

**NPSO Lustrum jaarlijkse dag 2018** – hand-out  
Rottterdam  
5 juni 2018

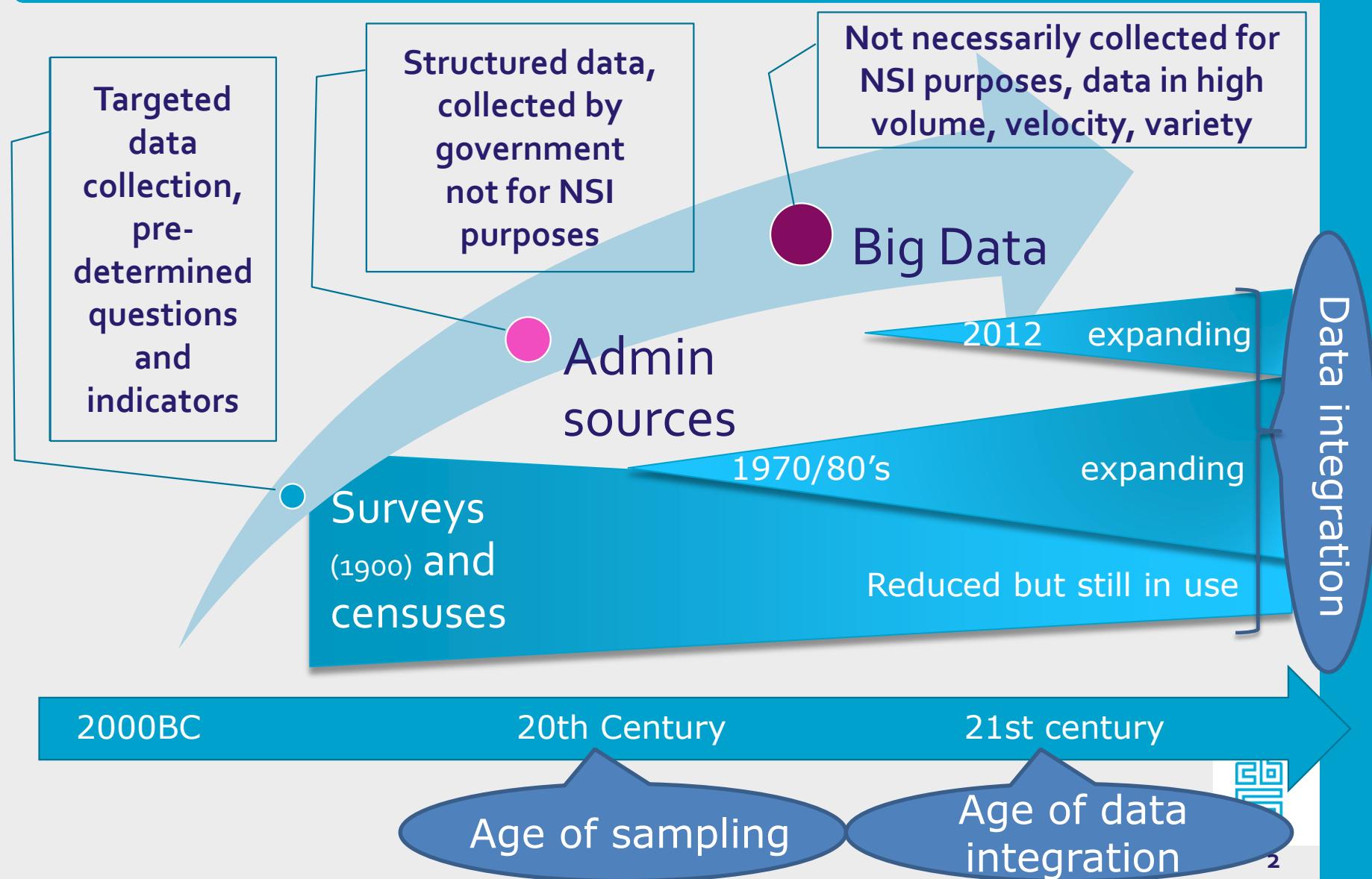
# Innovaties in het verzamelen van bedrijfsgegevens

Ger Snijkers en Sofie De Broe



Centraal Bureau  
voor de Statistiek

# Primary to Secondary to New Data sources



# Datafication of society

Ook voor  
het bedrijfsleven:

- Digitalisering en integratie van business information chain
  - Intelligent business architecture



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

EXPLOSION OF DATA

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

Source: CISCO, INTEL and Cukier and Mayer-Schoenberger "The Rise of Big Data", 2013



