Preconditions for effective "data science" deployment

Organisational and statistical challenges

Sofie De Broe, Johan van der Valk

17 December 2019

Mission and focus

Health, energy, labour market, mobility, safety, economic indicators

PROJECTS IMPACT 1

PROJECTS IMPACT 2

PROJECTS IMPACT 3

NEW STATS

GRANTS

Research: BD methods and techniques
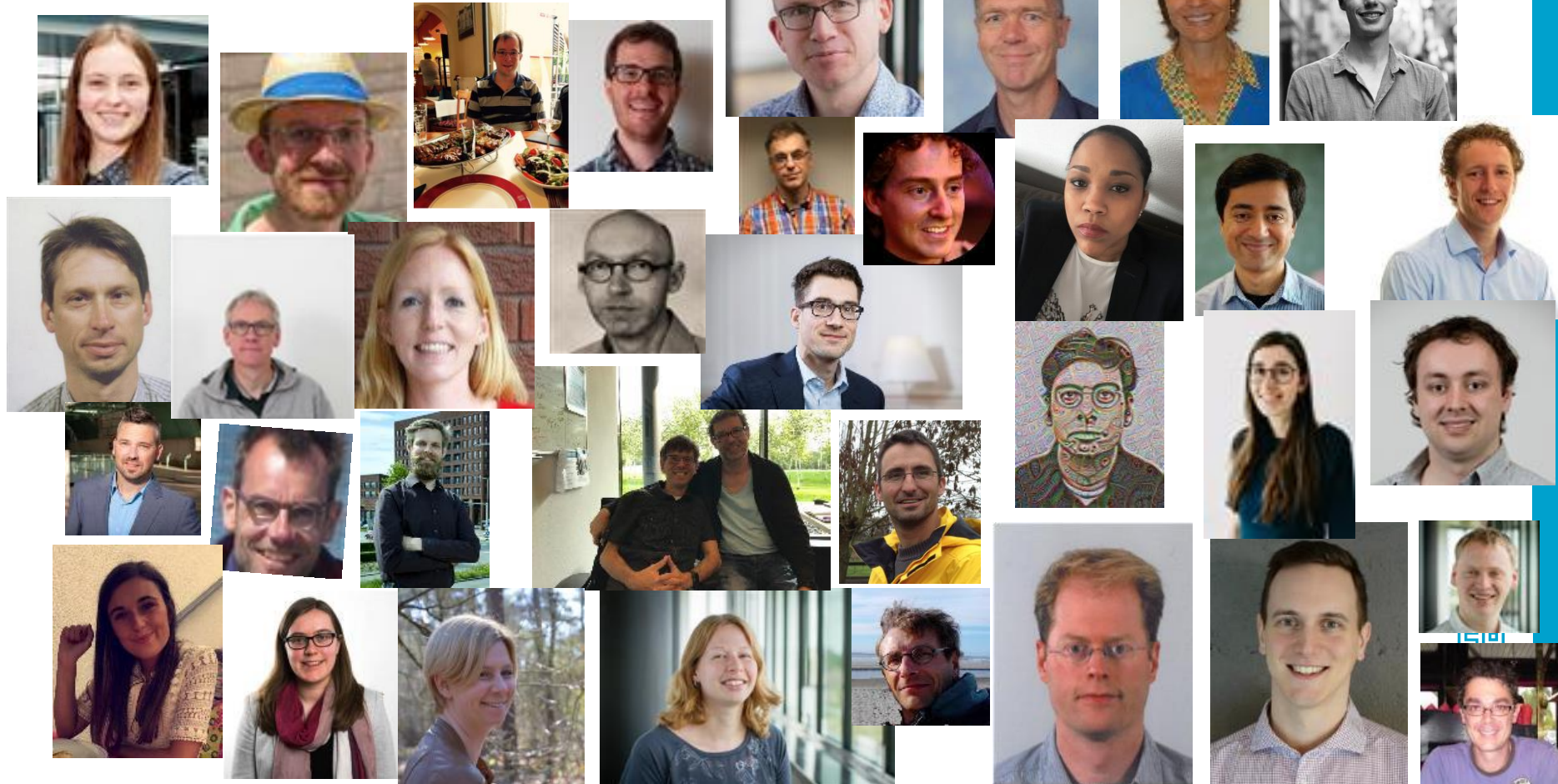
| Training | Data scouting | Ethics | IT |

