UNIVERSITY OF LJUBLJANA | SEB School of Economics and Business

EQUIS ACCREDITED | AACSB ACCREDITED | ASSOCIATION AMBA ACCREDITED

Data collection in official statistics

izr. prof. dr. Mojca Bavdaž (mojca.bavdaz@ef.uni-lj.si)

New Developments in Statistics, 14 May 2025

FDV, 17:15 – 19:45

Outline

Icons by Creative Stall, Johan H. W. Basberg , Noura Mbarki & Gung Yoga from Noun Project

UNIVERSITY OF LJUBLJANA | SEB School of Economics and Business

Institutional framework of data collection

Official statistics

- What is official statistics?
- Who is responsible for official statistics in Slovenia, Europe and elsewhere?
- What defines normative framework of official statistics?
- Who gets access to microdata?

A definition

Official statistics = statistics disseminated by a national statistical institute or body, except for those that are explicitly stated as not official.

These statistics include vital economic and societal indicators.

OECD. (2004). Official statistics. *Glossary of Statistical Terms.*

Classification of Statistical Activities

- Domain 1: Demographic and social statistics
- Domain 2: Economic statistics
- Domain 3: Environment and multi-domain statistics
- Domain 4: Methodology of data collection, processing, dissemination and analysis
- Domain 5: Strategic and managerial issues of official statistics

Classification of Statistical Activities (CSA Rev. 1 - October 2009). Retrieved from http://ec.europa.eu/eurostat/ramon/other_documents/csa/csa_rev_1_october_2009.pdf

Not only data:

+ institutions + statistical activity + scientific field

Official statistics

- What is official statistics?
- Who is responsible for official statistics in Slovenia, Europe and elsewhere?
- What defines normative framework of official statistics?
- Who gets access to microdata?

Institutions of official statistics in Slovenia

Main body and co-ordinator: SURS

Other national authorities:

- National Institute of Public Health (NIJZ)
- Bank of Slovenia (BS)

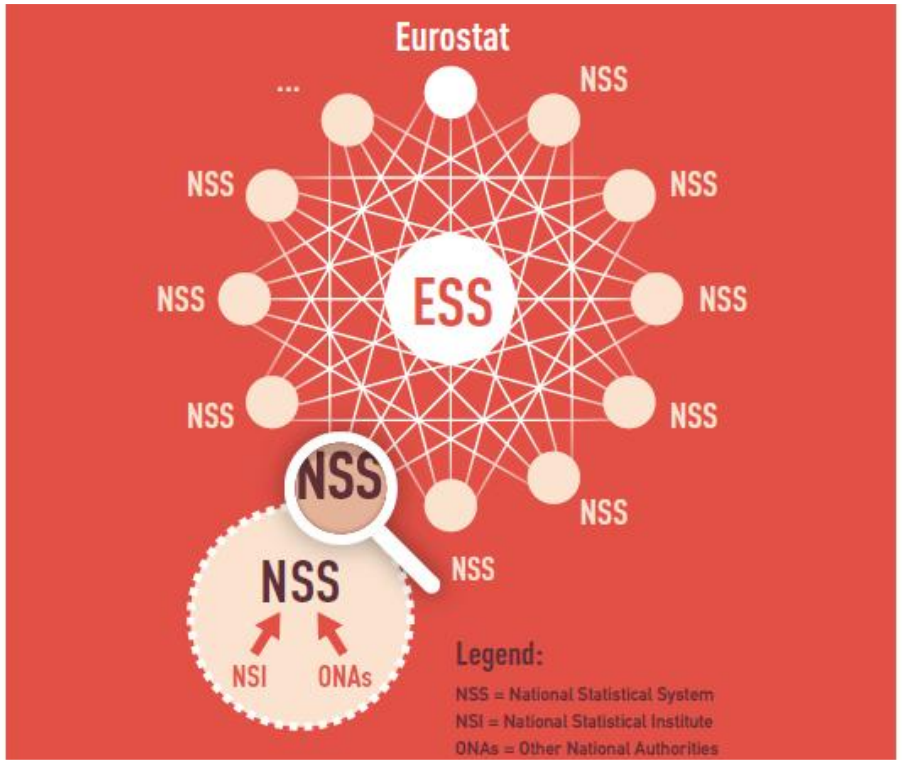
****

Other important institutions:

Ministry of finance, Agency of the Republic of Slovenia for Public Legal Records and Related Services (AJPES), Financial Administration (FURS), Pension and Disability Insurance Institute of Slovenia (ZPIZ), Employment Service of Slovenia (ZRSZ),...