# Technological innovations in collection of business data

- ▶ **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

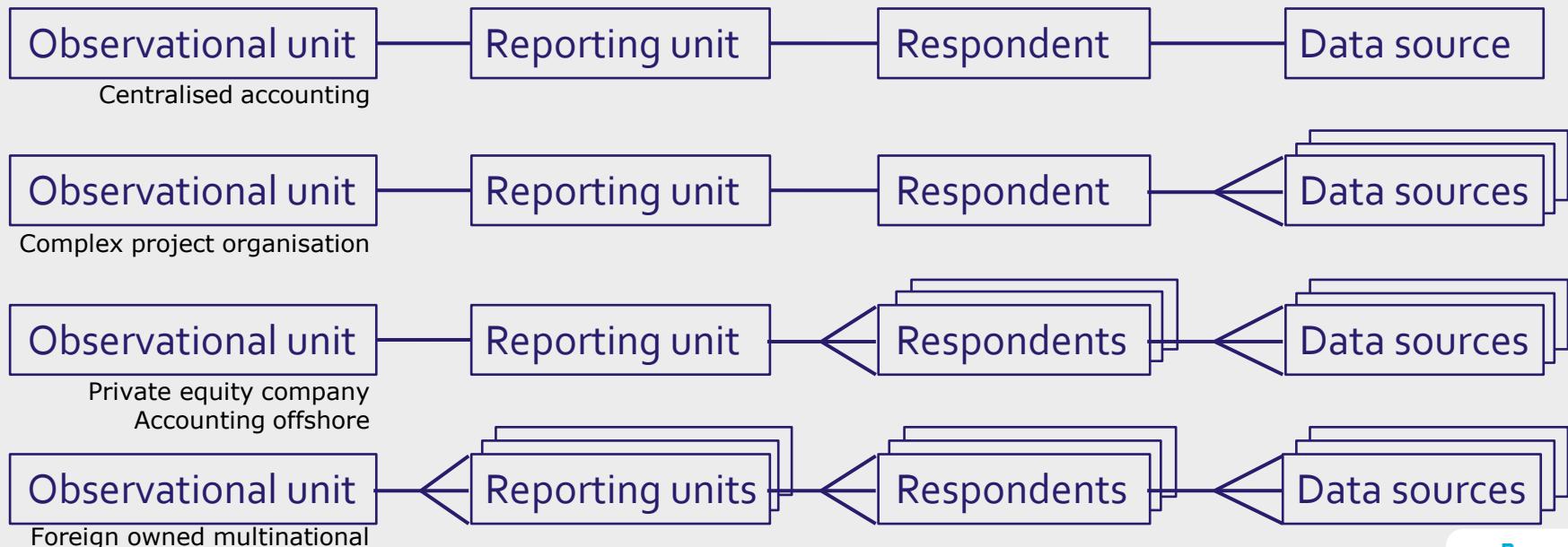
## Tailoring to the business context:

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



# 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*:
  - Questionnaire communication design
  - Usability issues / User-interface design / interaction design
- Pre-testing of questionnaires/completion process as soon as possible in the Q design process:
  - Feasibility studies
  - Usability + eye-tracking studies
- Apply '*influence principles*' (Cialdini) and '*nudging*' (Thaler & Sunstein) in survey communication to get response