Center for Big Data Statistics

Using state-of-the-art data science methods and techniques with new and combined datasources to obtain faster, cheaper and better statistical information on burning and relevant policy questions and challenges

# 1. Multi-disciplinarity

2. Eco-system

3. Datascouting

Bron	Data	Toepassingen
Cargonaut	Aerial cargo in NL	Economy and trade
Coosto	Social media posts	Desire to move, social tension
Enelogic/Liander	Smart meter data (testset)	Energy consumption
Tennet	Quantity of current	Energy consumption
KNMI	Solar radiation data	Solar power production
Tomtom	Floating Car Data	Traffic intensity
T-mobile	Signalling data	Movement/displacements

Example: Foreign persons in the Netherlands: Germans (left), Belgians (right) March-April 2018
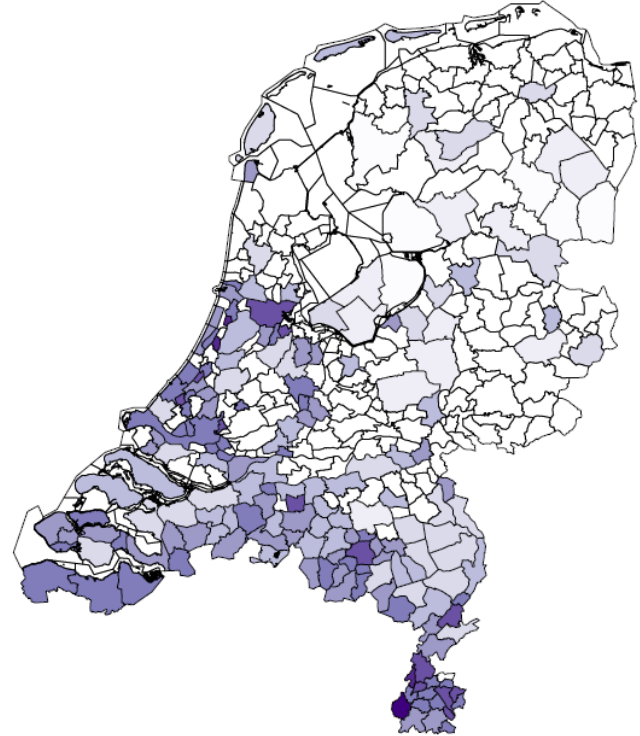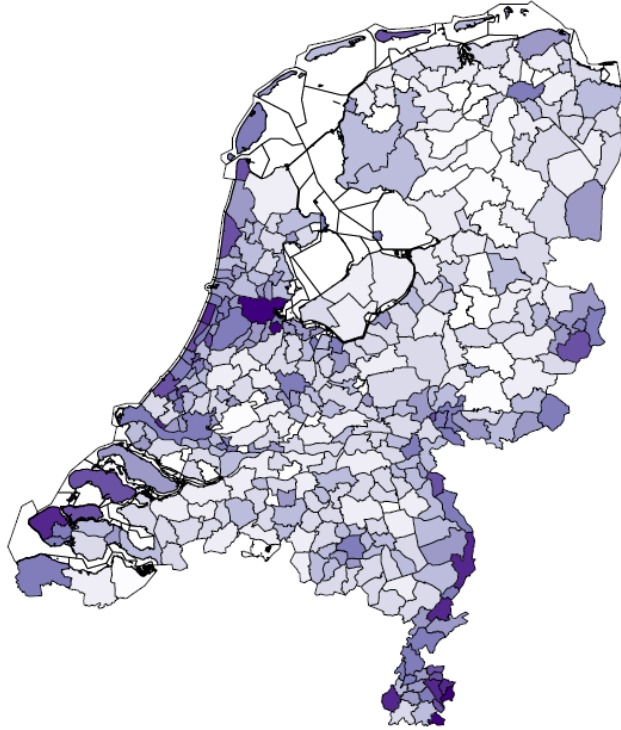
4. Funding: Deep Solaris Project