The European Statistical System

- ✓Partnership between:
  - Eurostat (ec.europa.eu/eurostat);
  - national statistical institutes (NSIs) of the member states and their other national authorities (ONAs);
  - including the EEA and EFTA countries (Iceland, Lichtenstein, Norway; CH). ***UK?***
- ✓ Defined in European Statistics Regulation 223/2009.
- ✓ Coordination of work with:
  - candidate countries;
  - ECB;
  - the Commission services;
  - international organisations.



Source: The ESS Report (2014)

9 October 2024

"This arrangement is built on a recognition that both the ONS and Eurostat will benefit by cooperating with data in areas such as GDP, National Accounts, Trade, and Foreign Investment."

"The UK Statistics Authority (UKSA) will meet regularly with Eurostat to discuss where additional collaboration could be beneficial. While the initial focus is on the transfer of GDP data we remain open to exploring other areas of statistical cooperation."

NSIs and ONAs

NL - Netherlands

NSI: Centraal Bureau voor de Statistiek

Other national statistical authorities:

-

SI - Slovenia

NSI: Statistical Office of the Republic of Slovenia

Other national statistical authorities:

1. National Institute of Public Health

SE - Sweden

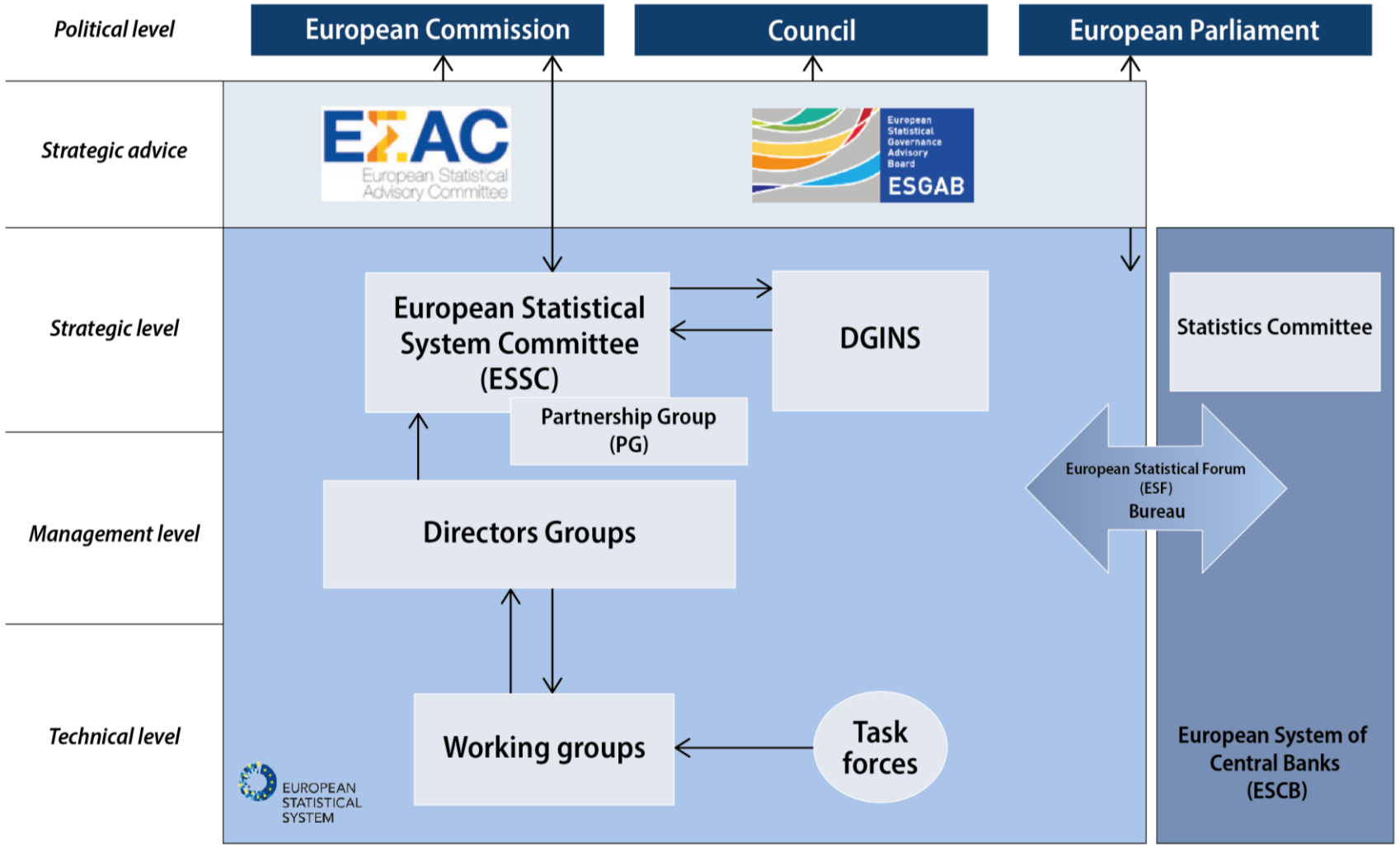
NSI: Statistics Sweden

Other national statistical authorities:

National Board of Health and Welfare
Swedish National Council for Crime Prevention
Swedish National Mediation Office