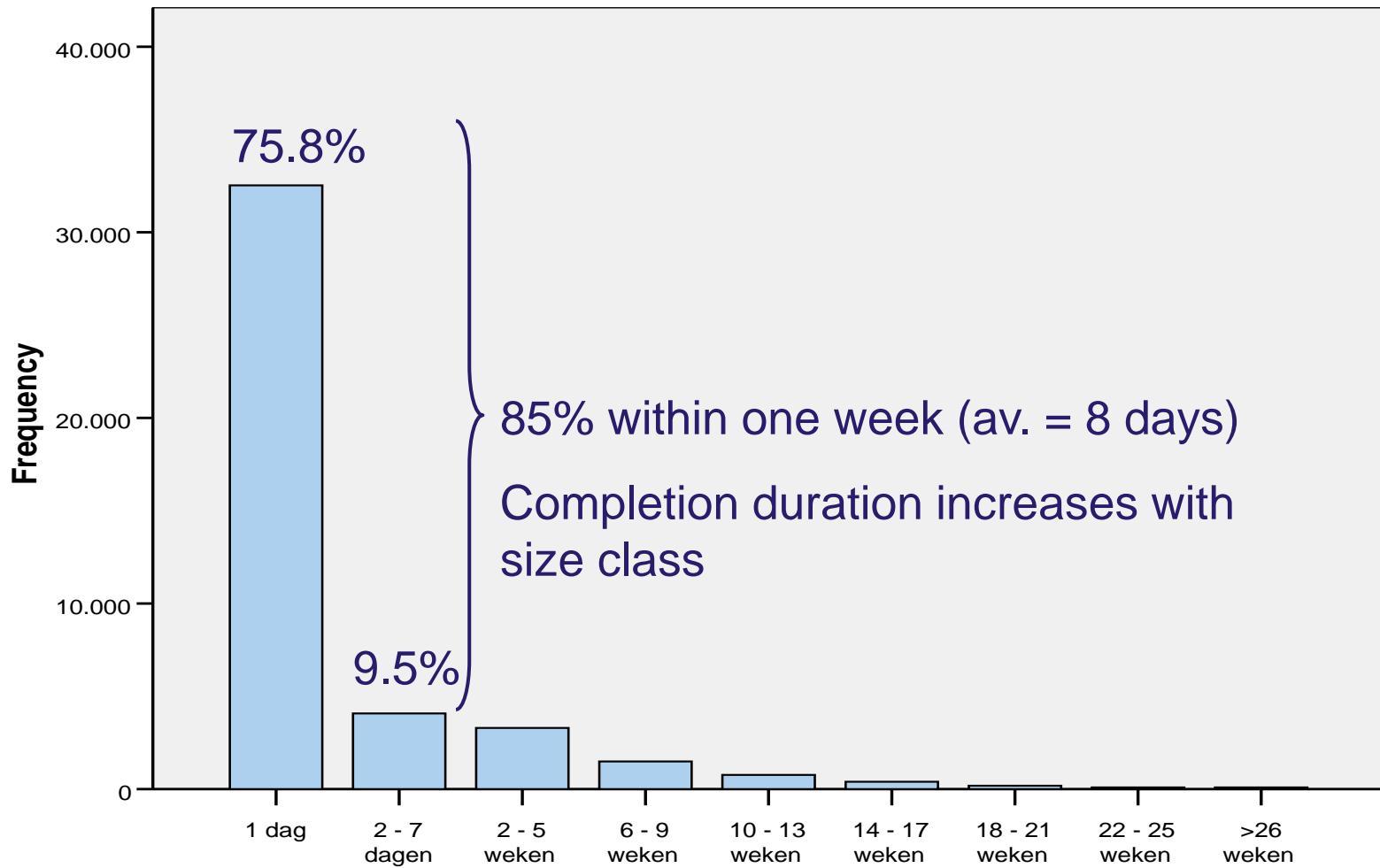
# 1b. Paradata in adaptive designs

- 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)