5. Relevance: Urban Data Centres
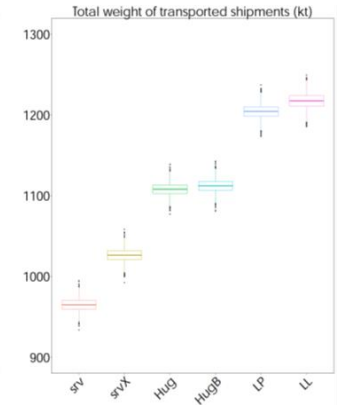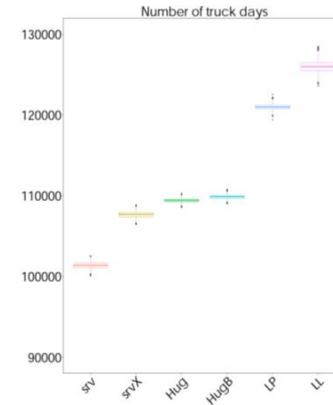
6. Communication

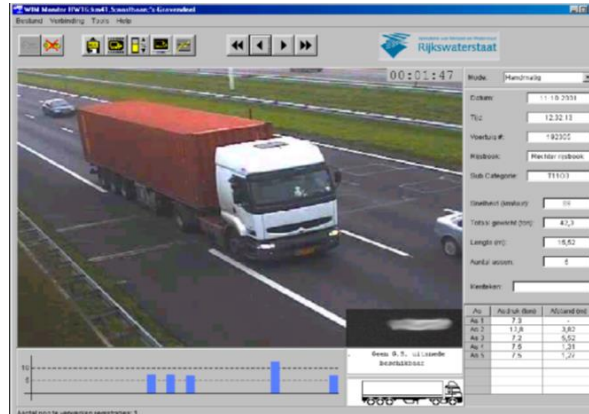Methods: ML to detect solar panels

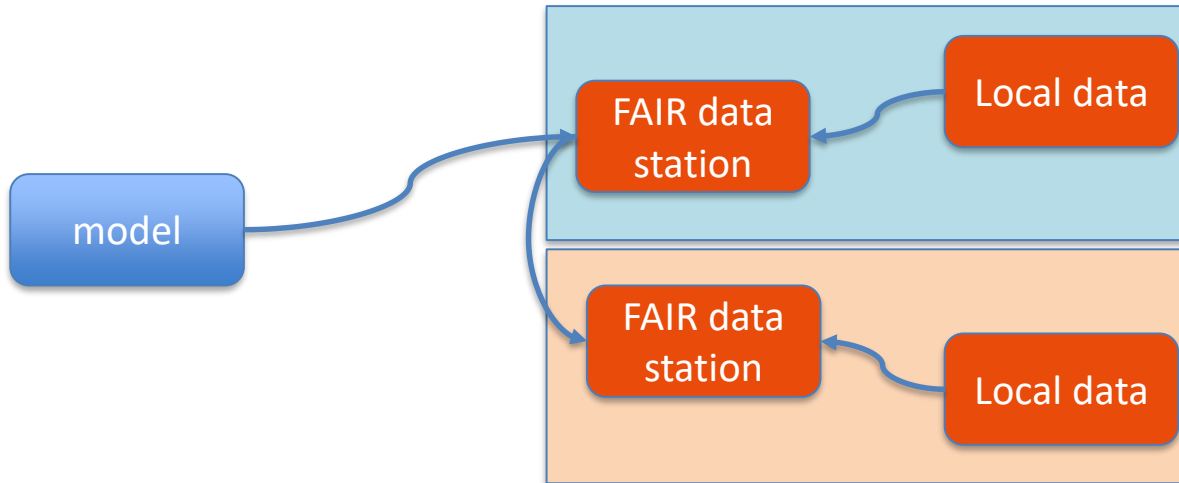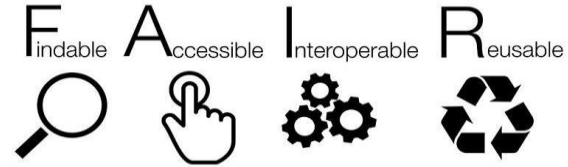Social media and text mining: cyber security