Swedish Agency for Economic and Regional Growth
Swedish Agency for Growth Policy Analysis
Swedish Agency for Marine and Water Management
National Board of Agriculture
Swedish Chemicals Agency
Swedish Energy Agency
Swedish Environmental Protection Agency
Swedish Financial Supervisory Authority
Swedish Forest Agency
Swedish National Agency for Education
Swedish Higher Education Authority
Swedish National Financial Management Authority
Swedish Work Environment Authority
Transport Analysis
Public Health Agency of Sweden
Swedish Migration Agency
Swedish Police

European Statistical System (18.11.2024). List of National statistical institutes (NSI) and other national authorities. Dostopno na https://ec.europa.eu/eurostat/web/european-statistical-system

ESS governance structure



Link to the European System of Central Banks (ESCB)

https://ec.europa.eu/eurostat/web/european-statistical-system/governance-bodies

International statistical institutions

- ✓ Eurostat
- ✓ OECD
- ✓ UN
  - UN (Statistics Division)
  - World Bank, IMF
  - ILO
  - WHO
  - UNECE
- ✓ WTO

…

Source: United Nations Economic Commission for Europe. Links to official statistical organizations active in the UNECE region. Available at http://www.unece.org/stats/links.html

Official statistics

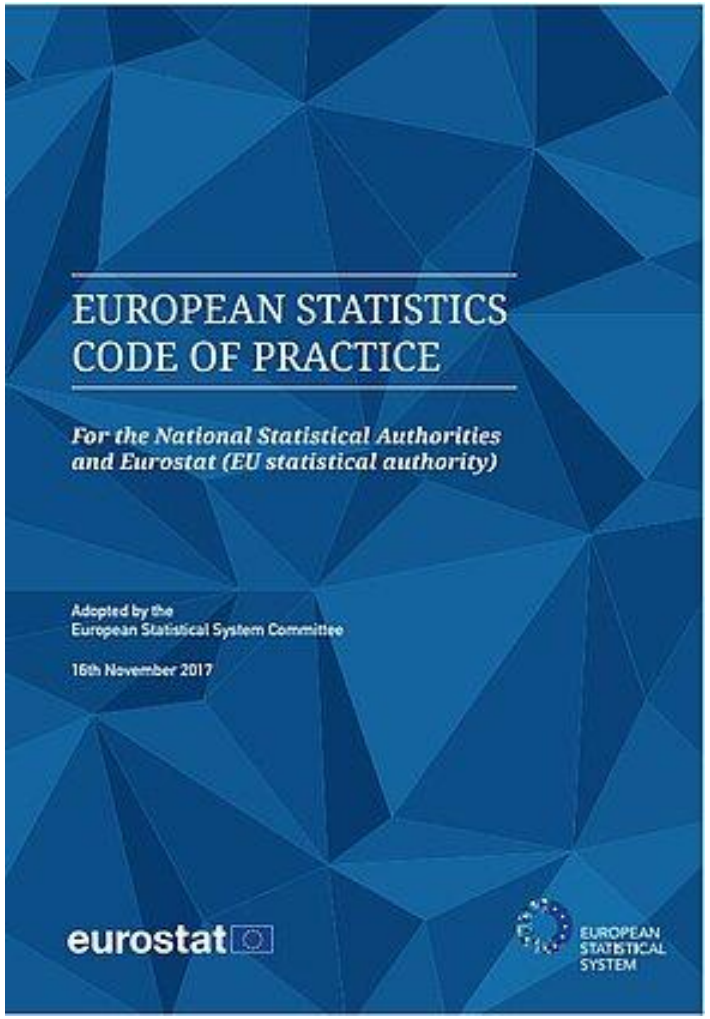
- What is official statistics?
- Who is responsible for official statistics in Slovenia, Europe and elsewhere?
- What defines normative framework of official statistics?
- Who gets access to microdata?