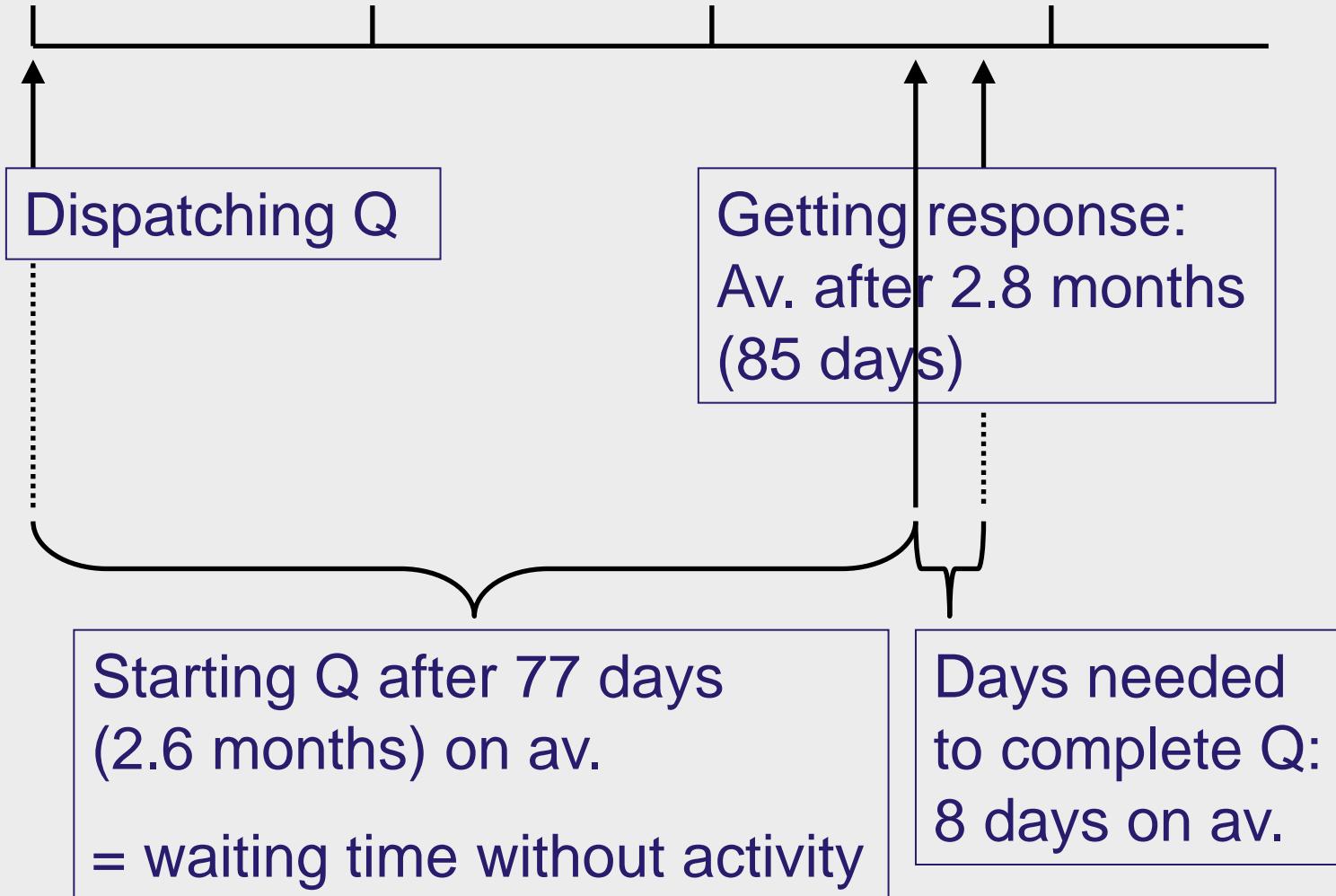


# Timeliness response for e-SBS2006

Time used to complete questionnaire



# Timeliness response for e-SBS2006



# 1b. Paradata



Help to:

- Get insights in these processes at both sides:
  - CBS side
  - Business side
- Tailor the survey design to the business context:
  - Questionnaire communication design
  - Business survey communication strategy to get response
  - Efficient sampling to reduce response burden
  - Adaptive designs
- Make the survey process more efficient:
  - 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