URLs: small innovative companies

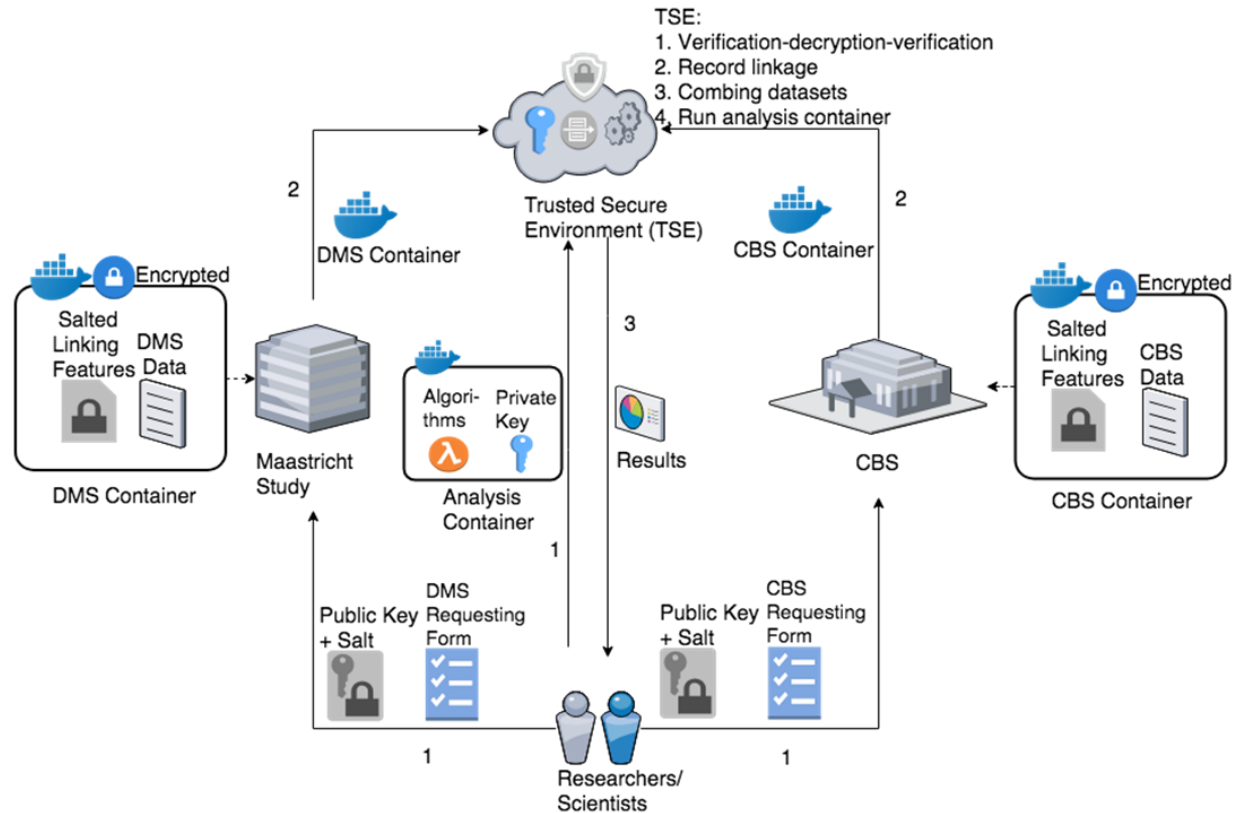Visualisation techniques: dotmaps

7. Research: combining data sources
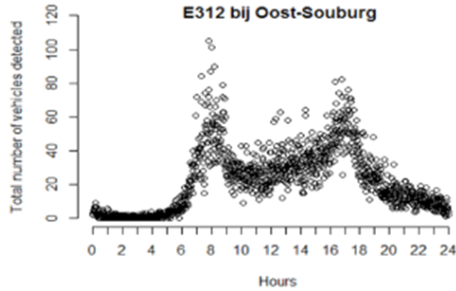
8. Privacy Preserved Data Sharing and Ethics

