



# Is there a future for statistical surveys?

Surveyföreningens jubileumkonferens  
Statistikfrämjandet

Barteld Braaksma (b.braaksma@cbs.nl)  
Innovation Manager, Statistics Netherlands (CBS)

Stockholm, 16 October 2024



# What is CBS (Statistics Netherlands)?



- Established 1899, 5 employees at Binnenhof
- 1949-1973: 23 locations in The Hague
- 1973-2008: Voorburg
- 2008-now : Leidschenveen
- 1974-now : Heerlen
  
- Since 10-10-10: Bonaire

Two (three) offices- or twenty?



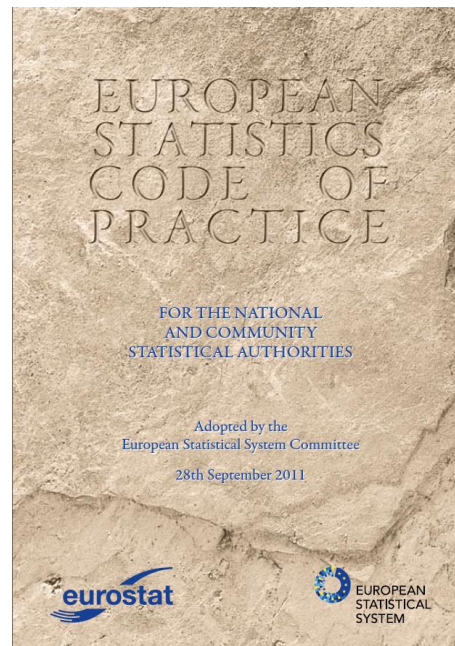
**Centraal Bureau voor de Statistiek**

- 1982: 3600 employees
- 2024: 2000 employees



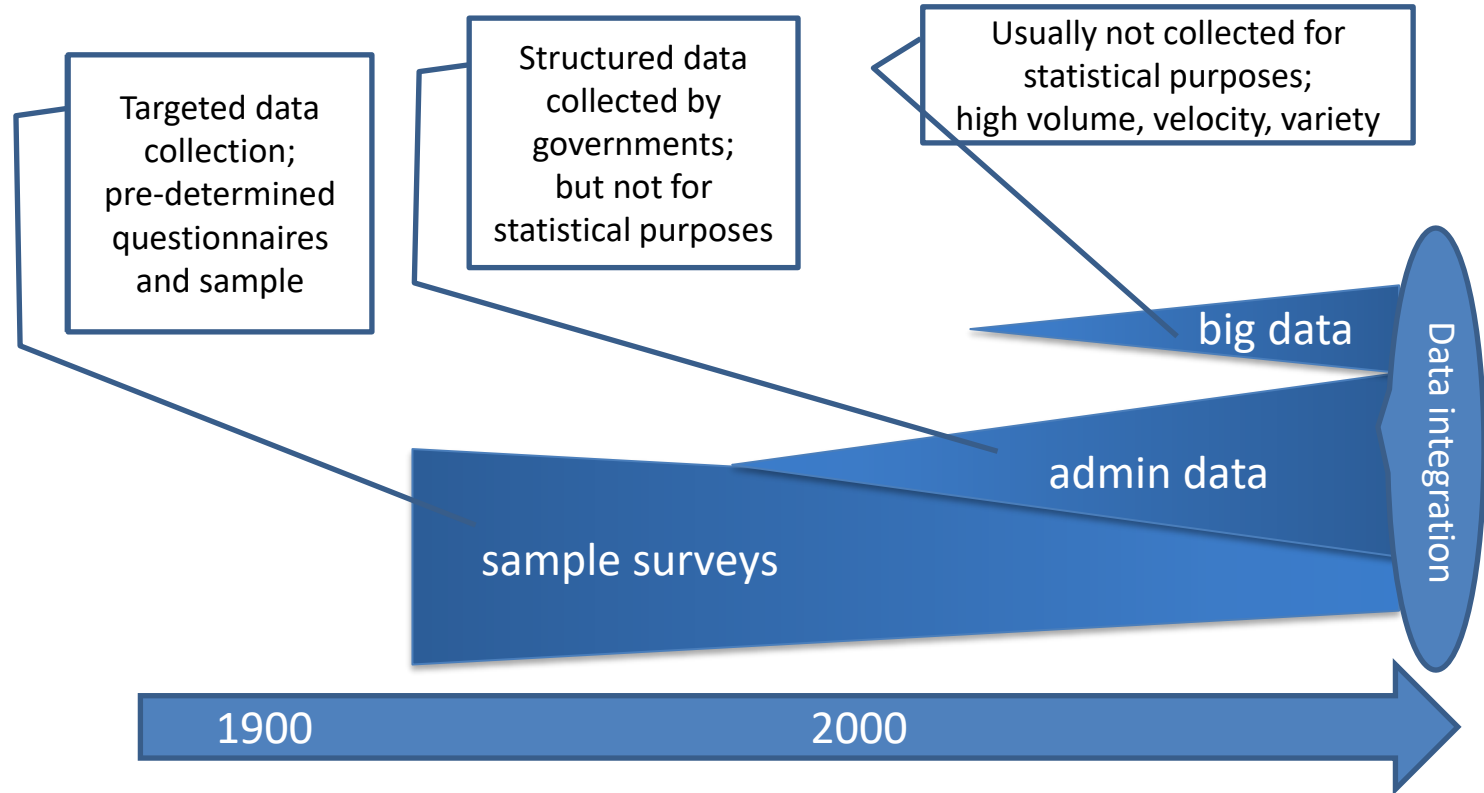
# Mission of the European Statistical System