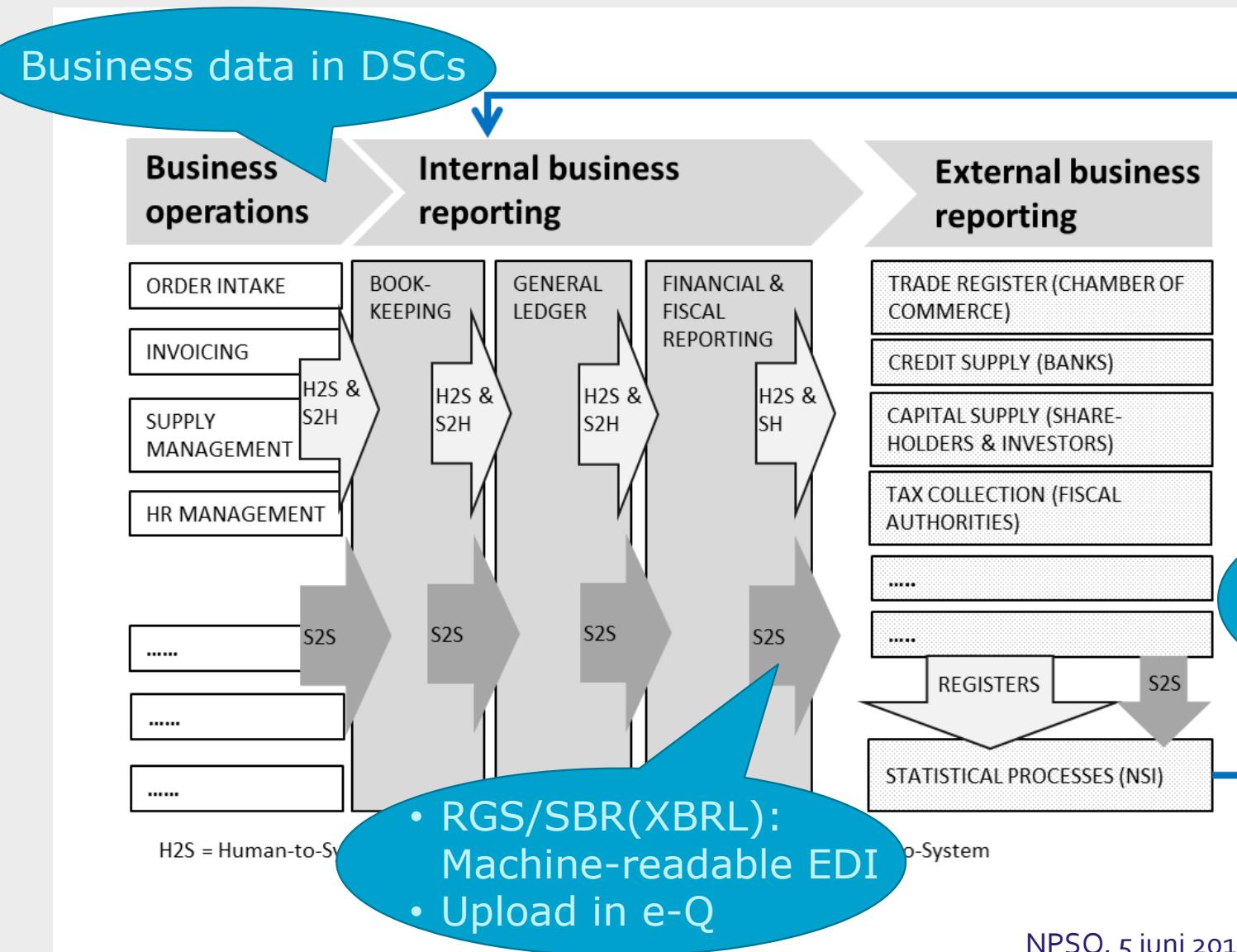


# Methodological approach

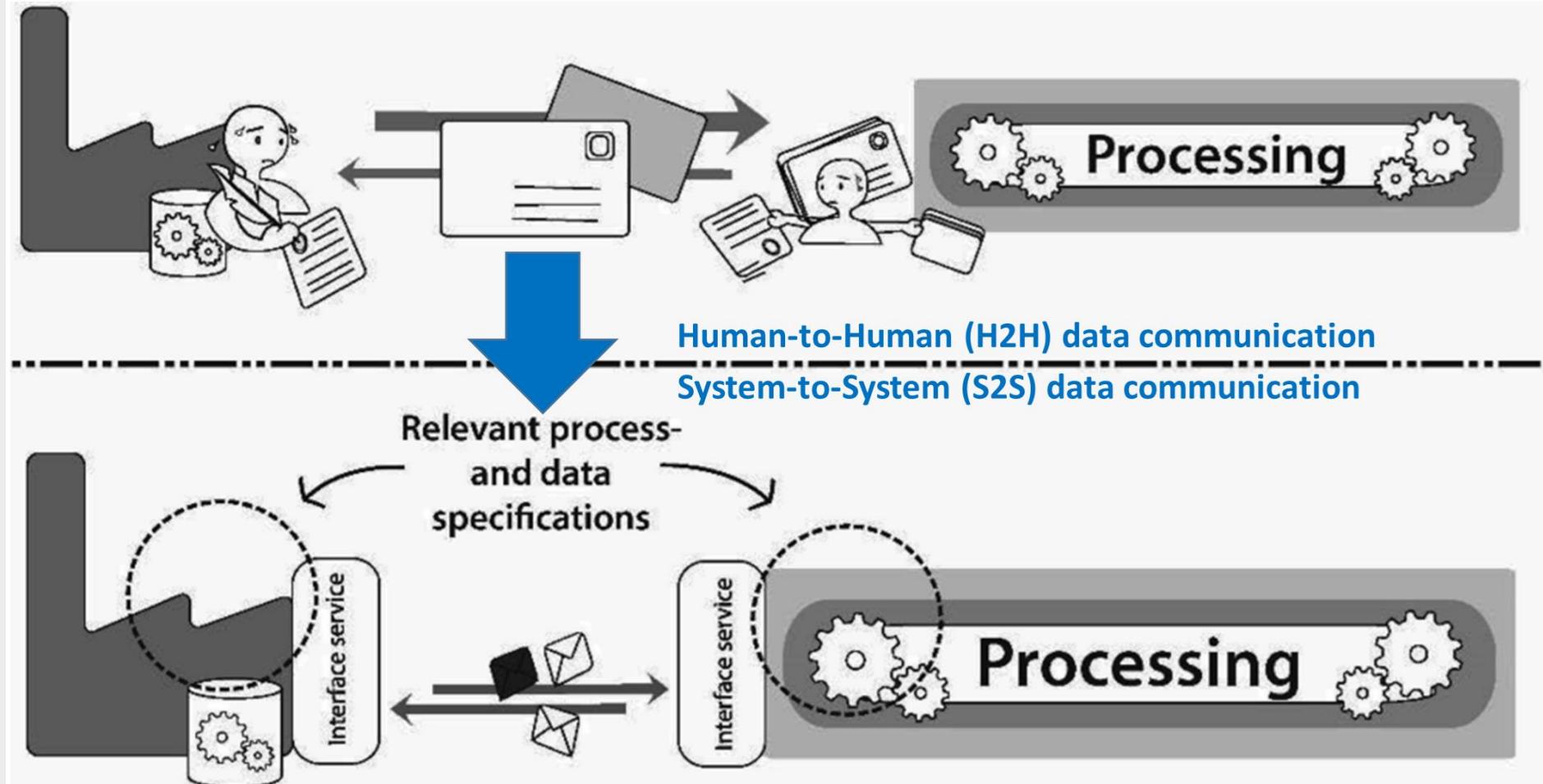


- Combination of **traditional** data sources with newly **emerging** sources and innovative '**Big Data**' analytics, e.g.
  - Traditional sources + innovative methods: XBRL-assisted scraping of company Annual Reports overcomes previous challenge of not being machine-readable
  - Emerging sources: Legal 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



# From Qs to S2S data communication



# 4. Internet of Things (IoT) – sensor data

Smart Systems Integration 2014 | 26.-27.03.2014

Waves of MEMS sensor proliferation

## Intelligent business architecture

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

Internet of Things and Services (IoTS)

Consumer Electronics

Automotive

3<sup>rd</sup> wave

2<sup>nd</sup> wave

1<sup>st</sup> wave

1990

2000

2010

2020



Bosch Sensortec

3

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.