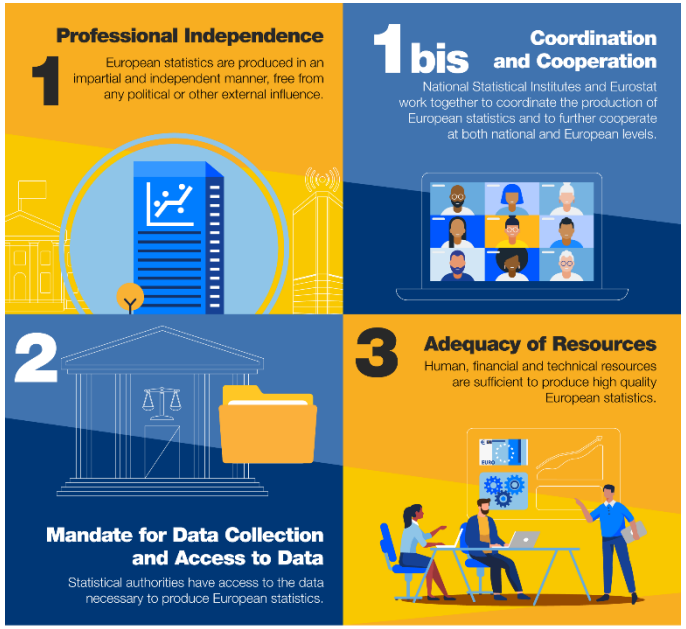
Some milestones

- 1994: *UN Fundamental Principles of Official Statistics* based on text developed by the Conference of European Statisticians
- 1995: as a response to the Mexican financial crisis (partly due to unreliable statistics), IMF launched the initiative *Special Data Dissemination Standard* to oblige national producers of economic statistics to improve the methodological information and respect the fundamental principles (see http://dsbb.imf.org/)
- 2005 European Statistics Code of Practice, in response to the Greek unreliable/falsified data.

GI – GO

Source: OECD (2008). Understanding economic statistics. Retrieved from https://www.oecd.org/std/41746710.pdf





ec.europa.eu/**eurostat**

ec.europa.eu/**eurostat**

ec.europa.eu/**eurostat**

ec.europa.eu/**eurostat**

UL | SEB

Normative framework in the EU

- European Statistics Regulation 223/2009

- Multi-annual programme (2021 - 2027)

- Annual programmes

+ Fundamental Principles of Official Statistics (UN)
+ European Statistics Code of Practice (ESS)

Source: European Statistical System. European Statistical Programme. Retrieved from http://ec.europa.eu/eurostat/web/ess/about-us/statistical-programmes

Official statistics

- What is official statistics?
- Who is responsible for official statistics in Slovenia, Europe and elsewhere?
- What defines normative framework of official statistics?
- Who gets access to data collected in the ESS (microdata)?

Access to microdata

Granted to:

- an employee of the research entity (or be working for them as a contractor, only in justified cases) or
- senior (Ph.D.) students under guidance of a supervisor employed by the research entity; supervisor must be identified in the research proposal as a principal researcher and a senior student as an individual researcher.

At SURS: https://www.stat.si/StatWeb/en/StaticPages/Index/For-Researchers

At Eurostat: https://ec.europa.eu/eurostat/web/microdata

Access to microdata at Eurostat

- o Adult Education Survey (AES)
- o Community Innovation Survey (CIS)
- o Community Statistics on Information Society (CSIS)
- o Continuing Vocational Training Survey (CVTS)
- o European Community Household Panel (ECHP)
- o European Health Interview Survey (EHIS)
- o European Road Freight Transport Survey (ERFT)
- o EU Labour Force Survey (LFS)
- o EU Statistics on Income and Living Conditions (EU-SILC)
- o Farm Structure Survey
- o Harmonised European Time Use Survey (HETUS)
- o Household Budget Survey (HBS)
- o Structure of Earnings Survey (SES)

Access to microdata is ***in principle*** limited to data collected by NSI (owner) but some exceptions are possible.

