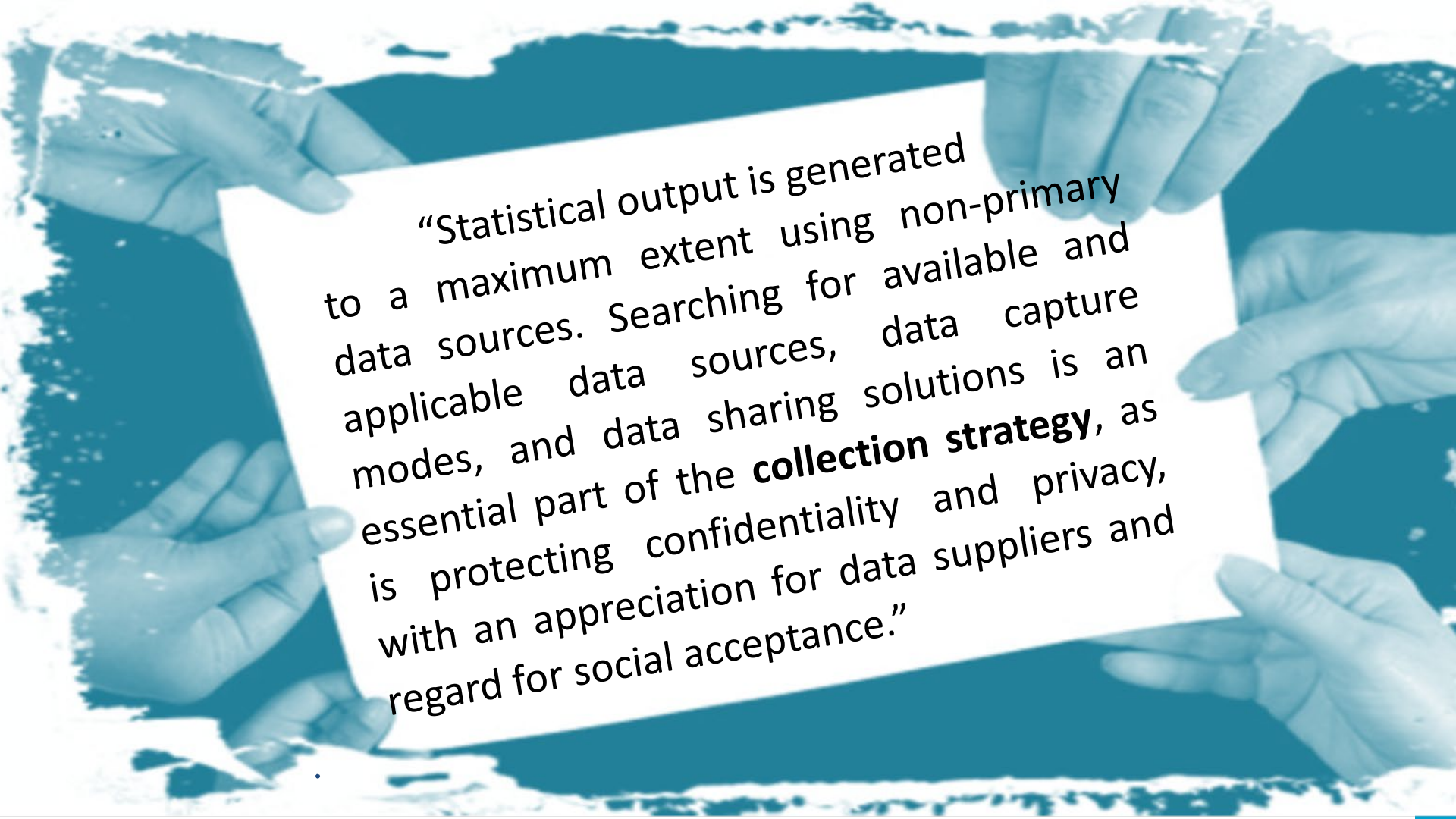


Future Advanced Data Collection

The Future is NOW

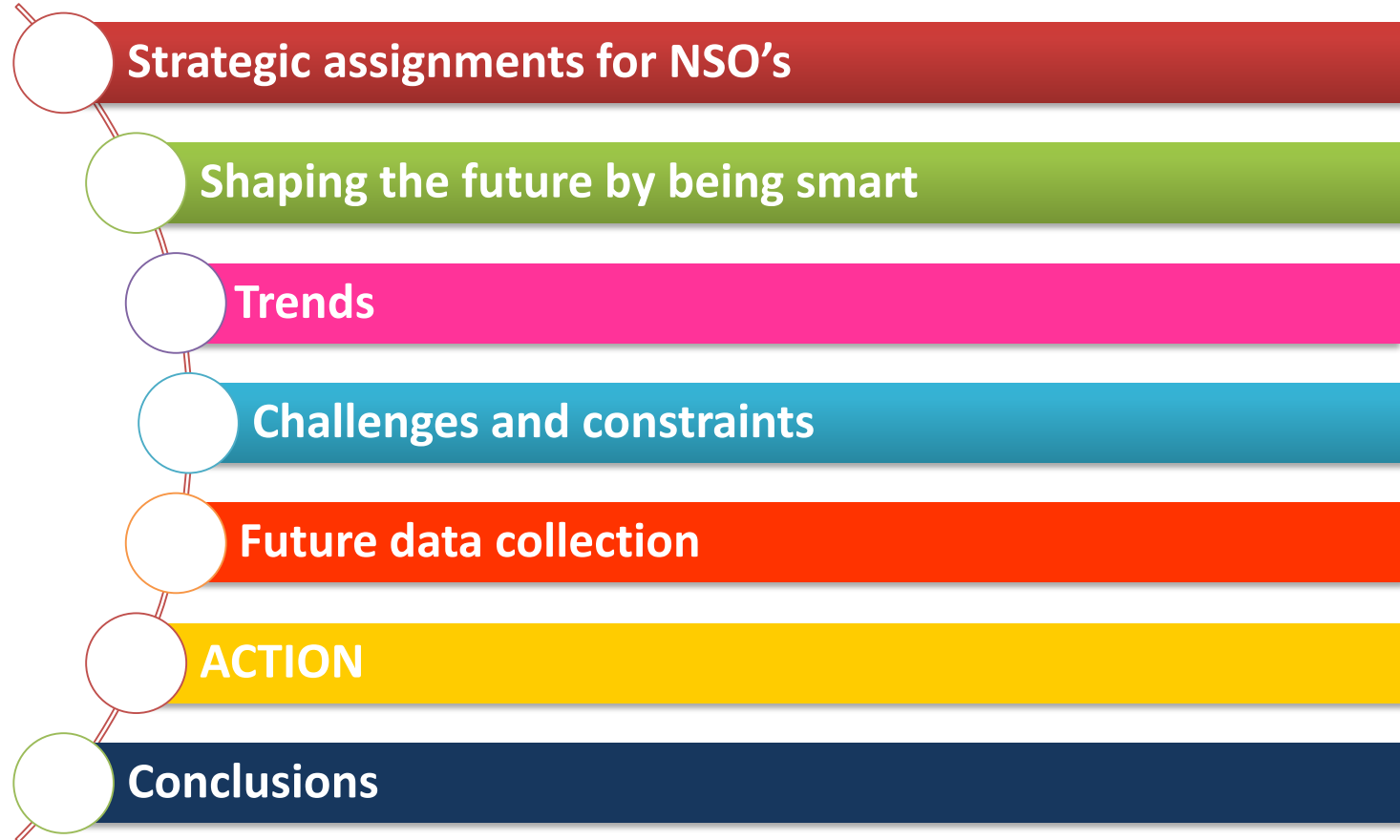
Irene Salemink

1 oktober 2019

A blue-tinted photograph showing four hands holding a white rectangular piece of paper. The paper is slightly tilted and contains a paragraph of text. The background is dark and textured, possibly a wall or a large sheet of paper. The hands are positioned at the corners of the paper, with fingers visible. The text on the paper is in a serif font, with the words "collection strategy" in bold.

“Statistical output is generated to a maximum extent using non-primary data sources. Searching for available and applicable data sources, data capture modes, and data sharing solutions is an essential part of the **collection strategy**, as is protecting confidentiality and privacy, with an appreciation for data suppliers and regard for social acceptance.”