“We provide the European Union, the world and the public with **independent high quality information** on the economy and society on European, national and regional levels and make the information **available to everyone** for decision-making purposes, research and debate.”



- Presupposes availability of **adequate** data sources (NOT necessarily surveys!)

# From surveys via admin data to big data



# Alternatives to surveys



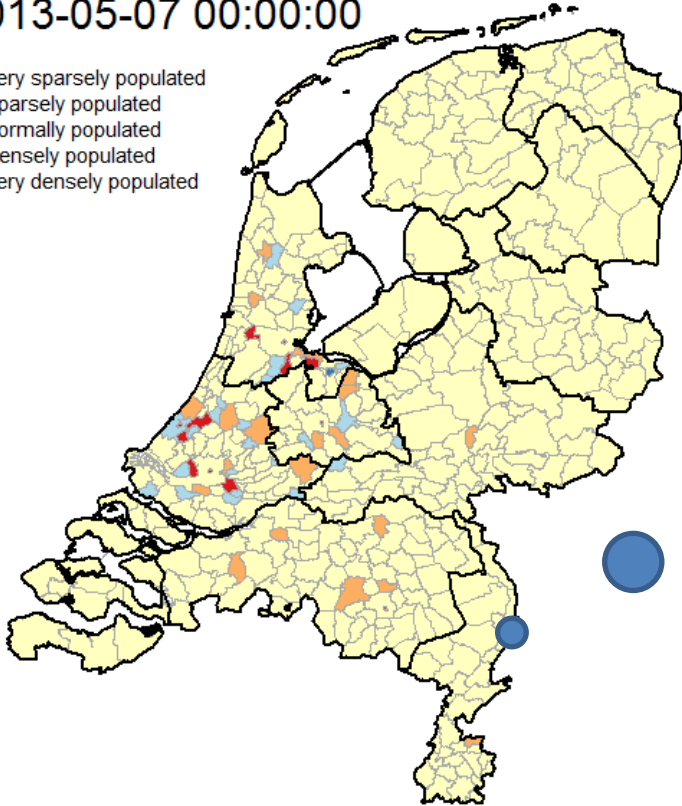
# Access to government registers by CBS



# Where is everybody during a typical Tuesday?

2013-05-07 00:00:00

- Very sparsely populated
- Sparsely populated
- Normally populated
- Densely populated
- Very densely populated



- ❖ Municipality level
- ❖ Red is increasing
- ❖ Blue is decreasing



# Mobile phone use - Ethnographic data & Big Data

DRIVACY FIRST



VACATURE PLA...

Nieuws Vernieuwing Het vak Loopbaan Agenda

Vacatures Stages Opgdrachten I

Trudy Brandenburg - van de Ven — vrijdag 3 juli 2020, 14:48 | 1 reactie, [praat mee](#)

## Autoriteit Persoonsgegevens benadrukt dat wijziging noodwet over delen telecomdata nodig is



Net binnen Uitg

Phishers doen zich v  
redactie LINDA.

Digitale journalistiel  
conserveren kun je l

Milieu-expert gedag  
optreden in uitzendi

Toeziethouder CTIV  
inlichtingendienster  
verzamelen te veel e  
te lang

Rechtszaak bedreigi  
Gargard: boetes en





**Many issues with surveys...**



# Are surveys still feasible?

- PAPI -> CAPI -> CATI -> CAWI -> CAMI
  - Tendency to cheaper modes
  - Decreasing response rates
  - **How to find respondents?**
  - How to motivate respondents?
  - CATI under pressure
- 
- Data collection and processing takes time
  - **Timeliness becomes more important than accuracy**
  - Less costly alternatives
  - 'Race to the bottom'



# Decreasing appetite to contribute

Menu **nrc**

## Ophef over het CBS-diversiteitsproject dat kansengelijkheid op de arbeidsmarkt moet vergroten

**Kansenongelijkheid arbeidsmarkt** De Universiteit van Utrecht schort deelname aan het project op nadat personeelsleden in de universiteitsraad zorgen hadden geuit. „Het lijkt me niet goed mensen in te delen in westers of niet-westers.“