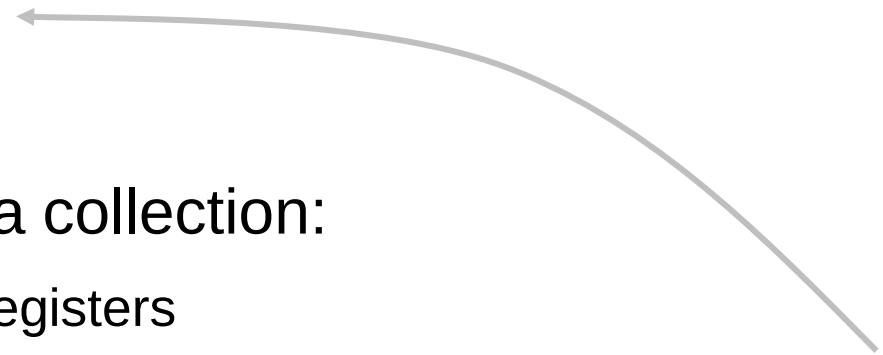
Source: Eurostat (n.d.). Access to microdata. Available at https://ec.europa.eu/eurostat/web/microdata/overview

UNIVERSITY OF L]UBL]ANA | SEB School of Economics and Business

Methods of data collection

Data collection methods for official statistics

- Primary data collection (directly from units of observation or their reporting units):
  - Censuses
  - Surveys
  - Web scraping
- Secondary data collection:
  - Admin data, registers
  - Big data, organic data, privately held data, web scraping



UNIVERSITY OF LJUBLJANA | SEB School of Economics and Business

Surveys

General trends in survey research

- On paper with interviewer ⇒ Electronic self-administration
- Fixed technology ⇒ Mobile technology
- One-off measurement ⇒ Longitudinal measurement
- Data ⇒ Metadata and paradata
- Single survey mode ⇒ Mixed modes

Do these trends apply to official statistics as well?

Types of surveys: By institution that conducts the survey

- Governmental surveys.
- Academic surveys.
- Commercial surveys.

⇒ Enormous differences in response rates!

Types of surveys: By observational units

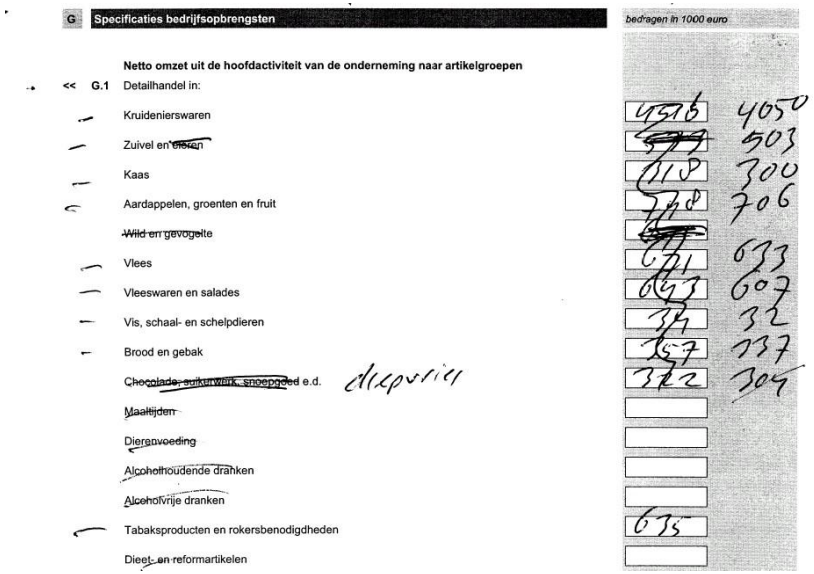
- Household surveys, surveys of individuals.

- Business / establishment / organizational surveys.

Specifics of business surveys

Common methodological bases of household and business surveys.
Specifics of business surveys:

- Differential importance of units.
- Identification and volatility of units of observation.
- Data availability essential (what, when, where, who)
- Questionnaires:
  - About organization, not individual
  - Terminology
  - Forms with item labels and detailed instructions
  - Often numerical, continuous variables
  - Self-administered mode due to data retrieval
- Several participants in the response process.
- Several sources of survey errors but a different meaning and importance of some survey errors.



