Innovaties in het verzamelen van bedrijfsgegevens

Ger Snijkers en Sofie De Broe
Primary to Secondary to New Data sources

- **Surveys (1900)** and censuses

- **Admin sources**
  - Targeted data collection, pre-determined questions and indicators
  - Structured data, collected by government not for NSI purposes

- **Big Data**
  - Not necessarily collected for NSI purposes, data in high volume, velocity, variety
  - 2012 expanding
  - 1970/80’s expanding
  - Reduced but still in use

- **Data integration**

- **2000BC**
- **20th Century**
- **21st century**

- **Age of sampling**
- **Age of data integration**
Datafication of society

EXPLOSION OF DATA
A city of 1M people will generate 200M GB of data per day by 2020

- SMART BUILDINGS: 55M GB/day
- SMART FACTORIES: 50M GB/day
- PUBLIC SAFETY SYSTEMS: 50M GB/day
- SMART VEHICLES: 40M GB/day
- SMART AIRPLANES: 4M GB/day
- SOCIAL MEDIA + OTHER: 2M GB/day

1. Digitisation of survey data collection
   a. Features of electronic questionnaires
   b. Paradata

2. Internet
   a. Internet as data source

3. Digitisation of the business information chain
   a. Electronic Data Interchange (EDI)

4. Internet of Things (IoT):
   Digitisation of business processes
   a. New data using EDI: Sensors
1. Digitisation of survey data collection

Instead of simply using paper lay-outs

Implementation of technological features of electronic business questionnaires (internet), like:

• Automated routing
• Built-in edit checks
• Complex questionnaires (matrices)
• Imputation of t-1 data (historic data)
• Web log-in portals
• Paradata: collection of paradata

How to use these features?
- methodological
- organisational issues
1a. Innovations in surveys

Tailoring to the business context:

• Opening the black box:
  o better insights in how businesses and people within these businesses think and operate
  o Reporting process: can be complex!
Complexity of reporting process

- Many data sources, at various locations
- Many people, at various locations
- Many sub-units
- Time: when data are available, and businesses have time

Centralised accounting

Complex project organisation

Private equity company
Accounting offshore

Foreign owned multinational

NPSO, 5 juni 2018, Rotterdam
1a. Innovations in surveys

Tailoring to the business context:

• Opening the black box: better insights in how businesses and people within these businesses think and operate

• Questionnaire *communication* instead of Q *design*:
  o Questionnaire communication design
  o Usability issues / User-interface design / interaction design

• Pre-testing of questionnaires/completion process as soon as possible in the Q design process:
  o Feasibility studies
  o Usability + eye-tracking studies

• Apply ‘*influence principles*’ (Cialdini) and ‘*nudging*’ (Thaler & Sunstein) in survey communication to get response
1b. Paradata in adaptive designs

- Paradata = process data
  = data about the own data collection process
- Digitisation of process -> easy to collect and analyse
  - At CBS side:
    - Dates when questionnaires are received and processed
    - Log-in information
    - Cost and quality indicators
  - At Respondent side (audit trails):
    - When respondents open questionnaire
    - How they complete the questionnaire: completion process

Example: SBS completion process (Structural Business Survey)
(Snijkers & Morren, 2010)
Timeliness response for e-SBS2006

Time used to complete questionnaire

85% within one week (av. = 8 days)
Completion duration increases with size class

- 75.8%
- 9.5%
Timeliness response for e-SBS2006

- Dispatching Q
- Getting response: Av. after 2.8 months (85 days)

Starting Q after 77 days (2.6 months) on av.
= waiting time without activity

Days needed to complete Q: 8 days on av.
1b. Paradata

Help to:

• Get insights in these processes at both sides:
  o CBS side
  o Business side

• Tailor the survey design to the business context:
  o Questionnaire communication design
  o Business survey communication strategy to get response
  o Efficient sampling to reduce response burden
  o Adaptive designs

• Make the survey process more efficient:
  o Applying the Deming cycle: Monitoring and improving
2. The internet as data source

Examples:

• Collecting Annual reports published on the internet and analysing the reports using text mining (chart of balance data) instead of using questionnaires
• Web Scraping: prices on the internet
• Measuring the internet economy, instead of conducting a survey on e-commerce (CBS)
• Measuring MNEs using Big Data (OECD), not possible using only surveys
• Finding information about innovative businesses
• Social media data
Methodological approach

• Combination of **traditional** data sources with newly **emerging** sources and innovative ‘Big Data’ analytics, e.g.
  • Traditional sources + innovative methods: XBRL-assisted scraping of company Annual Reports overcomes previous challenge of not being machine-readable
  • Emerging sources: Legal Entity Identifier (LEI)
  • Innovative sources + innovative methods: web-scraping MNE websites paired with NLP text analytics

• Careful **validations** and alignment with official statistical concepts

• Hence: results are **timely, scalable, and can be publicly disseminated**
3. Digitisation and integration of the business information chain

**Business data in DSCs**

- **RGS/SBR(XBRL):** Machine-readable EDI
- **Upload in e-Q**

** external business reporting**

- TRADE REGISTER (CHAMBER OF COMMERCE)
- CREDIT SUPPLY (BANKS)
- CAPITAL SUPPLY (SHAREHOLDERS & INVESTORS)
- TAX COLLECTION (FISCAL AUTHORITIES)
-......
-......
-REGISTERS
-STATISTICAL PROCESSES (NSI)

