Big Data and Road Sensors
Implementing a Big Data Statistics

Marco Puts
Statistics Netherlands and Big Data

Why a Big Data approach?

– Shorter time to publication
– Respond to current events
– Higher reliability
– More detail
– More efficient processes

Considerations:
- Infrastructure
- Competences
- Culture

“Simplicity is the ultimate sophistication.”
— Leonardo da Vinci
Road sensors

Road sensor data

- Passing vehicle counts for each minute (24/7) at about 60,000 sensors in the Netherlands
- Types of sensors:
  - Induction loop
  - Camera
  - Bluetooth
- Length categories (e.g. small, medium, long vehicles)
- Large volume: approx. 230 mln records/day
Dutch highways
Dutch highways with road sensors
Data journalism and (almost) real time statistics

Respond to current events

Within two days!
Statistics Netherlands and Big Data

Big Data Research (and development)
- Data Driven
- Case Based
- Roadmap

Bottom up Approach
- There is no Theory of Big Data yet
- Explorative Research

- Findings → Methodology
Big Data Processes
Data driven vs. output driven

Big Mess of Data → Database → Statistical Process → Stats → Big Mess of Data
How to get into production?
A Lean view on Statistical Processes

- Bring Processes to the data (T)
- Minimize human interaction (M)
- Added value of data cleaning (O/O)?
- Simple rules (O/O)
- Speeding up the processes: (H)PC (W)
- Process data in small chunks (I/W)
Examples of Big Data processing

Raw Data: 80 TB (2010 - 2014)

Transformed data: 70 GB

Microdata: 500 MB

Statistics: 6 KB

Selection + Transformation

Data Cleaning (filtering)

Estimation

Cultural Change
Process in small chunks
3V’s

BIG DATA

Volume
Variety
Velocity
Minimize Human Interaction
“traditional” approach
Minimize Human Interaction

Alternative Approach

Manually

Proces parameters

Automatic

Monitoring

Q

Q
Conclusion

– Big data is still a new phenomenon within official statistics
– No standard methodology
– Standard processes not efficient
– Big Data enables us to:
  - Publicize faster on more details
  - Respond to current events

– But we need to change!