- Example 1: Analysing combined data within a 'thrusted secure environment'
- The set-up of the systems guarantees that both parties will get no insight in the underlying datasets, but are only able to see the results

9. Implementation

1. Adjust statistical proces
2. Software/IT infra
3. Q and stability of the source
4. Train staff
5. ……
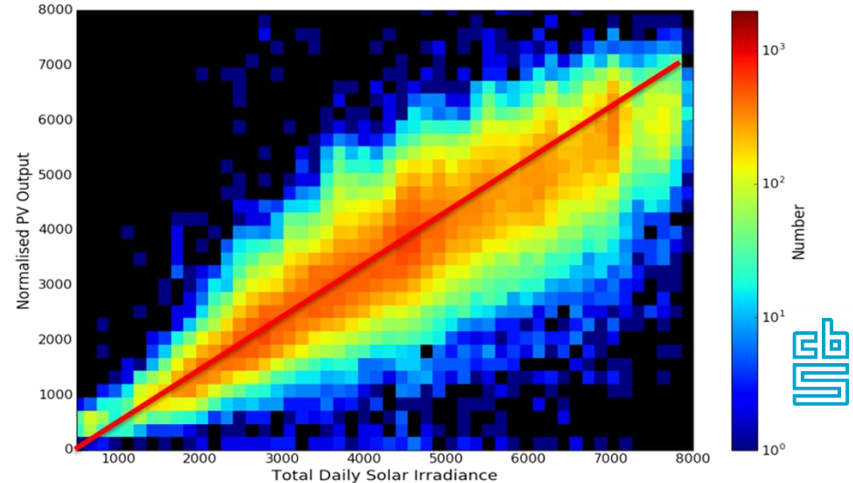
10. Questioning output and practises

Are we still making the right statistics?

Can we tackle policy questions nationally?

Are we still making the right statistics?
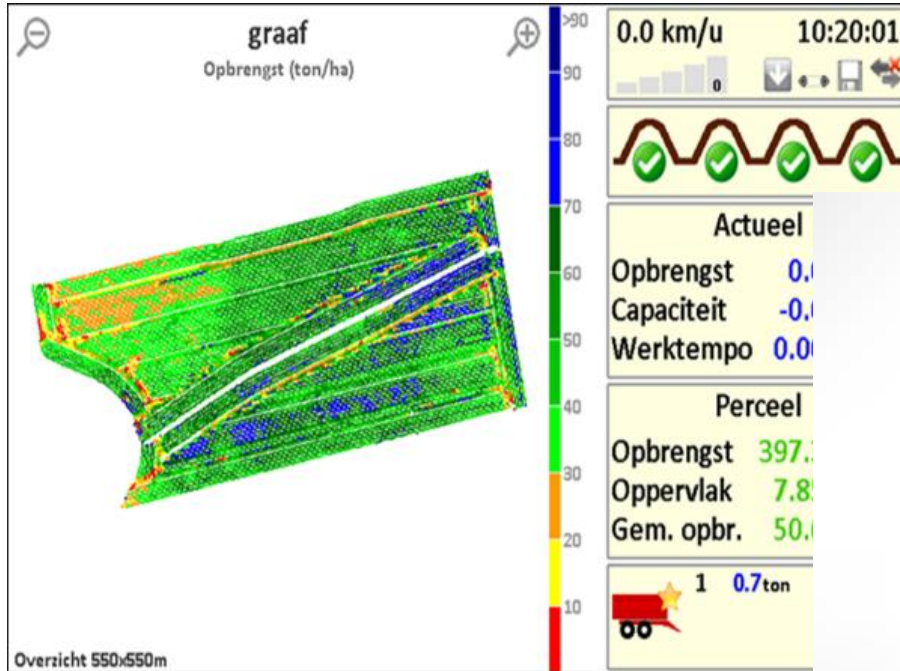