**BOSCH**

# 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



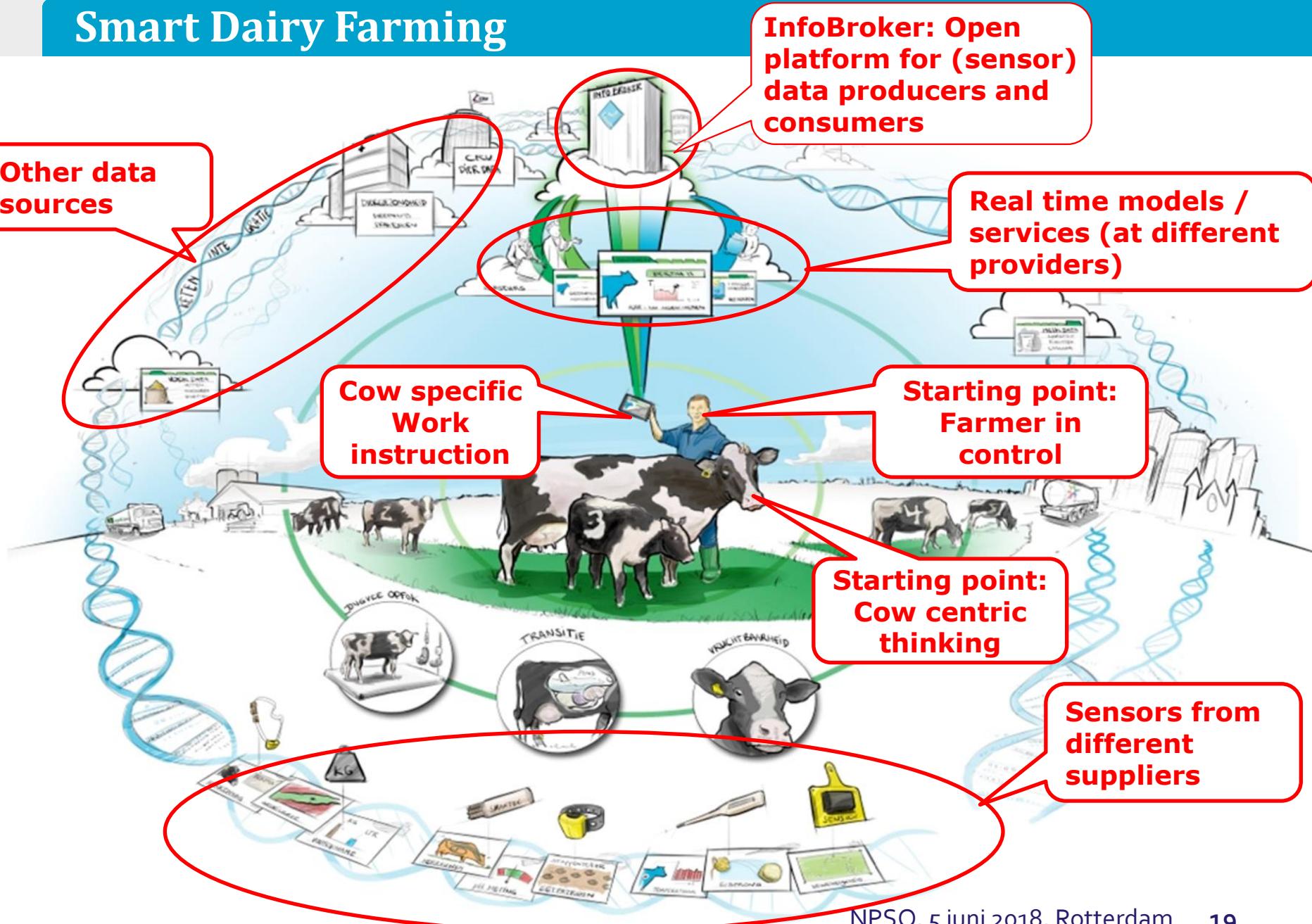
SmartDairyFarming

Vonder, M., 2017, Sensors going smart. Presentation at 'Big Data Matters' Seminar, Statistics Netherlands, 27 September 2017, Heerlen, Netherlands. (TNO Netherlands)

- Collecting environmental data:  
air quality, water quality, noise disturbance, etc.



