tidal water links
Coastal partitions
Partition edge conflicts
Testing

- Key model tests implemented
- Engineered test data
  - Input source data
  - Expected output data
- Use PyTest tests to:
  - Run source data through models
  - Compare actual and expected features
Land Parcel Polygon Statistics

10 Number of machines used
3 of 4 Number of cores used
150 million Number of input features
One button press will:

- Create Azure VMs & scale up databases
- Deploy models
- Load workflow config
- Run parent job – creates partitions and child jobs
- Create second set of conflict child jobs
- Publish (FME) to target format
- Tears down VMs & scales down databases

In:

14 hours
Summary
Core Principles

- Trust the source data
- COTS tools only
- Full end-to-end automation
- Data quality ‘good enough’

- Improve source data – only correct if necessary
- No bespoke development or workarounds
- No manual editing of product
- Understand the end-user’s needs – meet them but no more. Negotiate out constraining specifications
Trust the source data

- Initial feedback provided on source data quality
- Actively engaged in design of a new source data model

COTS tools only

- Using COTs:
  - ArcGIS Workflow Manager, Models and GP Tools
  - FME
  - Azure, VSTS
- Some workarounds— but – feedback to Esri where gaps or defects identified in toolset
Full end-to-end automation

- Successful for Land Parcel project
- Still a challenge for large volume products
  - Merging
  - Edge matching

Data quality 'good enough'

- Using techniques such as:
  - Timeboxing
  - Concentrating on customer value
- Earlier involvement in product development
- Challenging evolving product specification
- Trying to get closer to the customer
For discussion …

• Chain processing for different resolutions OR each resolution from same large scale source?
• Most effective partitioning strategy
  • Tiles or road based?
  • Same partitions for all themes?
  • Effective (and quick) edge matching
• Fewer big machines or many small machines
• ArcGIS Pro migration
Questions?

Derek Howland
Operations Domain Leader – Automated Generalisation
derek.howland@os.uk
+44 (0) 2380 055225