Display oogstmachine

Half of the EU-region consists of an international border

150 miljoen inhabitants representing 40 procent of the EU territory

Can we tackle policy questions nationally?
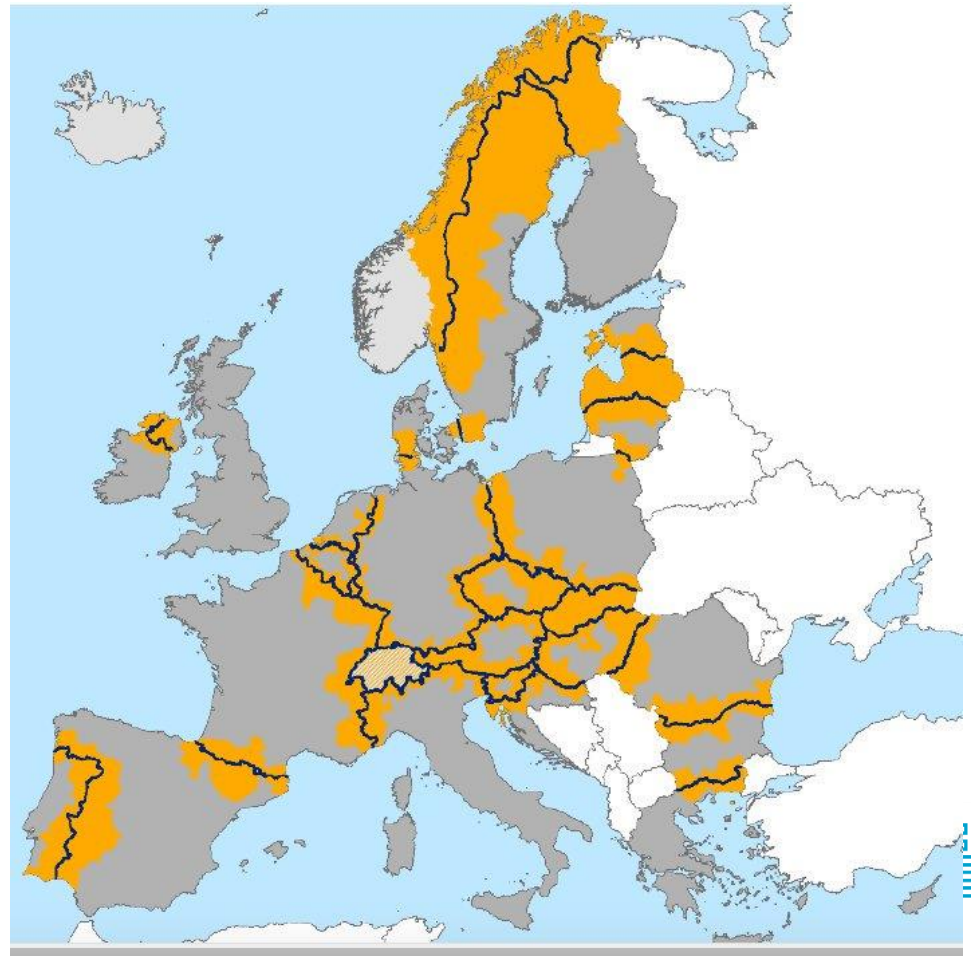
Producing cross-border statistics

Organisational and statistical challenges

Johan van der Valk

December 2019

Cross-border statistics: an example of new demands that impacts on how we work

Cross-border statistics not available?

➢ Need for comparable data on the border regions: low regional detail and harmonised

➢ **EU data**: Harmonised statistics of Member States with limited regional detail

➢ **National data**: regional detail with specific methods

Production information is challenging

Cross-border statistics are difficult to produce, because they …

- Require low regional detail,

- using international comparable methods and sources and

- measuring flows of persons & goods

➢ Strong incentive to explore new (BD) sources!