Future data collection



Strategic assignments NSO's



Strategic assignments NSO's

- Demand driven and user centric
 - Broad range statistics/information
 - Fit-for-purpose
 - Accurate and timely
 - Policymakers
 - Private sector
 - Society
- Fact based policy making
 - Actionable Intelligence

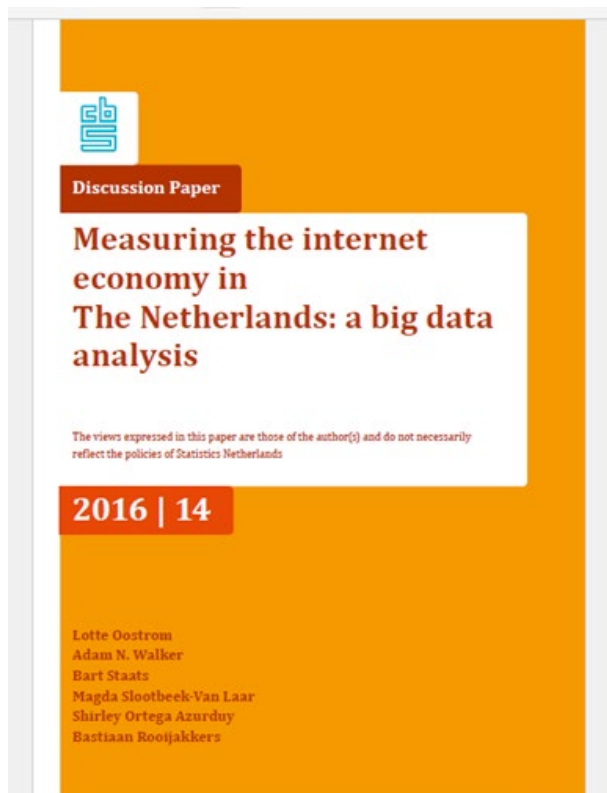


Strategic assignments NSO's

- Dealing with complexity
 - Complex societal and economic phenomena
 - Real time statistics
 - Tailored to all aggregation levels
 - More detail, regional, local
 - Fit –for-purpose quality

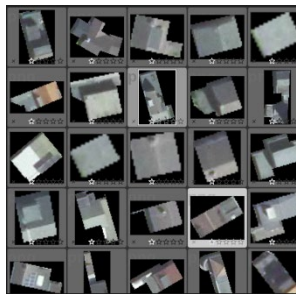
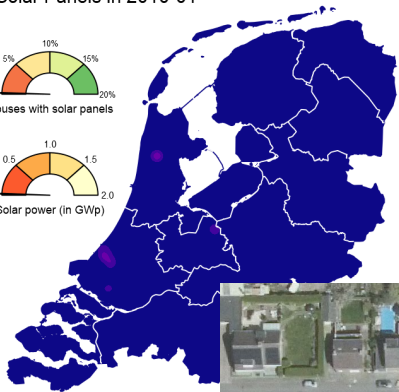
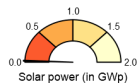
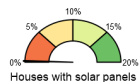