# Smart Dairy Farming



# Smart Dairy Farming



## Sensor data on:

	Farm 1	Farm 2	Farm 3	Farm 4	Farm 5	Farm 6	Farm 7
# cows/calves	459	186	315	239	706	202	351
Behaviour	5x				5x		
Temperature	1x				1x		
Activity	9x	9x	3x	6x	5x	13x	9x
Milk production	16x	20x			1x	2x	19x
Feed intake	24x	24x				10x	24x
Weight	10x	6x	6x	6x	7x	6x	10x
Water intake			3x	3x			
Milk intake			7x	11x			

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

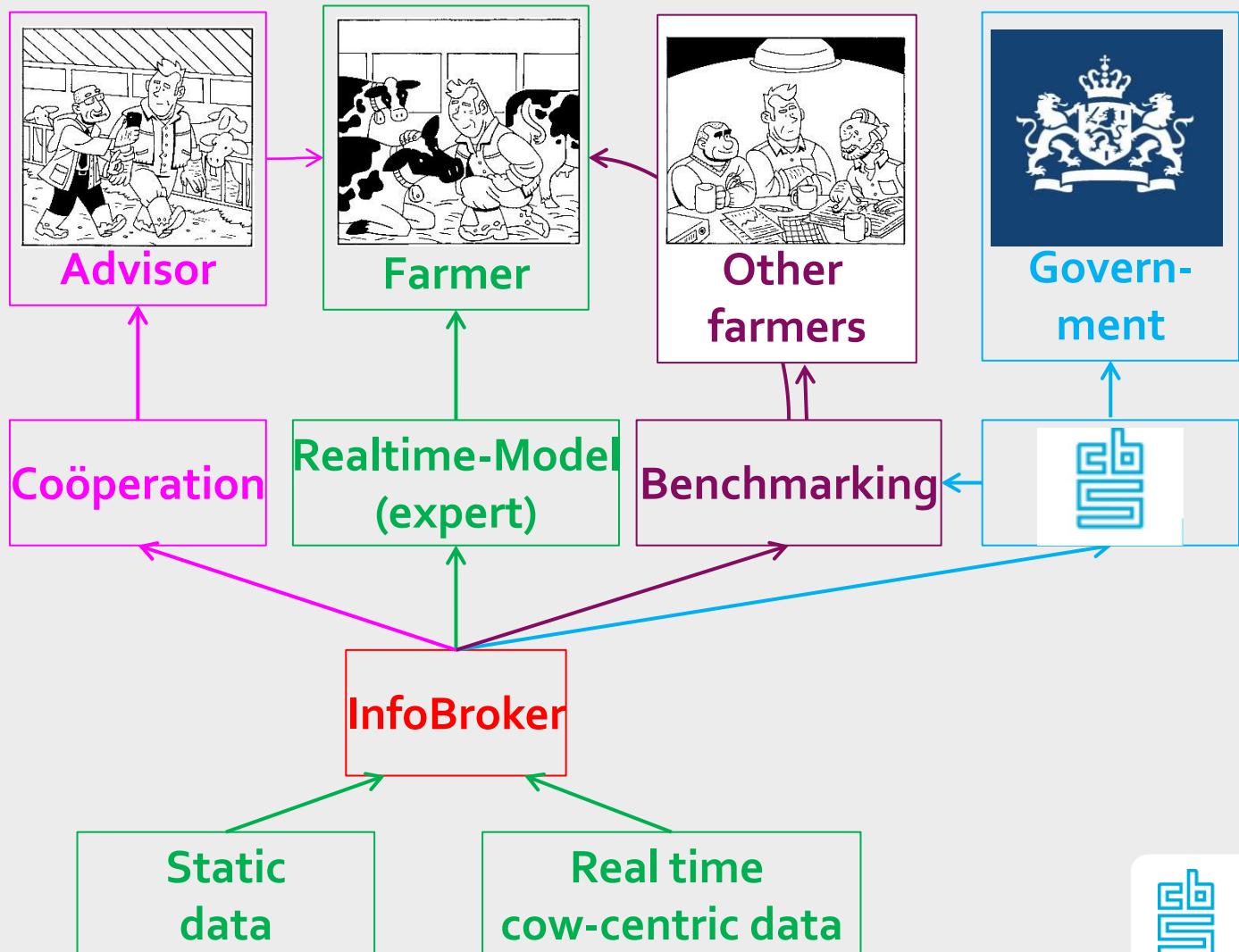
NB3: numbers in black are the sensor fields within a category (e.g. 3 fields related to waterintake)



# Smart Dairy Farming: InfoBroker



Some scenario's  
for using  
the  
InfoBroker:



# 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 dataverzamelingsmethodologie blijft in ontwikkeling,  
ook voor bedrijfsdata!*

En wat betekent dat voor de toekomst: ...