Data portal *Grensdata* for NL Border regions

Grensdata

Grensdata is een databank met cijfers over arbeidsmarkt, economie en samenleving voor de grensregio's van Nederland met de Duitse deelstaten Noordrijn-Westfalen, Nedersaksen en met de Belgische gewesten. U kunt zelf grafieken en tabellen maken, aanpassen en data downloaden.

Interreg

De databank is het resultaat van twee projecten rondom arbeidsmarkt data, die zijn uitgevoerd binnen de Europese INTERREG-programma's voor de grensgebieden DE-NL en VL-NL. Het gaat om arbeidsmarkt in grensregio's DE-NL en Werkinzicht.

Datum	Publicatie
22-11-2019	Grenspendel werknemers; nationaliteit, woonland, werkregio (gemeente)
15-11-2019	Bevolking; leeftijd, geslacht

Borderdata portal

Open data to download

Users vs Producers of cross-border data

Users: Diverse

➢ In needs

➢ Different region levels

Producers: Organised

➢ NSI's within ESS

➢ Eurostat, UNECE

Specific

Policy

Recommen dations

Analysis

Visualisation tools

Open data

General
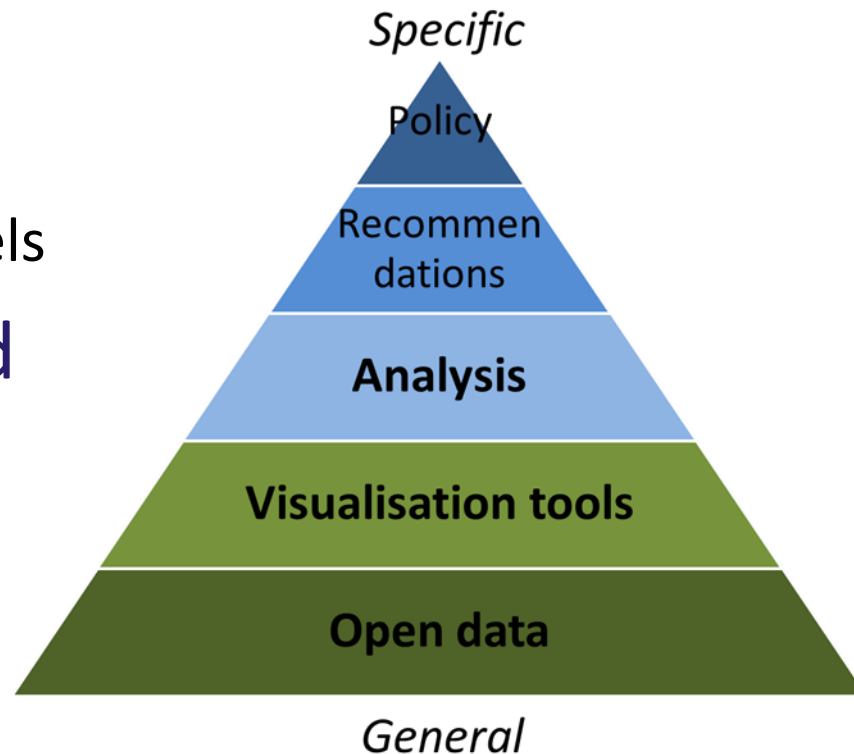
SI's should cover X-border stats

- Methods should be harmonised (inter)nationally
- Is efficient using existing data and expertise
- Equipped for setting up a sustainable system
- Cross-border data is relevant from national perspective
  - Stop depicting a country as an island!
- Fits into mandate of the ESS
  - Cannot rely on Eurostat because of lack of resources and expertise

How should we develop x-border stats?

