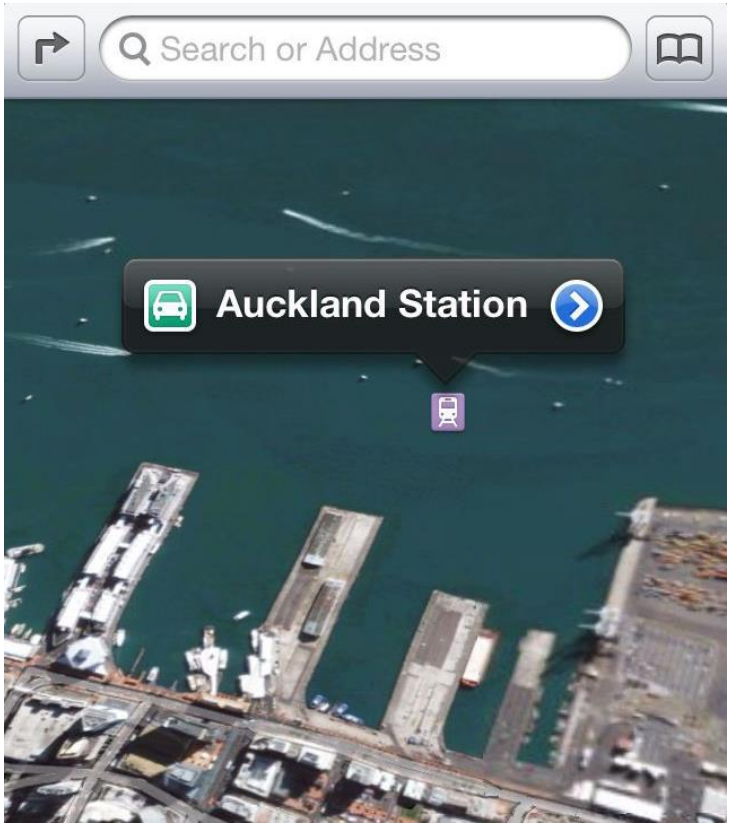
# International Workshop on Spatial Data and Map Quality

Dan Warner

7th February 2018

# How Data Quality Automation Delivers Improved Productivity:

- **Automation Drivers**
- **Automation v Spatial Data Quality**
- **Case study: United Utilities**
- **Case Study: US Census**
- **Approach to Data Quality**



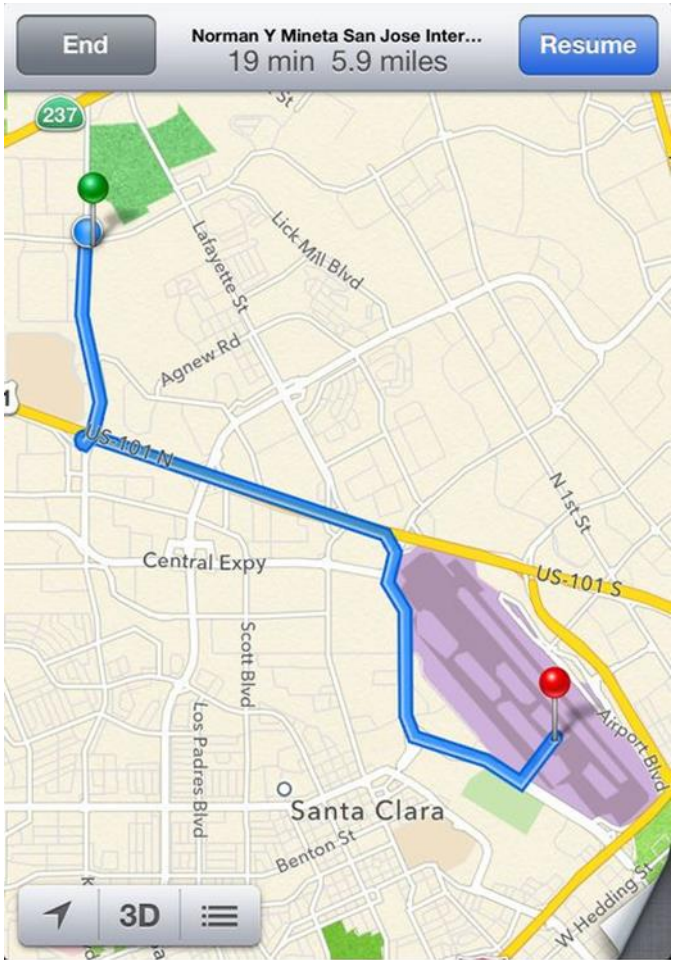
# Drivers for Automation:

- **Deliver more for less**
- **Consistent and Repeatable**
- **Measureable outcomes**
- **Real time data**
- **Complements our abilities**

# Automation v Spatial Data Quality

- **Automation Drivers:**
  - Real time data, greater currency
  - Bulk data processing
  - Derive new product, data on demand
  - More analytics, enrich data
  - Innovation in capture methods

- **Causes of Data Quality Issue:**
  - Repurposed
  - Legacy data and legacy data conversions
  - Historic conflation
  - Manual capture, varying capture standards, capture hardware



# Case Study: United Utilities

**Repurpose** ✓ **Legacy Data** ✓ **Different capture standards** ✓ **Manual data entry** ✓

- 76,000 km sewers, 42,000 km of clean water
- Inherited assets and records (data)
- Data incomplete and inconsistent
- Data maintained in multiple systems
- 400 field engineers with their copy of the network
- legislation changed responsibility for private sewers

- **Goal: Accurate view of assets**
- Reduce operational costs, and response times
- Improving customer service
- KPI's based on network length
- Estimating costs
- View of erroneous billing

# Case Study: United Utilities

**Repurpose✓ Legacy Data ✓ Different capture standards✓ Manual data entry✓**

- Data quality audit and data fix-up
- Inferred data for omissions/gaps
- Inference rules bases on:
  - Engineers knowledge
  - known network data (pipe age, pipe size, etc)
  - Topographic and Census
- Type and location of sewer inferred
- Inferred network infrastructure, calculated land parcel entry point
- Confidence rate on contextual information

# Case Study: United Utilities

**Repurpose✓ Legacy Data ✓ Different capture standards✓ Manual data entry✓**

Outcomes:

- Fully mapped private sewer stock
- Additional 6,000 km of private sewer stock identified
- Improved accurate deterioration models
- More accurate predictive maintenance plans
- Fully Connected assets inter-connect: reservoirs to pipes, drainage and sewers to treatment plants

# Case Study: US Census

**Repurpose** ✓ **Legacy Data** ✓ **Different capture standards** ✓ **Data Broker** ✓

2010 Census:

- 135 million address
- 140,000 people to validate address records
- 600,000 people to trace non-respondents
- Cost = $12 billion

- **Goal: Accurate view of the geography**
- $400 billion of government spend dependent on census
- Streamline the integration of data from 3,200 agencies
- Deliver census data to government faster
- Reduce size of field validation task
- Reduce projected cost of 2020 census ($17 billion)

# Case Study: US Census

**Repurpose** ✓ **Legacy Data** ✓ **Different capture standards** ✓ **Data Broker** ✓

- Scrubbing
- Topology Health Check
- Fix Tolerance issues
- Change detection between Partner and Census datasets
- Conflation proposal
- Commit accepted proposals

# Case Study: US Census

**Repurpose** ✓ **Legacy Data** ✓ **Different capture standards** ✓ **Data Broker** ✓

Outcomes:

- Improved data reduced field canvassing to 25% of 2010's 140,000
- Deliver census data to government faster
- Greater integration and analytics with commercial dataset (postal, roads)
- Improved data collection, base data and use of automation saving $5 billion

# Approach to Data Quality

- Data quality management system/strategy
- Automation of data validation
- Application independent
- Data validation at point of capture
- Centralised rules repository
- High performance/scalable
- Automatic correction
- Data transformation
- Pinpoint reporting
- Define what is correct
- Quality reporting, KPIs