Wilmer Heck 14 april 2021 om 22:51 Leestijd 3 minuten



Goal: describe diversity of employees of Utrecht University (reduce discrimination)

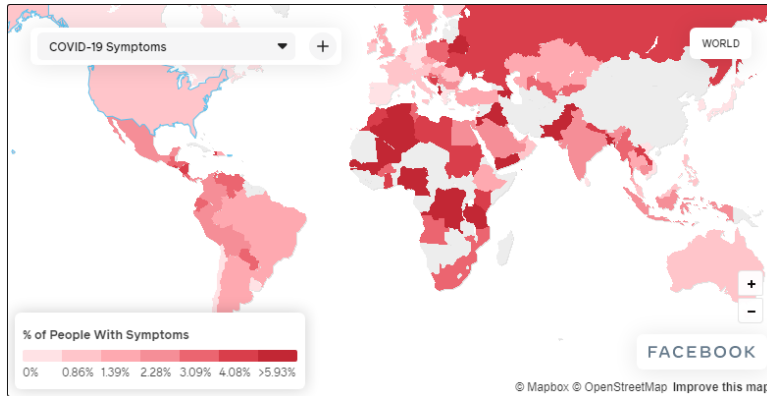
Concerns about CBS-project that aims to increase equality on labour market

‘It is not OK to divide people in ethnic groups’



# Big Tech is taking over

## Facebook COVID19 symptoms survey



<https://dataforgood.fb.com/docs/covid19/>

### Publicly Available Aggregated Data

- [Aggregated US survey data](#) are made available to the public by Carnegie Mellon University's Delphi Research Group.
- [Aggregated non-US survey data](#) are made available to the public by University of Maryland.

### Request Access to Non-Aggregated Data for Research Purposes

Could we collaborate with them?

# AI is taking over

Nobel Prize 2024 for Physics for neural network research

Nobel Prize 2024 for Chemistry for 3D molecule structures using AI

~~Nobel Prize 2024 for Literature for a novel written by ChatGPT~~

Can we use AI with surveys?

- Input: Explain questions
- Throughput: AI-enhanced processing
- Output: Find results (StatGPT)

Surveys on AI use and perception

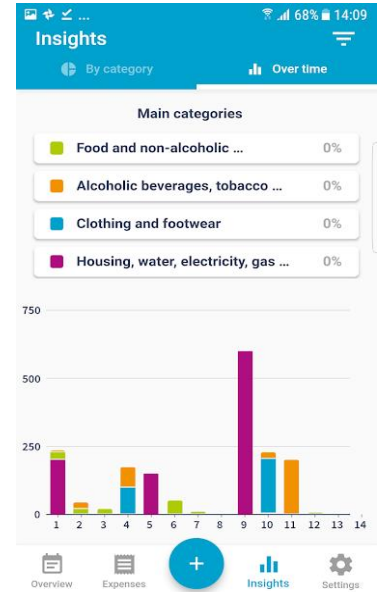
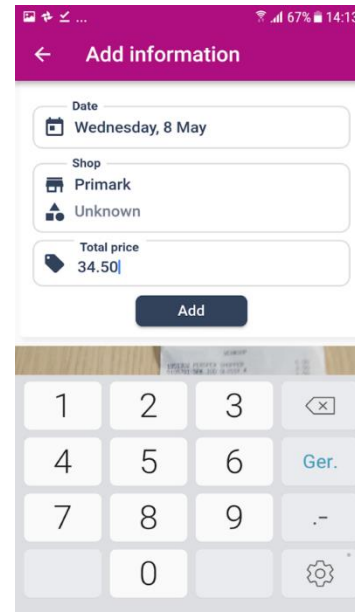
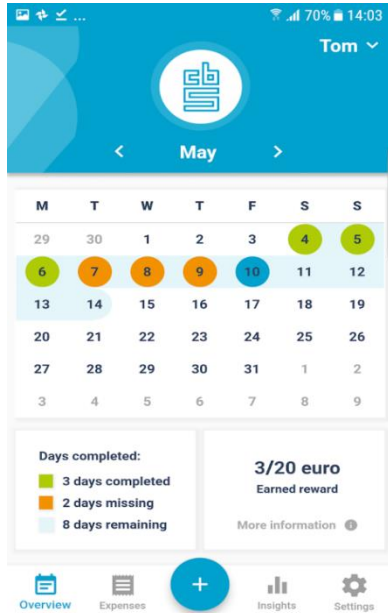


"IT FIGURES. IF THERE'S ARTIFICIAL INTELLIGENCE, THERE'S BOUND TO BE SOME ARTIFICIAL STUPIDITY."