- Make use of existing data and infrastructures
- Use strategy of constructive collaboration
  - Start with small group of volunteering SI's
  - Cooperating with stakeholders and relevant organisations
  - Aligned with other initiatives
  - Meet to exchange practices

New demands: change the way we work

Change organisation of producing statistics

➢ Production

  ✓ One source per Statistic-> Multiple sources
  ✓ Micro data available -> not available
  ✓ Few Models -> Many Models

➢ Dissemination

  ➢ Tables & publications -> Open data & Tools & Dashboarding
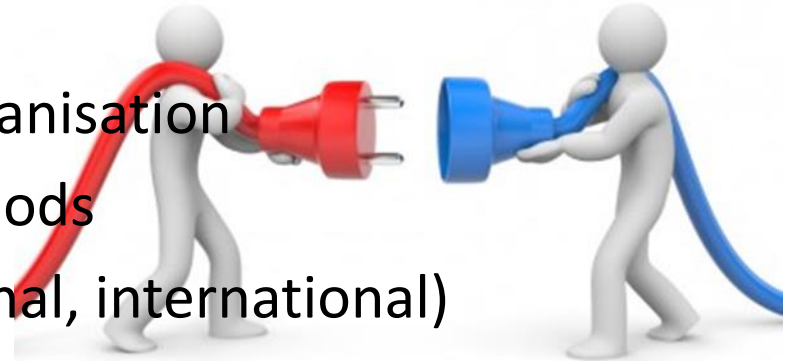
➢ Different roles within ESS

  ✓ ESS: SI's have to take more responsibility
  ✓ Eurostat: Coordinating -> Facilitating

Product of statistics: cooperation

➢ Different forms of cooperation is required with

- ✓ Other SI's, international organisation
- ✓ Academics to develop methods
- ✓ Users (local, regional, national, international)
- ✓ Data suppliers
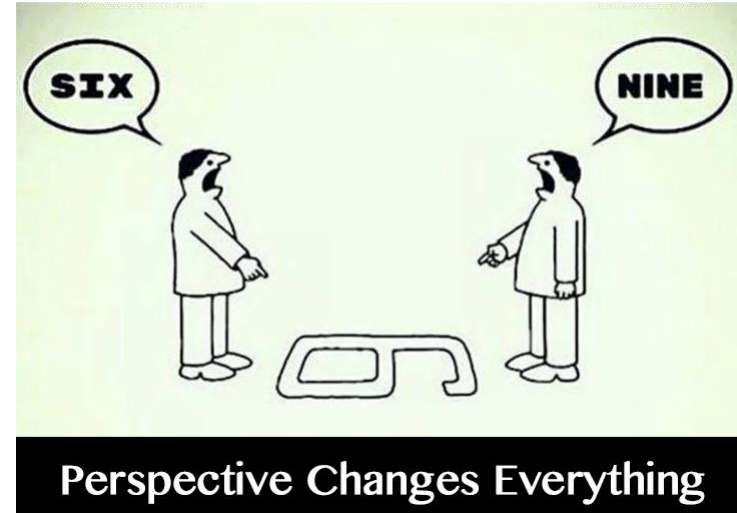- ✓ Private parties for dissemination

Requiring new skills

➢ New hard skills:
- ✓ Data science
- ✓ Visualisation

➢ New soft skills
- ✓ Communication
- ✓ Relations
- ✓ Collaboration rather than cooperation

Perspective Changes Everything

Towards Collaboration

From Cooperation to Collaboration

Cooperation...	Coordination...	Collaboration...
Short term	Longer term	Long term
Informal Relations	More formal relationships	More pervasive relationship
No clearly defined mission	Understand mission	Commitment to a common mission
No defined structure	Focus on a specific effort or program	Results in a new structure
No planning effort	Some Planning	Comprehensive planning
Partners share information about the project at hand	Open communication channels	Well defined communication channels at all levels
Individuals retain authority	Authority still retained by individuals	Collaborative structure determines authority
Resources are maintained separately	Resources and rewards are shared	Resources are shared
No Risk	Power can be an issue	Greater risk: power is an issue
Lower intensity	Some intensity	Higher intensity
***********************************	***********************************	***********************************
--informal, no goals are defined jointly, no planning together, information is shared as needed.	--some planning is required and more communication, thus, a closer working relationship is developed.	-working together, having shared commitment and goals, developed in partnership. Leadership, resources, risk, control and results are shared. More accomplished than could have been individually.