New information products



Complex phenomena

Solar Panels in 2010-01

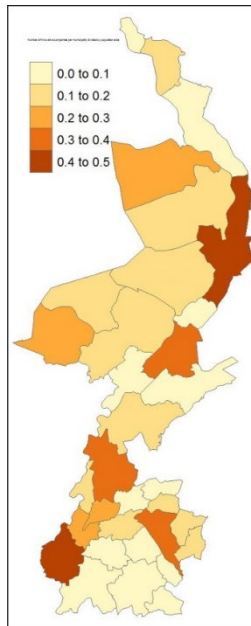
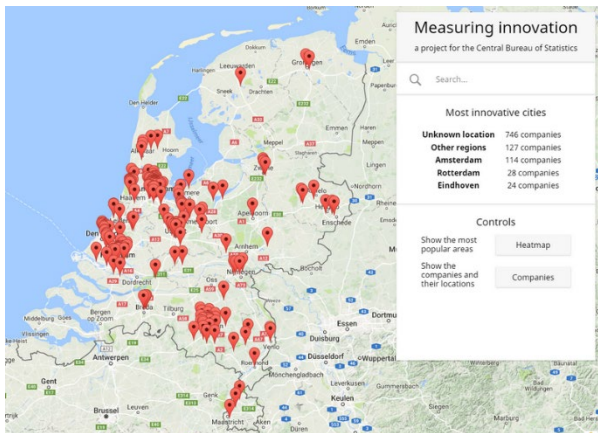


Energy transition

Support local government with insights to support sustainable development goals

Production Installation Register (solar panels)
VAT registers from Tax authority
Basic register Addresses and buildings

Complex phenomena



Innovative enterprises

Web scraping and text mining to identify small innovative enterprises

Classification

Linking to background characteristics

Shaping the future by being smart



Shaping the future by being smart

- Official statistics become “smart statistics”
 - Smart technologies
 - Smart data
- Guaranteed confidentiality
- Privacy by design

Trusted Smart Statistics



Trusted Smart Statistics

- Facts for evidence based policy making
- Quantitative monitoring of development and progress of policy
- Society oriented
- Reliable and innovative
- Protect confidentiality and privacy



Nature of data collection is bound to change

Trends



Trends

- Datafication



- Data storage and access

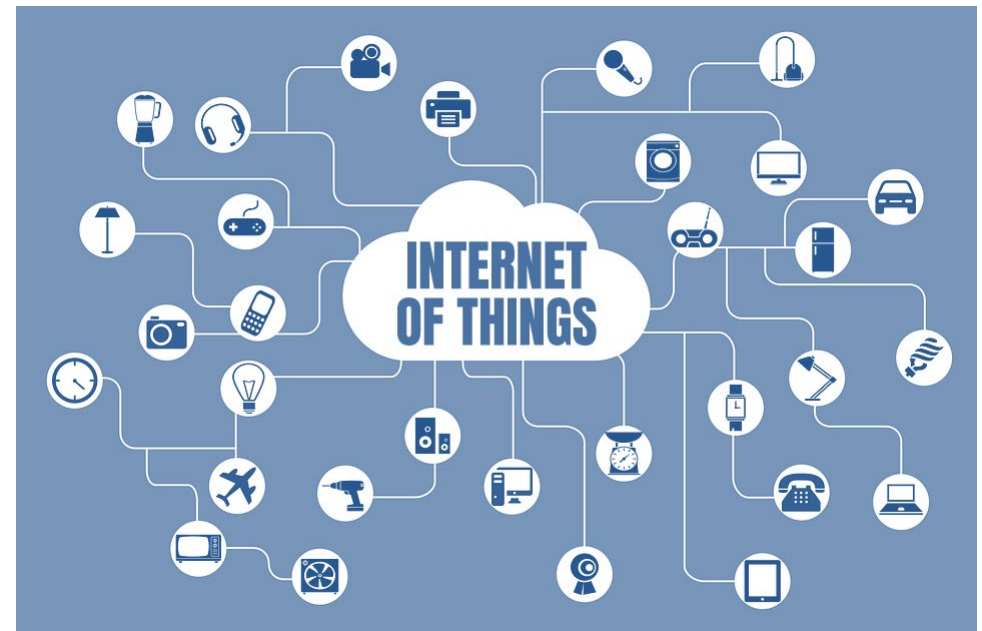
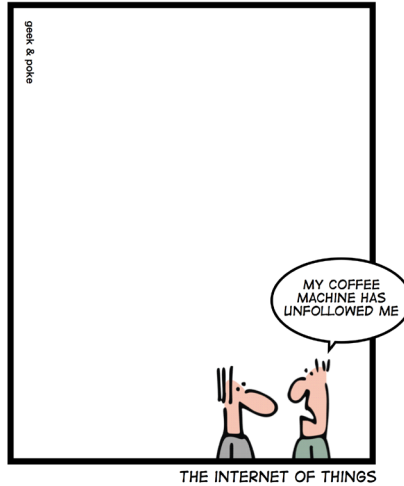
- Computing power and analytical capabilities



Administrative data use at CBS



Sensors and IoT



We Live in Exponential Sensor Times

Period	Annual sensor volumes	New industry consuming sensors
1960s	10,000s	Defense, Avionics
1970s	100,000s	Process Control, Automotive
1980s	1,000,000s	Medical
1990s	10,000,000s	Consumer
2000s	100,000,000s	Mobile
2010s	10,000,000,000s	Wearables, mHealth, Internet of Things, Big Data
2020s	10,000,000,000,000s	Internet of Everything, Social Cloud