# Many opportunities for surveys



# Going Mobile



Office for National Statistics



Statistics Finland



University of Essex



Household Budget Survey (experimental)

@HBS >An app-assisted approach for the Household Budget Survey

'Trusted Smart Surveys' (Eurostat)

# Collaborations and citizen engagement



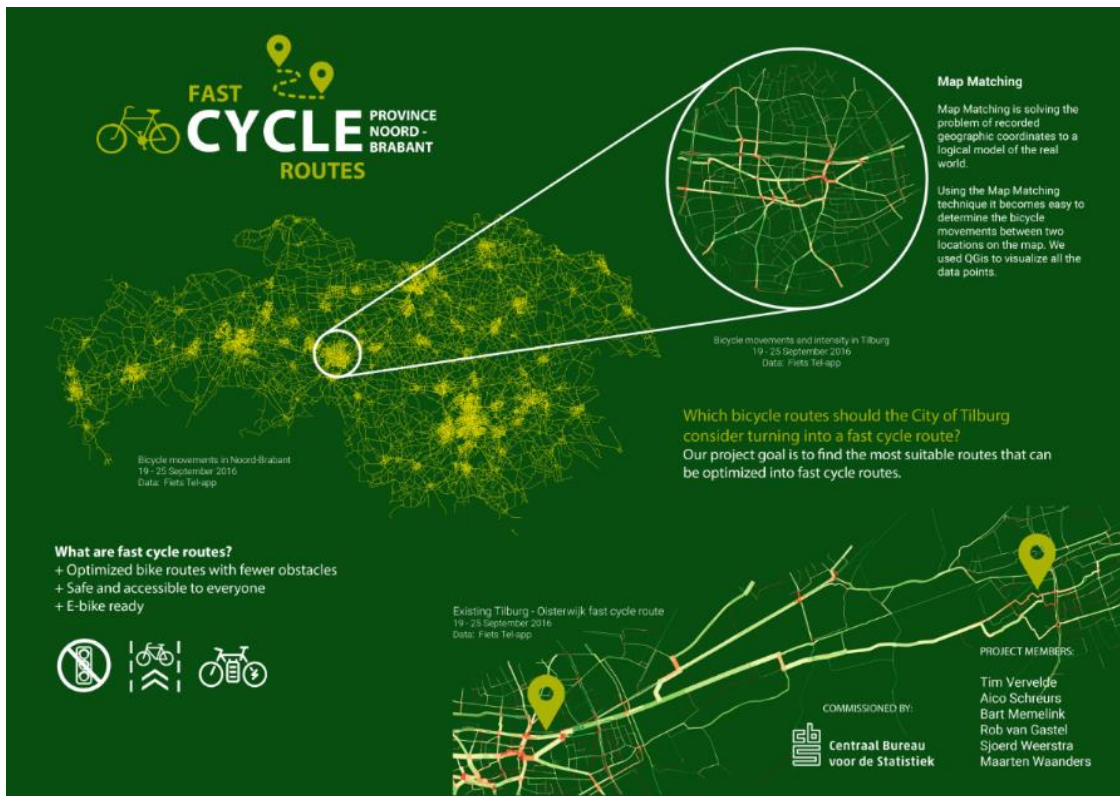
National Bicycle Counting Week using an app (40 thousand participants)

Can we turn this into statistics?





# Yes we can!



# New roles for surveys?

Rapid Survey Systems (apps, sensors, ...)

Quality control for other data sources

Training data for machine learning algorithms

Focus on what secondary sources cannot do  
(attitudes, perceptions and sentiments)

Combine quantitative and qualitative analysis  
(follow up survey with interview)



# Embrace advanced methods

- Non-probability sampling
  - Small area estimates
  - Nowcasting
  - Integration techniques
  - Machine learning
- ...All possible within the survey realm

# What is acceptable?



Organisation for Economic Co-operation and Development

For Official Use

SDD/CSSP(2024)5

English - Or. English

23 April 2024

STATISTICS AND DATA DIRECTORATE  
COMMITTEE ON STATISTICS AND STATISTICAL POLICY

**Imputations, assumptions, and models - how far and by whom for official statistics?**



ental

tors

f

es,

l,

elp



# Do it 'the survey way'

- Surveys typically start from a statistical philosophy
  - What do we want to measure? What is the target population? Which concepts to define? How to process? How to characterize quality? ...
- Well-established methods, language and tradition
- Philosophy remains relevant when dealing with other sources, methods etc may change
- How to transpose the philosophy to data science and AI?
  - TSE approach for machine learning (Puts & Daas, under construction)?