Snijkers, G., Bavdaž, M. (2011). Business surveys. In: Lovrić, M. (ed.). International encyclopedia of statistical science. Berlin; Heidelberg: Springer, pp. 191-194.

Types of surveys:
By survey mode

- Survey mode = a set of data collection procedures that determine the basic principles of communication and information transmission between the respondent and the survey questionnaire.

- Basic survey modes:
  - Face-to-face survey
  - Telephone survey
  - Mail survey
  - Web survey

- Key differences:
  - Type of contact with the respondent
  - Involvement of interviewer
  - Degree of computer assistance

Mixed modes (1)

- Parallel/contemporary design: different survey modes for different population groups.

  PANDA (NIJZ)

  Web survey for people aged up to 44
  Mail survey for people aged 45 +

- Sequential design: start with the cheapest mode, more expensive modes for nonrespondents.

  Survey on ICT use in households (SURS)
  Web survey ⇒ Face-to-face survey
  Web survey ⇒ CATI (covid-19 epidemics)

Mixed modes (2)

- Personal mode with an interviewer for the first completion of the questionnaire, cheaper modes for repetitions.

  Household budget survey (SURS)

  Face-to-face survey (CAPI) (about household and members)

  Self-administration of diaries on expenditure (14 days)

  Labor force survey (SURS)

  Wave 1: Face-to-face survey (CAPI)
  Wave 2: CATI if the number obtained in Wave 1; otherwise F2F again

  During Covid-19: matching population register data with publicly available telephone numbers already for Wave 1

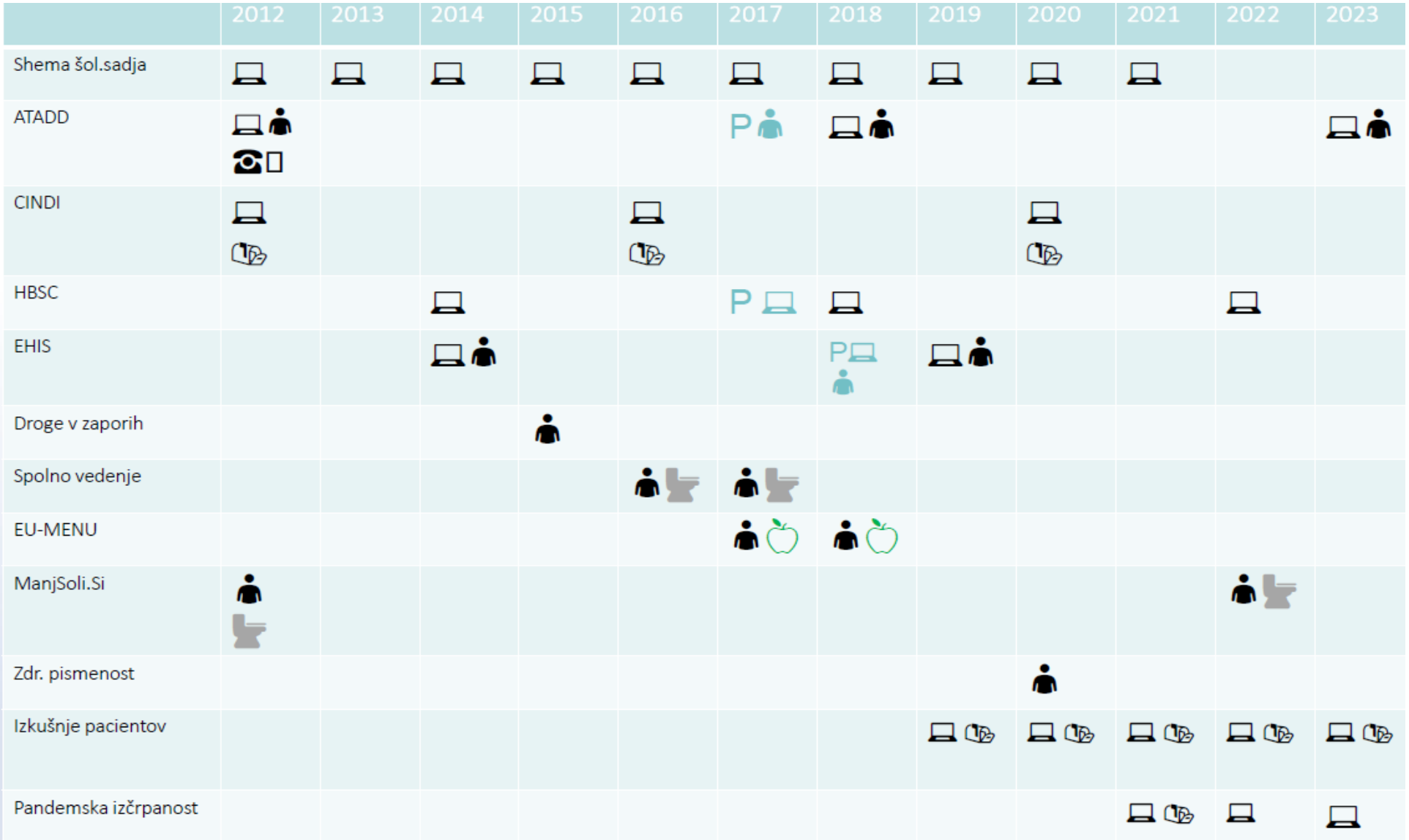
- Parallel independent samples, each with own survey mode.