Mobile market accelerated the sensor growth by an order of magnitude

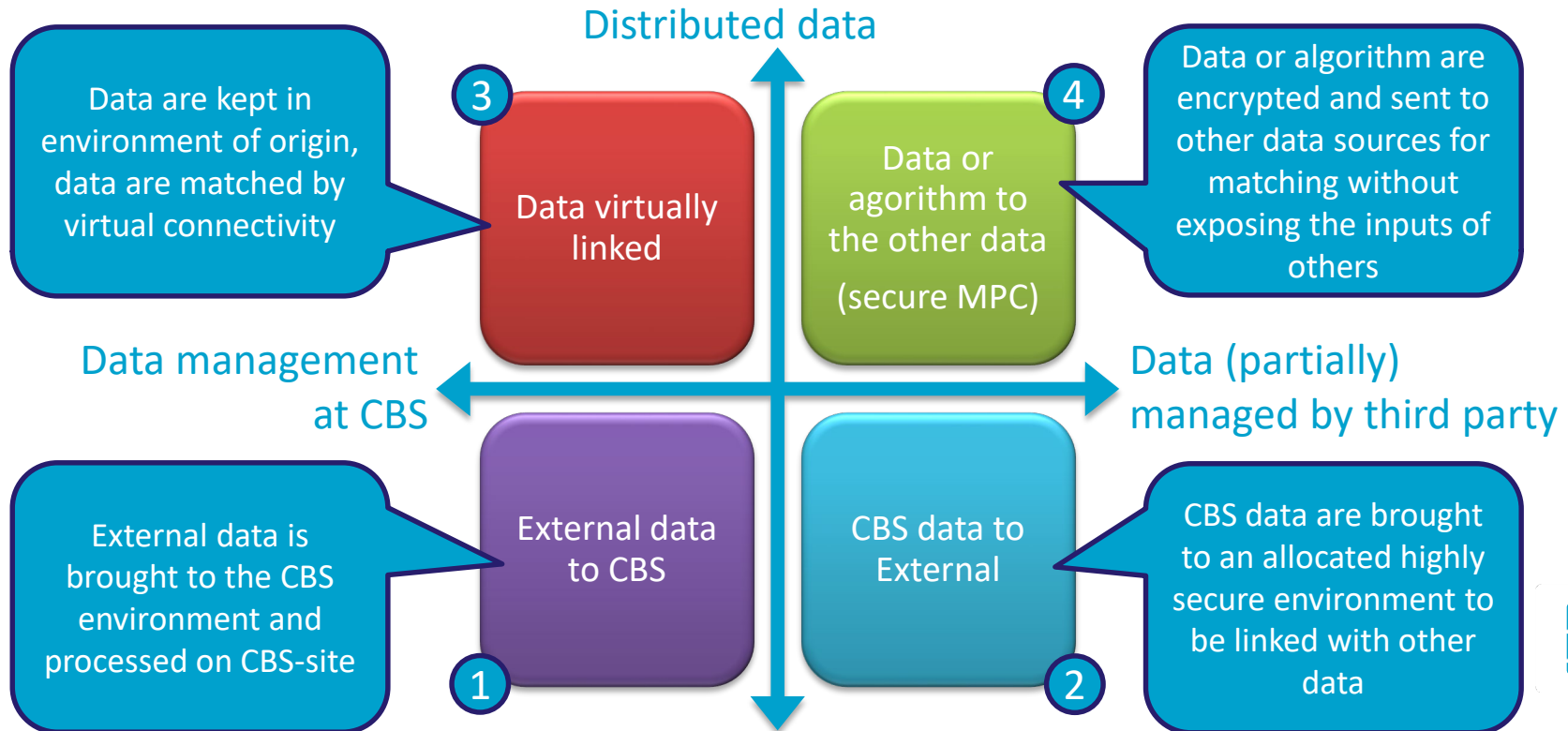


Data storage and access

- Need for :
Fast, easy, and free access to all relevant data
- Reality:
Datasets; too big to copy, not allowed legally to “leave the building”, need matching between multiple (different) sources, require knowledge, only the proportion that is needed can be accessed...
- Solution:
Dependent on type of data sharing