# Data integration and multi-source statistics



# Is the time of statistics over?



In a post-truth world, statistics could provide an essential public service  
**John Pullinger** (National Statistician UK)

By combining the best of both worlds

Statistics Netherlands

<https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy>

**Data, data everywhere**

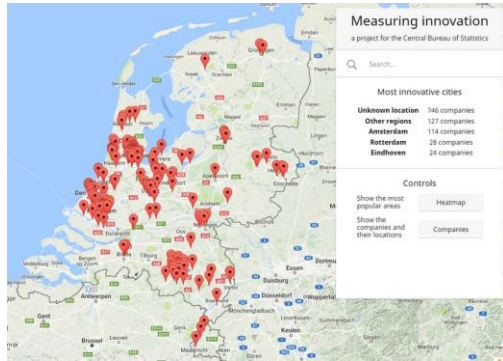
**Information has gone from scarce to superabundant.**

**The Economist**



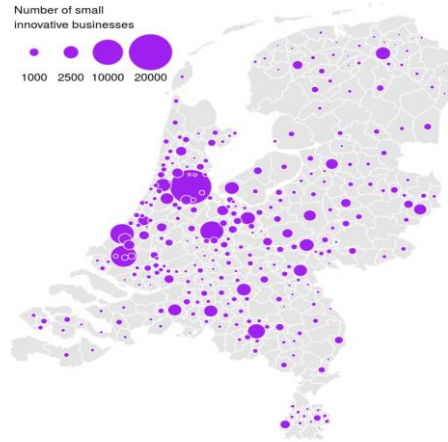
# Measuring innovation via websites

First a students' assignment



Then a beta product:

## Innovation in small businesses



Based on data from 500,000 SMEs  
(CIS: 10,000 non-SMEs)

Now a project for the  
Ministry of Economic Affairs

- Goals
  - ✓ Better view of innovation landscape
  - ✓ Better targeting of subsidies
- Project focused on two pilots
  - ✓ Energy transition-related innovation (North of NL)
  - ✓ University spin-offs (Eindhoven region)

➤ Joint work with  Innovatiespotter

Many spin-offs





# Statistical description of online platforms

Arbeid en inkomen Economie Maatschappij Regio Corporate Cijfers

Deel deze pagina

## CBS brengt online platforms in kaart

19-5-2019 12:00 / Auteur: Maja de Ree

Lees meer over

economische verwachtingen  
innovatie  
internetwaarneming

Meer in nieuws

Regionale verschillen in huizenprijzen groter  
gisteren 00:00

45-plussers wisselen relatief weinig van baan of beroep  
gisteren 00:00

Bijna 5 procent meer investeringen in januari  
22-5-2019 06:30

Paper

## Measuring online platforms

In cooperation with UvA  
UNIVERSITEIT VAN AMSTERDAM

Nico Heerschap (CBS)  
Nicky Psoou (UvA)  
Cybele Alm (UvA)

December 2018

tegenwoordig af op online platforms. Niet alleen op  
ok of Twitter), maar ook bij economische activiteiten.  
en Marktplaats bemiddelen tussen aanbieders en  
ten. Het CBS zet stappen om deze online platforms beter in  
het programma 'Adequaat meten van de economie'.

bemiddelaar tussen een aanbieder en een gebruiker van  
nderzoeker Nico Heerschap uit. 'Op steeds meer gebieden  
e platforms een rol. Dat kunnen aanbieders en gebruikers  
, maaltijden, muziek, games of burenhulp, maar ook  
eeld kennis over innovaties uitwisselen.' Die online  
terrein effect hebben. 'Zo kunnen ze bijdragen aan

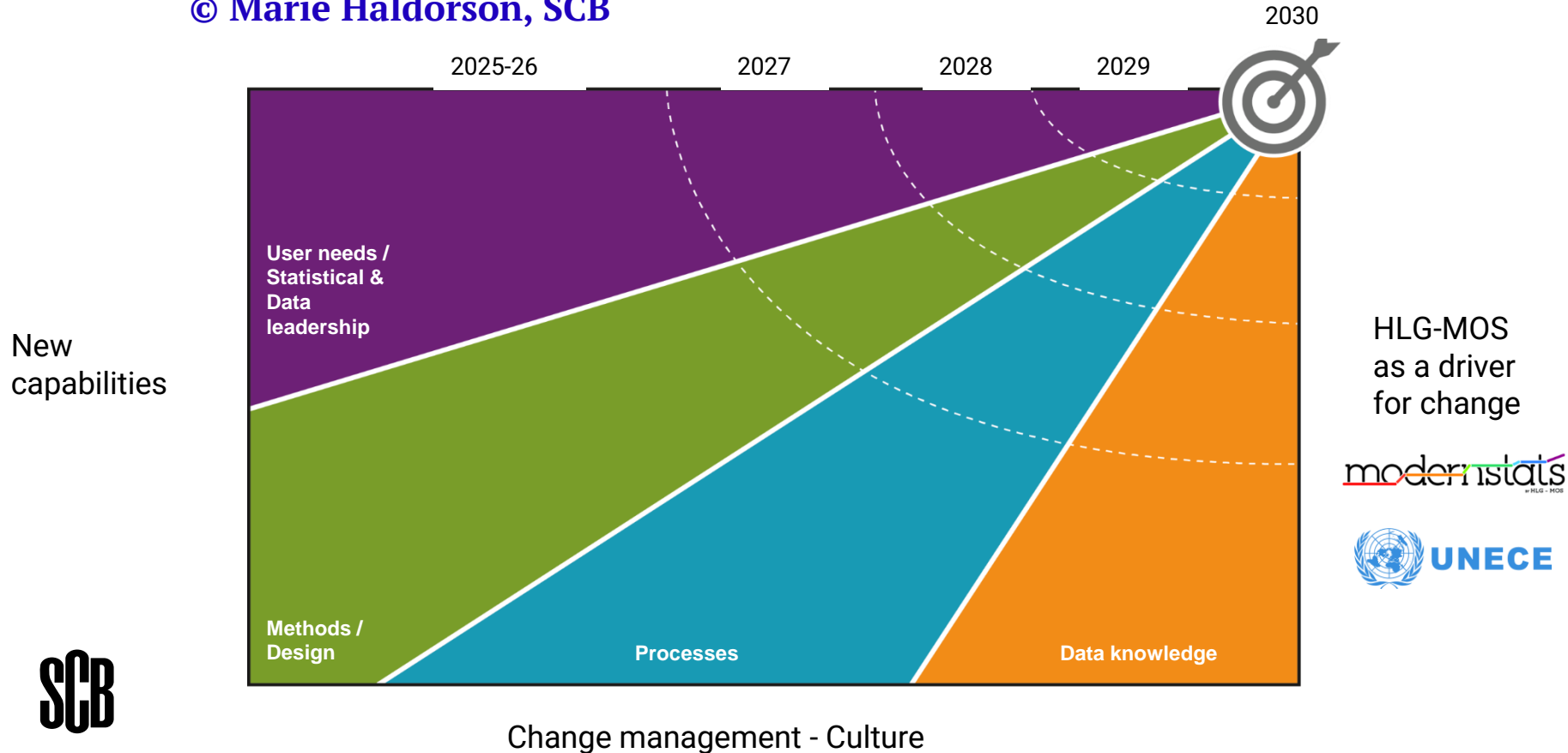
Transportation	Finance	Consumer goods
Space	Personal services	Professional services