  Daily mobility survey (SURS)

  Sample A: Face-to-face survey (CAPI)

  Sample B: Web survey

Overview of NIJZ surveys @Metka Zaletel, NIJZ



Survey from quality perspective

Error for any variable consists of sampling and nonsampling error.

Sampling error can easily be quantified if probability sampling is used, nonsampling error is much more difficult to quantify or not quantifiable.

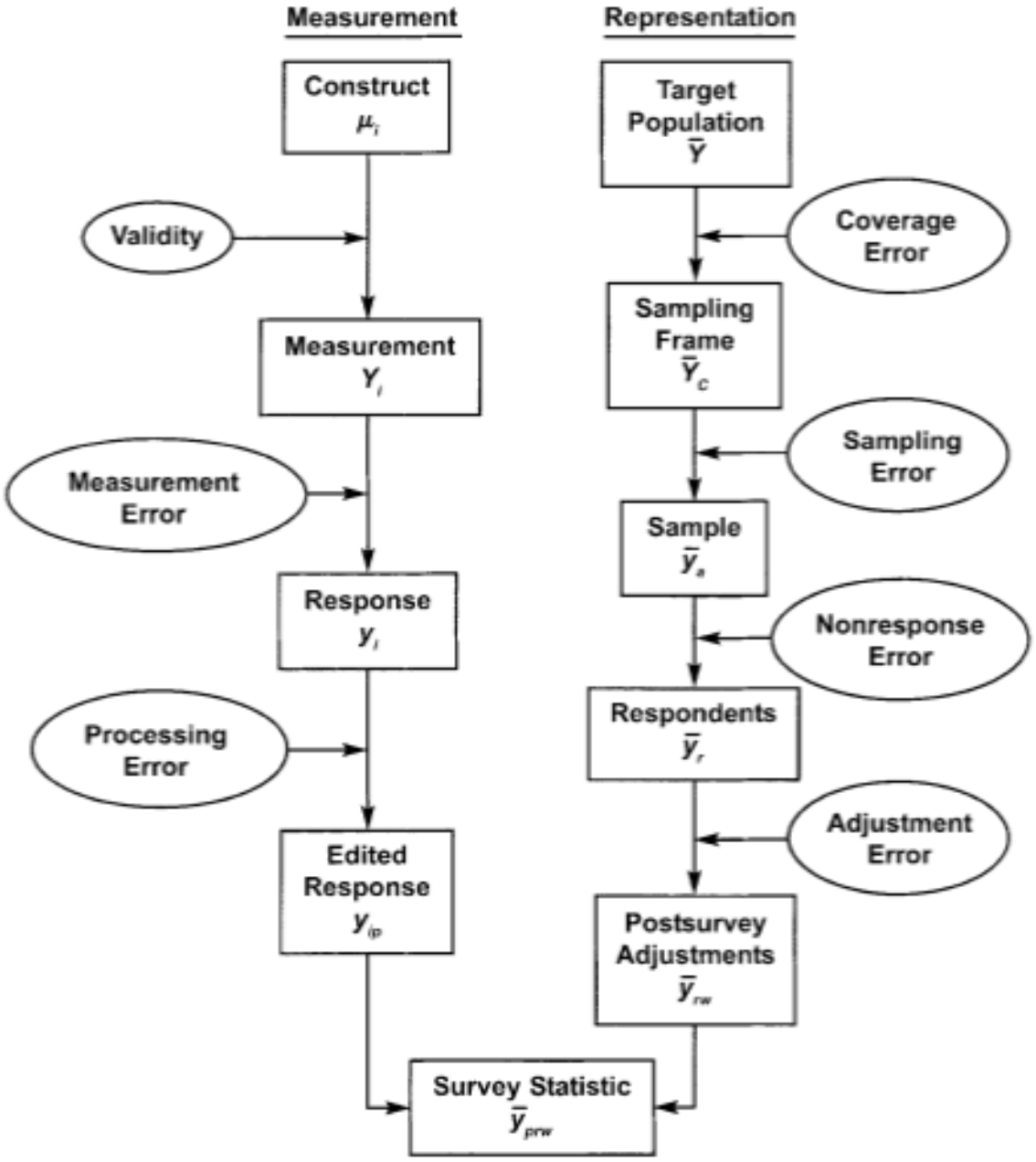


Figure 2.5 Survey life cycle from a quality perspective.

Groves et al. (2004). *Survey Methodology.* Hoboken: Wiley.

Sampling error

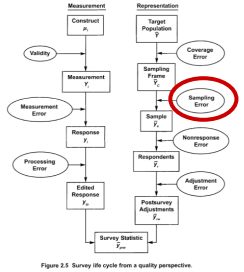


Figure 2.5 Survey life cycle from a quality perspective.