Data architecture patterns

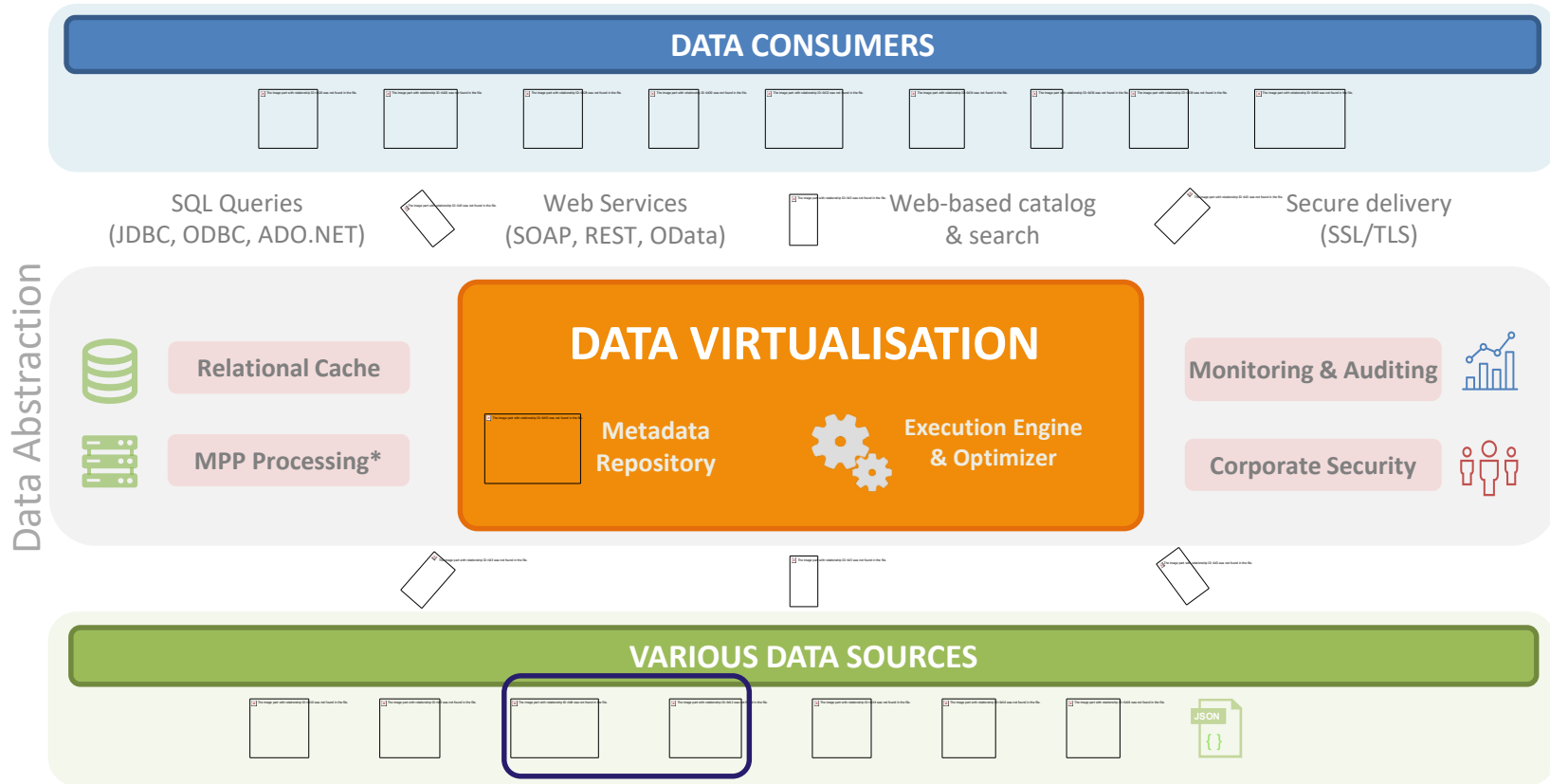


Data storage and access

- These 4 patterns come with capabilities that need further investigation
 - Privacy preserving analytic techniques
 - Secure multi party computation
 - Data virtualisation and data abstraction
 - Metadata management



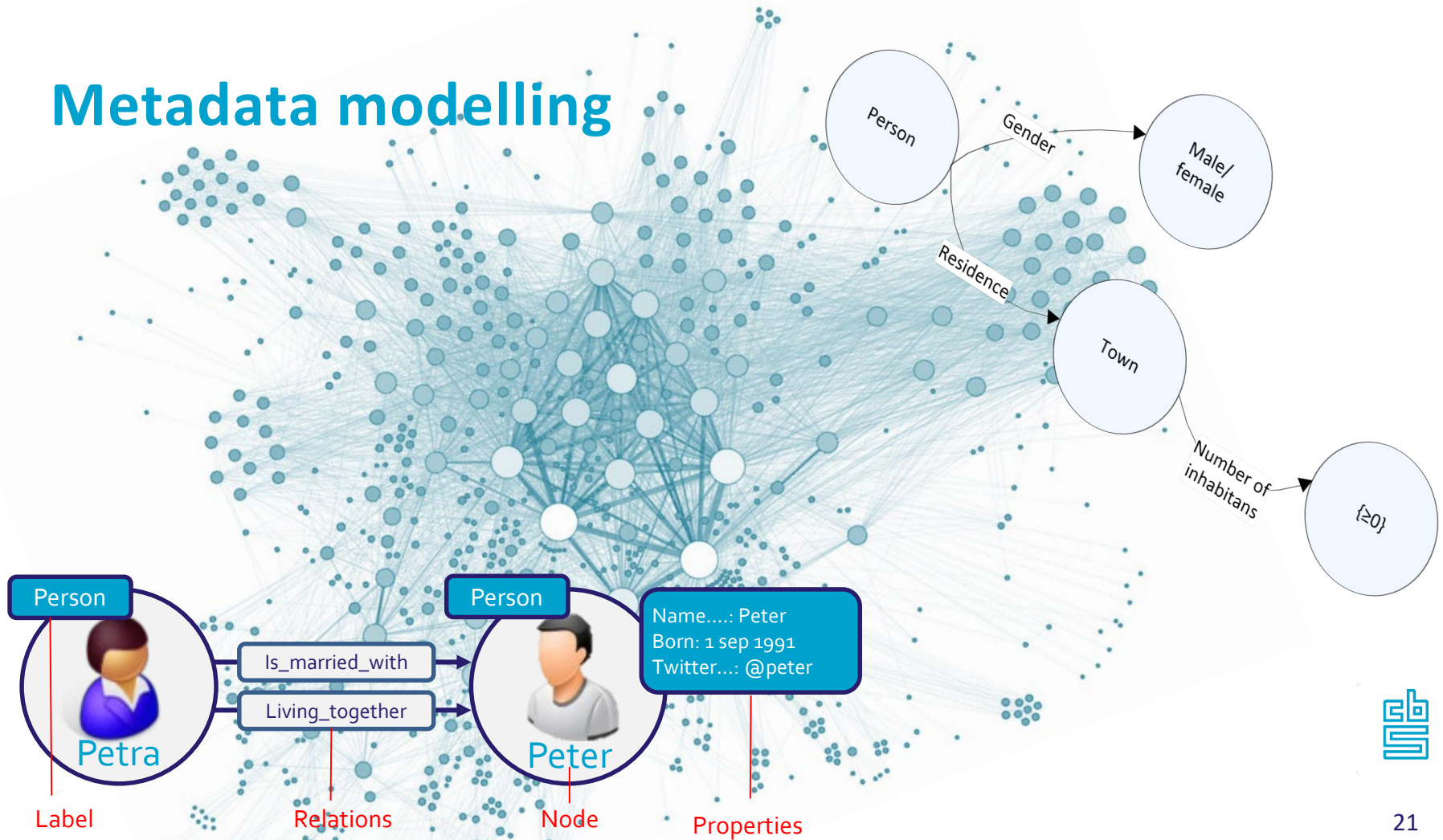
Data virtualisation architecture



* Massive Parallel Processing



Metadata modelling



Computing power and analytics

- Explore and Confirm
 - Deductive; data analysis to explain, check or validate ideas
 - Inductive; data analysis to generate new ideas
- Edge analytics
 - Analysis and data quality framework are brought to the data gathering devices instead of moving the data to the (centralized) analytics and quality frameworks