- No population frame existed
- Created through analysis of websites
- Then: survey among identified companies



# A roadmap to multi-source statistics

© Marie Haldorson, SCB



# Common European Data Spaces

What about surveys?

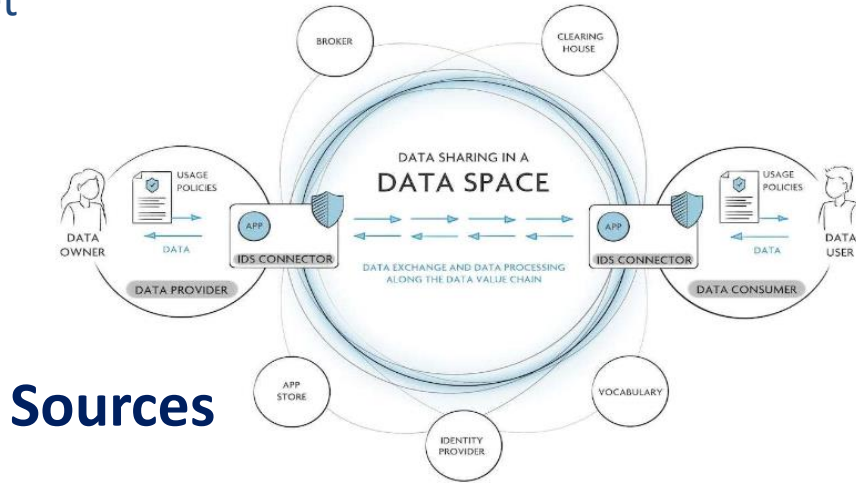


# What is a Data Space?

Interoperability  
Agreements  
'Data Market'