**Internal business reporting**

- BOOK-KEEPING
- GENERAL LEDGER
- FINANCIAL & FISCAL REPORTING

**Business operations**

- ORDER INTAKE
- INVOICING
- SUPPLY MANAGEMENT
- HR MANAGEMENT

H2S = Human-to-System

NPSO, 5 juni 2018, Rotterdam
From Qs to S2S data communication

Human-to-Human (H2H) data communication
System-to-System (S2S) data communication
4. Internet of Things (IoT) – sensor data

Intelligent business architecture
- Smart sensing technology
- Explosive growth in data
- Real-time integrated data communication

Smart Systems Integration 2014 | 26.-27.03.2014

Waves of MEMS sensor proliferation

1st wave
2nd wave
3rd wave

1990  2000  2010  2020

Internet of Things and Services (IoTS)

Consumer Electronics

Automotive

Bosch Sensortec
Marketing | 03/22/2014 | © Bosch Sensortec GmbH 2014. All rights reserved, also regarding any disposal, exploitation, reproduction, editing, distribution, as well as in the event of applications for industrial property rights.
4. Internet of Things (IoT) – sensor data

Sensor data in businesses:

• Transportation: tracking packages
• Satellite images to estimate crop yields
• Precision or smart farming, like E.g. Smart Diary Farming


• Collecting environmental data: air quality, water quality, noise disturbance, etc.
Smart Dairy Farming

- Other data sources
- Cow specific Work instruction
- Starting point: Farmer in control
- Starting point: Cow centric thinking
- Real time models / services (at different providers)
- Sensors from different suppliers
- InfoBroker: Open platform for (sensor) data producers and consumers

Smart Dairy Farming

InfoBroker: Open platform for (sensor) data producers and consumers

Real time models / services (at different providers)

Starting point: Farmer in control

Starting point: Cow centric thinking

Sensors from different suppliers

Other data sources

Cow specific Work instruction

NPSO, 5 juni 2018, Rotterdam
# Smart Dairy Farming

## Sensor data on:

<table>
<thead>
<tr>
<th>Sensor data categories</th>
<th>Farm 1</th>
<th>Farm 2</th>
<th>Farm 3</th>
<th>Farm 4</th>
<th>Farm 5</th>
<th>Farm 6</th>
<th>Farm 7</th>
</tr>
</thead>
<tbody>
<tr>
<td># cows/calves</td>
<td>459</td>
<td>186</td>
<td>315</td>
<td>239</td>
<td>706</td>
<td>202</td>
<td>351</td>
</tr>
<tr>
<td>Behaviour</td>
<td>5x</td>
<td></td>
<td></td>
<td></td>
<td>5x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>1x</td>
<td></td>
<td></td>
<td></td>
<td>1x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>9x</td>
<td>9x</td>
<td>3x</td>
<td>6x</td>
<td>5x</td>
<td>13x</td>
<td>9x</td>
</tr>
<tr>
<td>Milk production</td>
<td>16x</td>
<td>20x</td>
<td></td>
<td>1x</td>
<td>2x</td>
<td>19x</td>
<td></td>
</tr>
<tr>
<td>Feed intake</td>
<td>24x</td>
<td>24x</td>
<td></td>
<td></td>
<td>10x</td>
<td>24x</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>10x</td>
<td>6x</td>
<td>6x</td>
<td>6x</td>
<td>7x</td>
<td>6x</td>
<td>10x</td>
</tr>
<tr>
<td>Water intake</td>
<td></td>
<td>3x</td>
<td>3x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk intake</td>
<td></td>
<td></td>
<td></td>
<td>7x</td>
<td>11x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NB1: blue numbers are animals; not all animals are monitored for SDF (e.g. 3 and 4 only calves)

NB2: the left column gives a list of “sensor data categories” at a farm

NB3: numbers in black are the sensor fields within a category (e.g. 3 fields related to waterintake)
Some scenario’s for using the InfoBroker:

- Advisor
- Farmer
- Other farmers
- Government

Coöperation

Realtime-Model (expert)

Benchmarking

InfoBroker

Static data

Real time cow-centric data

Smart Dairy Farming: InfoBroker

NPSO, 5 juni 2018, Rotterdam
Methodological and organisational Challenges

- Data generating process
- Harmonisation of data and metadata
- Stability of the (meta) data delivery
- Standardisation of IT systems
- Conceptualisation
- Unit issues
- The use of sampling
- Data quality issues: sensor data drift
- Data-ownership, privacy and security
- Costs
- Resistance
Conclusions

Technological innovations make things possible; the applied methodology and the organisational context make it work.

De dataverzamelingmethodologie blijft in ontwikkeling, ook voor bedrijfsdata!
En wat betekent dat voor de toekomst: ...
Visie bedrijfswaarneming CBS

• Reductie van vragenlijsten door:
  ○ Betere ontsluiting van secondaire gegevens
  ○ S2S data communicatie voor financiële data, sensor data:
    ▪ Individuele bedrijven (grotere)
    ▪ DSCs voor bedrijfscoöperaties
  ○ Gebruik van nieuwe data bronnen (internet)
• Ontwikkelen van een nieuwe methodologie:
  ○ Conceptualisatie (statistische concepten): nog belangrijker
• Speciale aandacht voor grote/complexe bedrijven blijft belangrijk, met tailor-made rapportage-tools (vragenlijsten)
• Surveys blijven bestaan voor afzonderlijke onderwerpen:
  ○ Een professioneel, getailored design wordt extra belangrijk om respons en goede kwaliteit data te krijgen tegen acceptabele kosten

NPSO, 5 juni 2018, Rotterdam
ARE THERE ANY QUESTIONS? FEEL FREE TO ASK ANYTHING AT ALL.