Challenges and constraints



Challenges

- Data gap



- Burden to society and response rates



- Privacy protection and difficult access to data



- Methodology



Future data collection

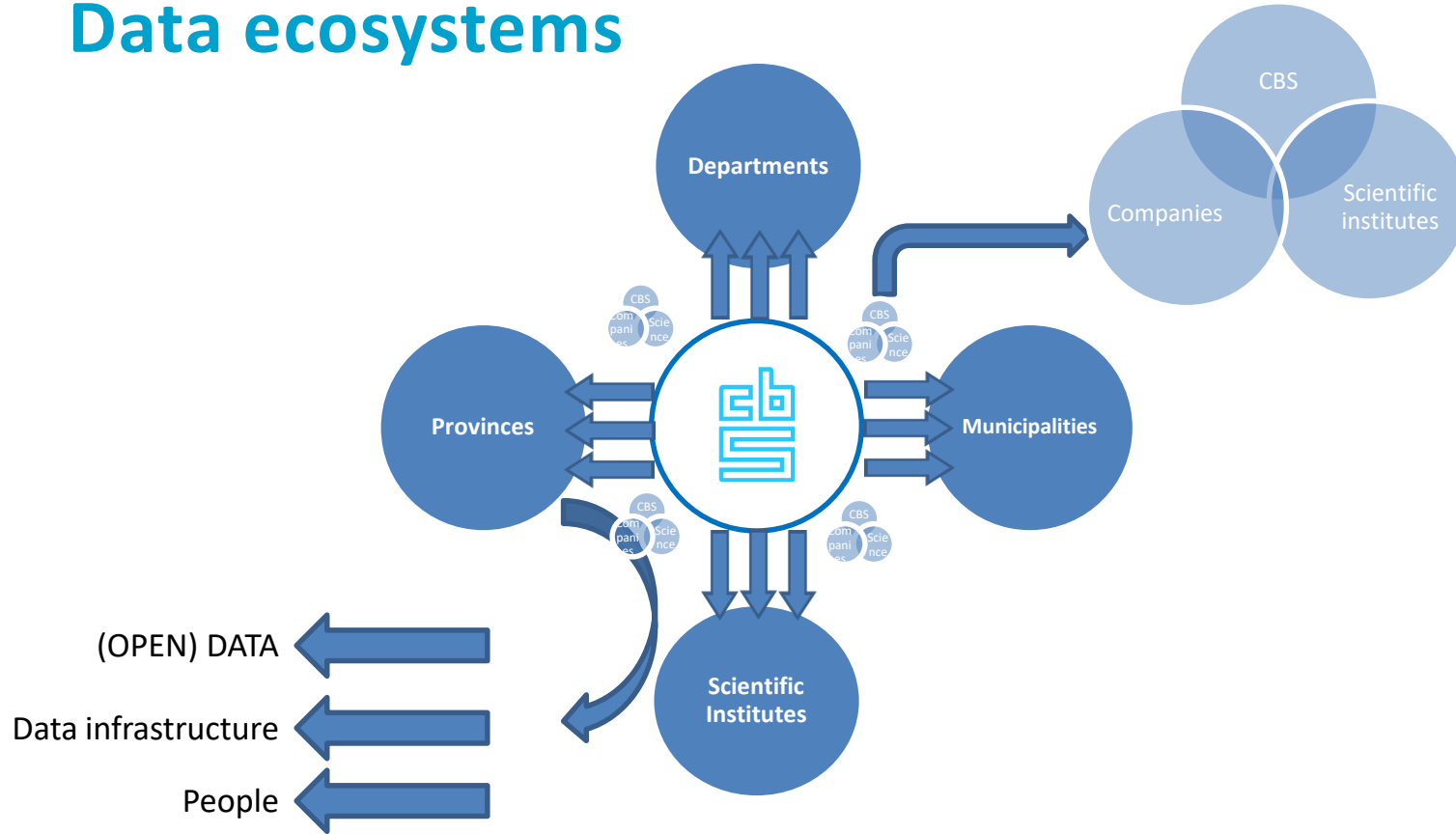


CBS in relation to data owners & end users

- Signalling trends and potential future needs
- Meet the public demand for information
- Data hub
- CBS fulfils role of a platform
- Enabling broad cooperation between
 - Governmental bodies
 - Municipalities
 - Companies
 - Scientific institutes



Data ecosystems



Tapping into and unlocking new data sources

- Datascouting
 - Awareness of data sources being there
 - Usability and availability of the data source
 - Organization and facilitation of the acquisition, tapping into and unlocking of new data sources
- Data scout en community
 - Link between internal and external stakeholders



Surveys – primary data collection

- CAWI hybrid mode
- Multi mode and multi source
- Custom-fit data collection
- Experimenting
 - Crowdsourcing
 - Interactive Voice Response
- Validation of alternative sources / data



Call to ACTION



Call to action

- Methodology
- Collect – Connect – Link
- Metadata
- Proprietary sensor networks
- Legal frameworks and social acceptability