STAKEHOLDER

## Governance



© International Data Spaces

## Infrastructure

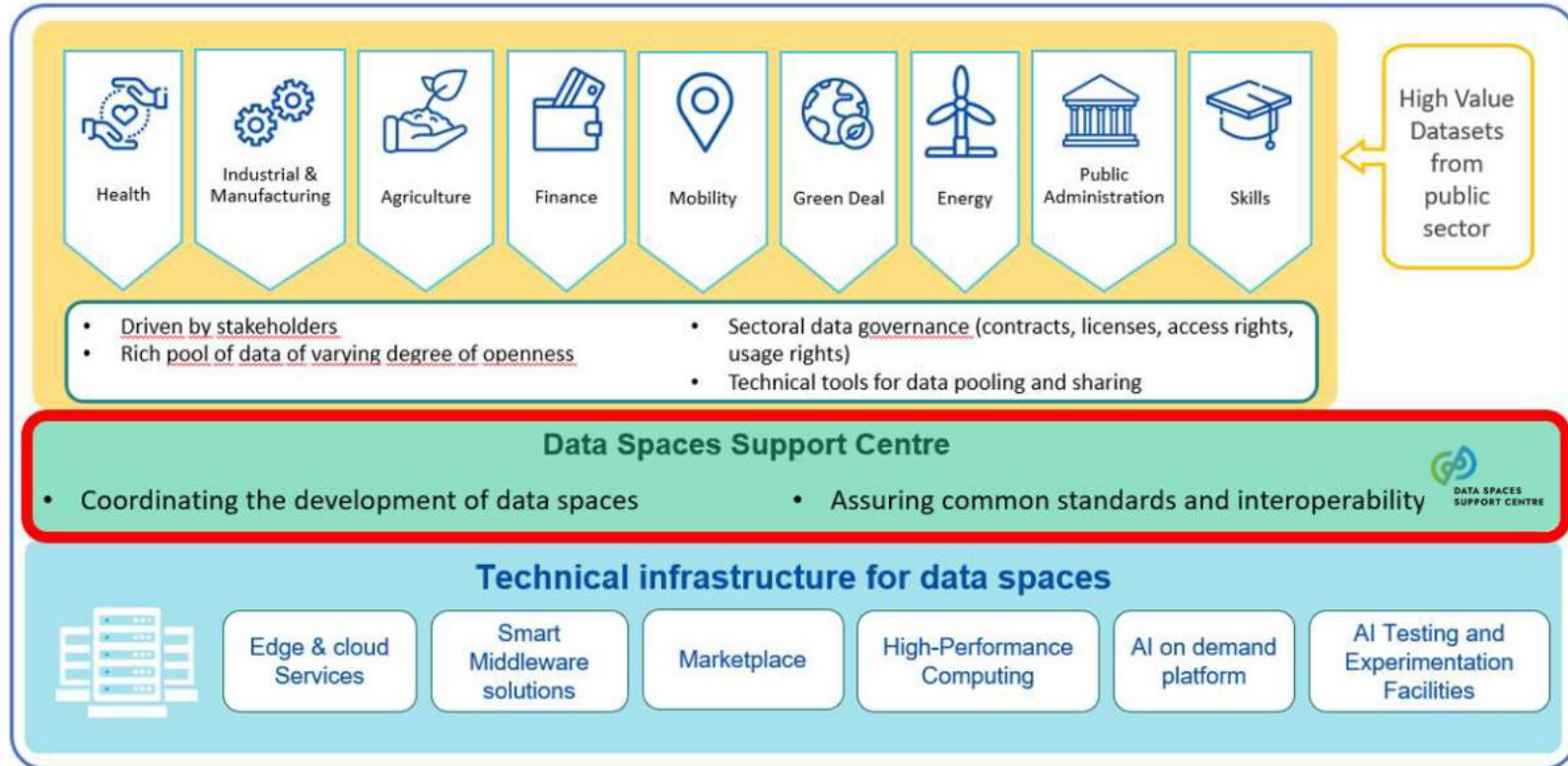
Four-corner principle of data space concept (International Data Spaces)



Users



# Common European Data Spaces



# The European Health Data Space



[Home](#) > [Press](#) > [Press releases](#)

Council of the EU | Press release | 15 March 2024 01:10

## European Health Data Space: Council and Parliament strike deal

*This press release was updated on 22 March 2024 to include the link to the compromise text*

The Council of the EU and the European Parliament have reached a provisional agreement on a new law making it easier to **exchange and access health data** at EU level. The agreement will now need to be endorsed by both the Council and the Parliament.

The proposed regulation for a **European Health Data Space (EHDS)** aims to improve individuals' access to and control over their personal electronic health data, while

## [Provisional agreement on a new law](#)

After months of hard work and dedication, the Council and the Parliament have a deal that will strongly support the development of digital health and scientific research in Europe. The agreement agreed on today will allow for the exchange and access to health data wherever it is held, providing scientific and medical research with the reasons of public interest. This will be a major step towards data that will greatly benefit the development of digital health policies.

— Frank Vandenbroucke, Belgian Deputy Prime-Minister and Minister of Health

## Easier access to health data for individuals

Under the new rules, individuals will have **faster and easier access to electronic health data**, regardless of whether they are in their home country or another member state. They will also have **greater control over how their data is used**. EU countries will be required to set up a **digital health authority** to implement the new provisions.

## Greater research potential





## JRC SCIENCE FOR POLICY REPORT

# EUROPEAN DATA SPACES

Scientific insights into data  
sharing and utilisation at scale

## 2023

Farrell, Eimear; Minghini, Marco;  
Kotsev, Alexander; Soler-Garrido, Josep;  
Tapsall, Brooke; Michei, Marina;  
Pozzida, Monica; Signorelli, Serena;  
Tartaro, Alessio; Bernal, Jaime;  
Vespe, Michele; Di Leo, Margherita;  
Carballa-Smichowski, Bruno;  
Smith, Robijn; Schade, Sverre;  
Pogorzelska, Katarzyna;  
Gabrielli, Lorenzo; De Marchi, Davide

(80 pages)



## HOW CAN STAKEHOLDERS BENEFIT FROM SYNTHETIC DATA IN A DATA SPACE?

TECHNICAL

Synthetic data are data randomly and artificially generated from an original dataset and maintain a certain (given) similarity with it, while not including any of the original records but still preserving analytical value. Synthetic data can help overcome similar challenges in many fields of applications.