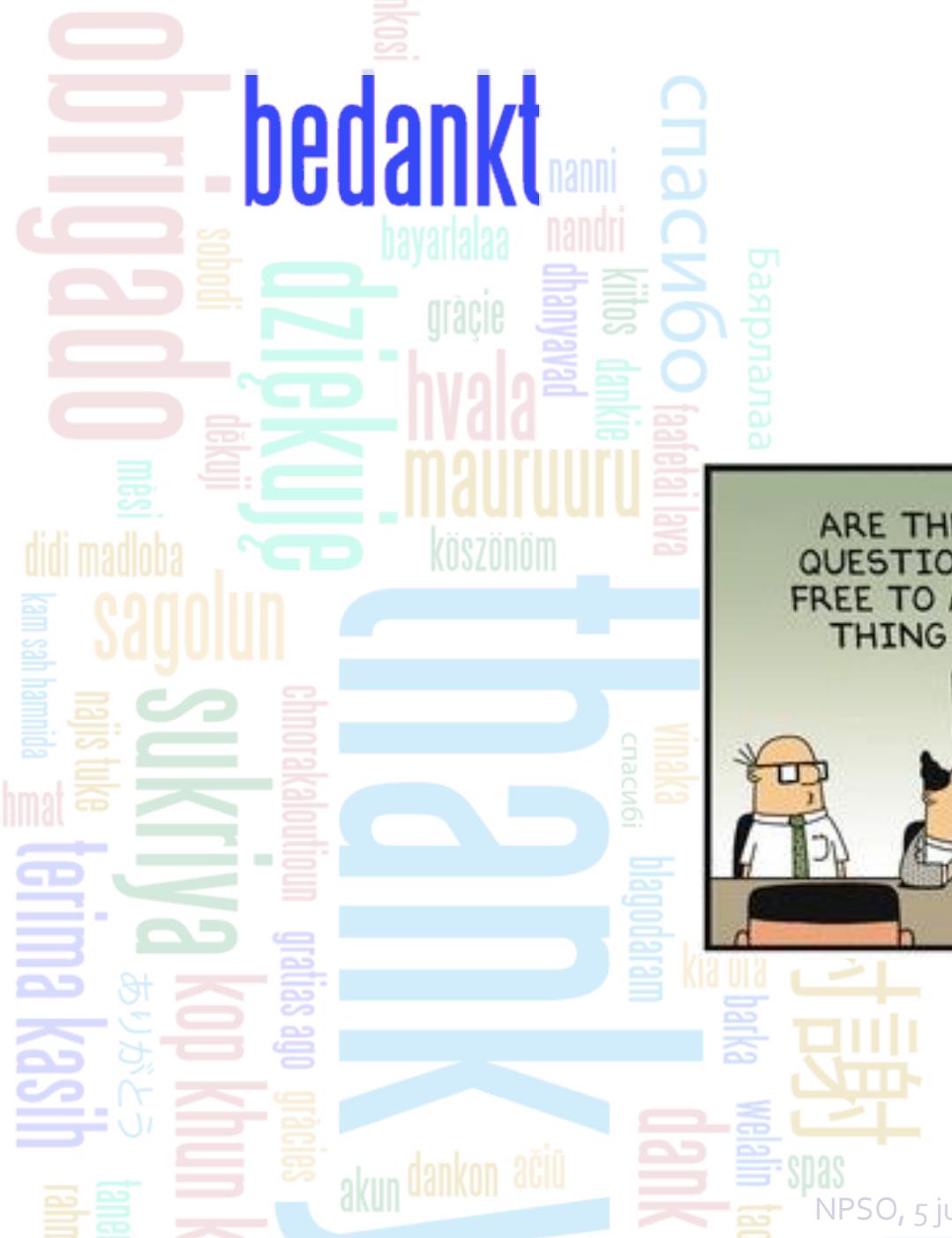
# Visie bedrijfswaarneming CBS

- Reductie van vragenlijsten door:
  - Betere ontsluiting van secondaire gegevens
  - S2S data communicatie voor financiële data, sensor data:
    - Individuele bedrijven (grote)
    - DSCs voor bedrijfscoöperaties
  - Gebruik van nieuwe data bronnen (internet)
- Ontwikkelen van een nieuwe methodologie:
  - Conceptualisatie (statistische concepten): nog belangrijker
- Speciale aandacht voor grote/complex 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



감사합니다

তোমাকে ধন্যবাদ



Баярлалаа

спасибо

vinaka  
 спасибі

谢

dank

благодарю

спасибі

спасибі

спасибі

ARE THERE ANY  
QUESTIONS? FEEL  
FREE TO ASK ANY-  
THING AT ALL.