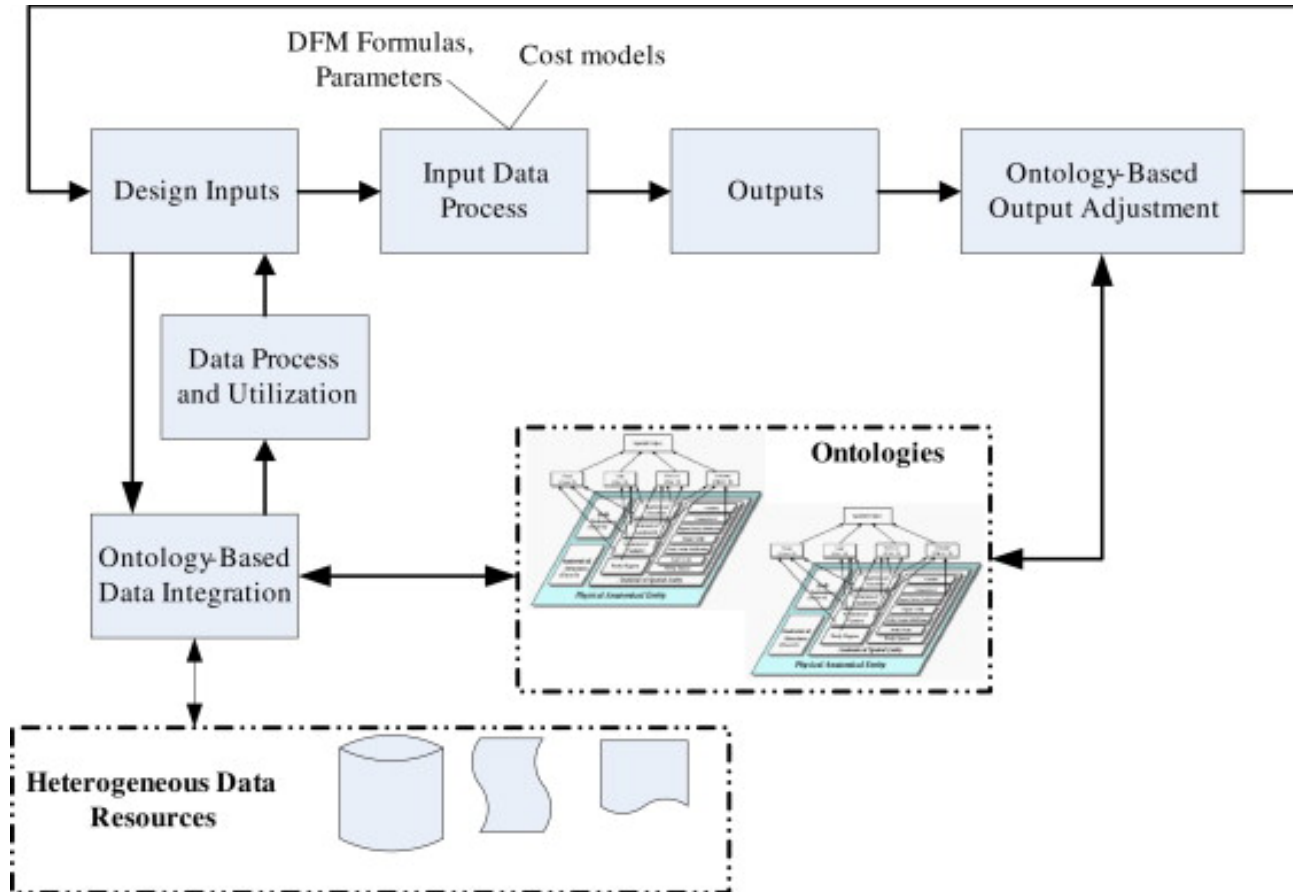


Methodology on combining data

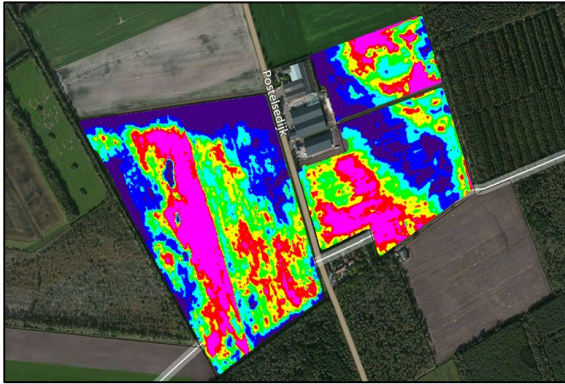
- Multiple collection modes and different data sources
 - units to be matched do not equal source units
 - sources do not contain overlapping units
 - matching errors
 - variables in multiple sources with different measurement errors
- Complex mixed mode designs
- Extent and/or combine existing techniques
 - Probabilistic matching
 - Matching with supervised machine learning
 - Synthetic matching
- Integration by design



Integration by design

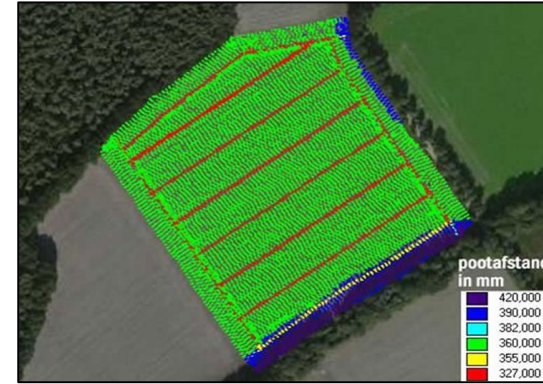


Sensor networks - precision agriculture cycle



Winter

- Draw parcels
- Yield potential
- Tractor lanes



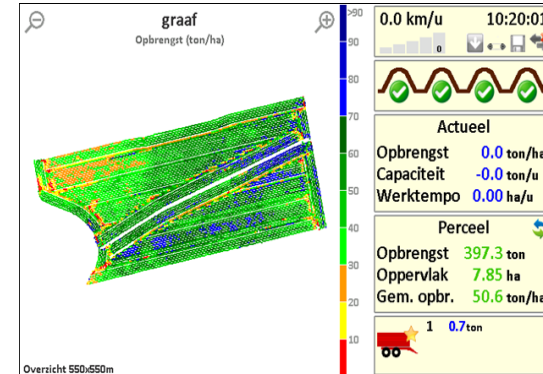
Spring

- Fertilization
- Variable planting



Summer

- Additional fertilization, pesticides & water
- Based on sensordata



Autumn

- Harvest
- Storage



...and also....