In the gap representation scenario, artificial profiles can be generated by exploiting technologies similar to those used by Big Tech companies for tailoring their advertising based on user preferences. Statistical personae are synthesised by crossing census data with additional data sources (e.g., behavioural [surveys](#)), and are used as input in model simulations to gain insights on habits, responses to policy scenarios, mobility patterns, contact tracing in epidemiological emergencies, etc.

## WHICH ACTORS ARE PROVIDING WHAT TYPES OF DATA WITHIN SCOPE OF A DATA SPACE?

ORGANISATIONAL

The agency was either not aware of, or was aware of, yet did not use, additional data existing at the national level, such as, (i) a national university has already researched land use/cover changes in the Member State with a yearly frequency and produced openly-licensed change maps for each two-year period in the last 10 years; (ii) a private company has recently performed a [survey](#) of land use/land cover on a 10 km grid covering the whole country; (iii) an EU agency has released classified land use/land cover maps derived from satellite imagery; (iv) within a national citizen science project, volunteers have recently collected thousands of land use/land cover samples (geo-located photographs and descriptions) using a dedicated app. If these datasets were also used, they could have provided an essential contribution to complement, enrich and validate the government maps as well as to train and apply machine learning models assessing climate changes.



# Synthetic data in official statistics (EU-SILC, 2014)

The screenshot shows the CROS website interface. At the top, there is a search bar and the CROS logo. Below the navigation bar, the breadcrumb trail reads: European Commission > Eurostat > CROS > Centre of Excellence on Statistical Disclosure Control > Public use files for Eurostat microdata > PUF - Public Use Files > Background information for EU-SILC. The main heading is "Background information for EU-SILC".

**Centre of Excellence on Statistical Disclosure Control**

- Public use files for Eurostat microdata
  - Project documentation
  - PUF - Public Use Files
    - Background information for EU-SILC
    - Background information for LFS
      - PUF EU-SILC - the data
      - PUF LFS - the data
- User support for and maintenance of SDC tools
- Harmonised protection of Census data
- Perturbative confidentiality methods

CoE on SDC 2020-2024

**Public microdata for EU-SILC have moved**

The public microdata (Public Use Files) for EU-SILC have now moved from the project environment to the Eurostat production environment: <http://ec.europa.eu/eurostat/web/microdata/public-microdata>.

Gradually more data will become available (more countries and more years). Access no longer requires registration.

**Background information for EU-SILC**

The EU-SILC (EU Statistics on Income and Living Conditions) Public Use Files (PUFs) were created to enable interested parties to become familiar with using of EU-SILC microdata while protecting the privacy of data providers (Principle 5 of the European Statistics Code of Practice). The Public Use Files consist of fully synthetic data simulated using statistical modelling and statistical distributions in the original data. They have specifically been produced for testing purposes and education. The data structure of the EU-SILC PUFs is the same as for the EU-SILC Scientific Use Files (SUFs) although not all variables have been simulated (like Flag variables) and are hence empty in the PUFs. They provide researchers and trainers the opportunity to develop programs using identical formats and variable names to those that appear in the actual EU-SILC SUFs.

The PUFs are tailored to the following purposes:

- to enable accredited researchers to develop and write a code that may eventually be applied to the actual EU-SILC SUFs after the access is granted;
- to train accredited researchers on the use and complexity of conducting analyses with the EU-SILC PUFs prior to initiating the process to obtain access to the actual EU-SILC SUFs;
- for statistical training.

The data cannot be used for any statistical inference to a wider population. With these data it is not possible to make generalised statements about individual characteristics or relationships between different personal or household characteristics. These files have limited research value; only the EU-SILC SUFs provide a basis for valid analysis. Any calculations based on PUFs or the conclusions drawn from them are to be produced at the sole risk and responsibility of the user with no warranty for accuracy or reliability given by the NSI that produced these EU-SILC PUFs or by Eurostat. Furthermore neither the NSI that produced these EU-SILC PUFs nor Eurostat can be held responsible for any improper





# Concluding remarks



# Hal Varian 'on workers and managers'



The McKinsey Quarterly, Januari 2009:

I keep saying the sexy job in the next ten years will be statisticians (...)

Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

I think statisticians are part of it. (...) Managers need to be able to access and understand the data themselves.

- Chief economist at Google
- Emeritus professor at Berkeley

# Is there a future for statistical surveys?

- ❖ Surveys are here to stay
- ❖ But innovation is a must
- ❖ If you can't beat them, join them
- ❖ Non-statistical questions become more important
  - Governance, ethical, legal, perception, alternative facts, ...
- ❖ Yes, there is a future for surveys
- ❖ And it looks different from the past





**Facts that matter**