- Hotspots
- Portal
- Reference Ledger Scheme

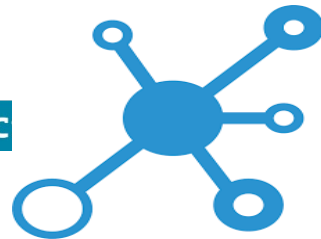


rijksuniversiteit
groningen

- Post Doc



Center for
Big Data Statistics

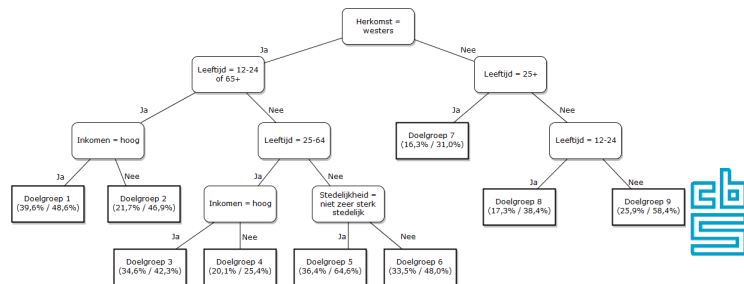


- Redesign HBS
- Essnet smart surveys



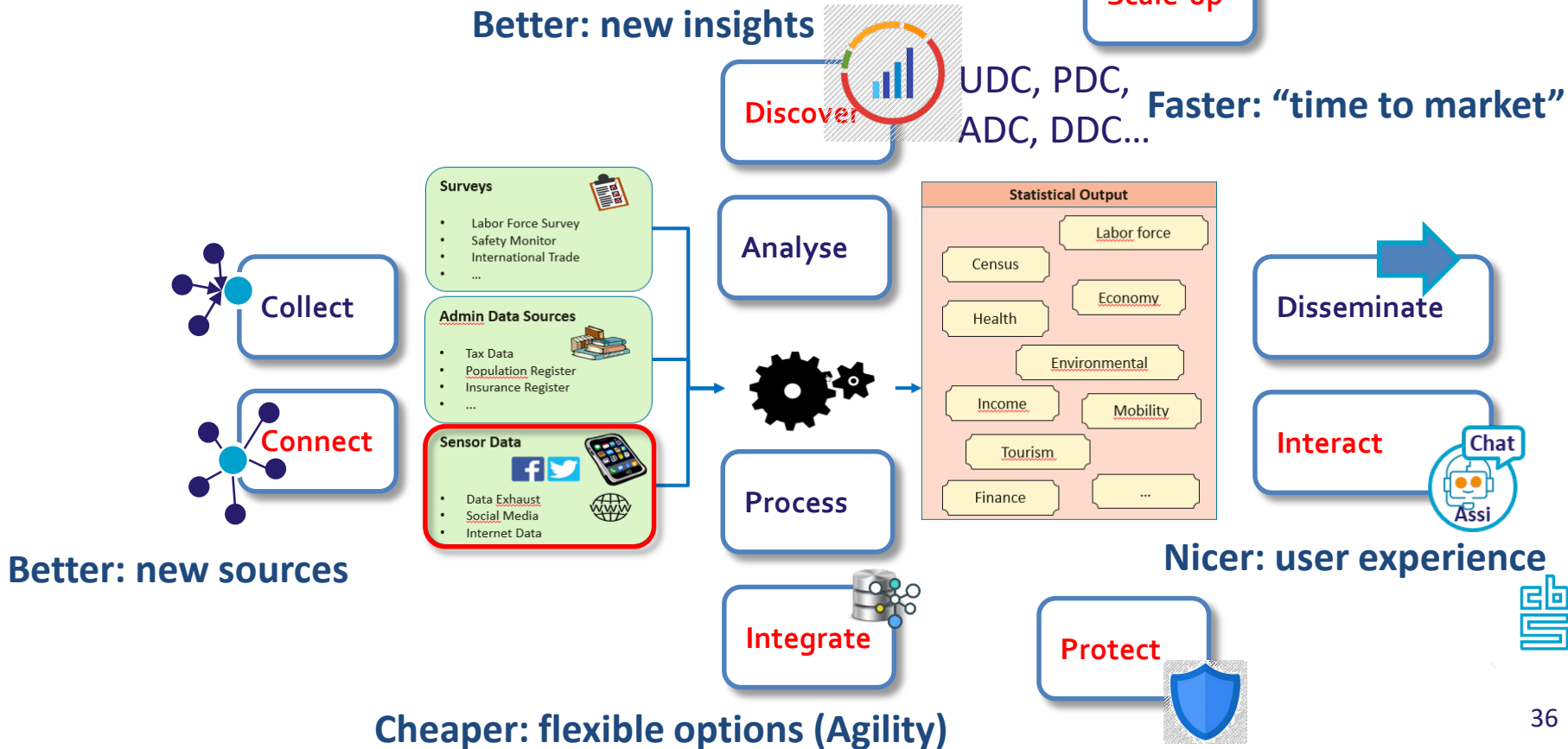
- DNB – e-recognition
- Bd – Digi coupling
- UWV – Apache NiFi
- Bol.com - Cloud

CDL



- Primary data collection 2.0B5

Summary



Conclusions

- Influence of NSI on content of captured data diminishes
- Availability and technology greatly determine the data to be collected
- New sources increase complexity
 - Validation, measuring errors & biases, mixed modes and multiple sources are a challenge
- AND possibilities
 - New – More – Cheaper – Faster

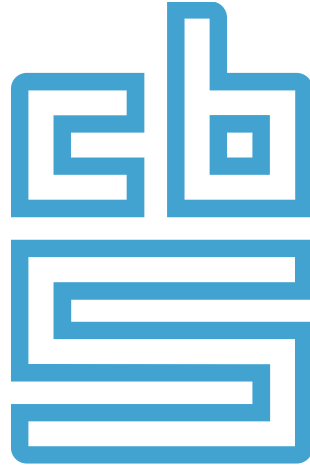


Conclusions

- Data collection becomes Advanced
- Data collection through data connection
- Statistical production process has to follow
- Integration by design







Facts that matter