- Condition for drawing a probability sample: existence of a sampling frame or other procedures that ensure the probability of selection (e.g. area sampling).

- Types of probability sampling: simple random sampling, systematic sampling, stratified sampling, cluster sampling, multi-stage sampling

- Typical sampling designs for business surveys:
  - stratified samples
  - cut-off samples
  - a combination: Take all + Take some + Take none

  - *Or: adjusted definitions of population covered!*

- Typical sampling designs for household surveys: multi-stage sampling

Representation side less problematic in official statistics than elsewhere

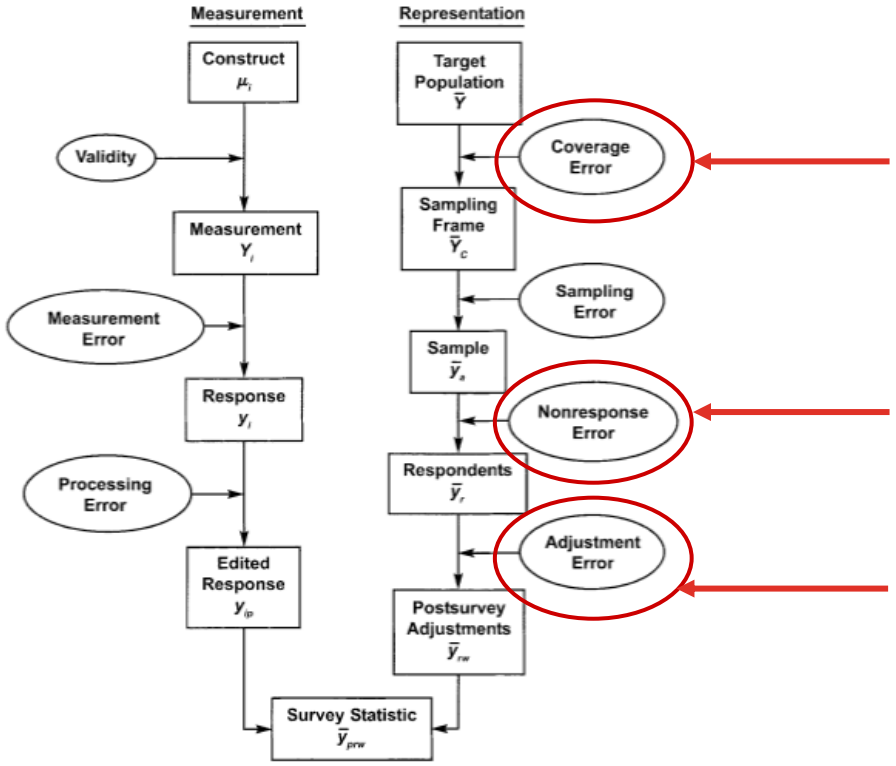
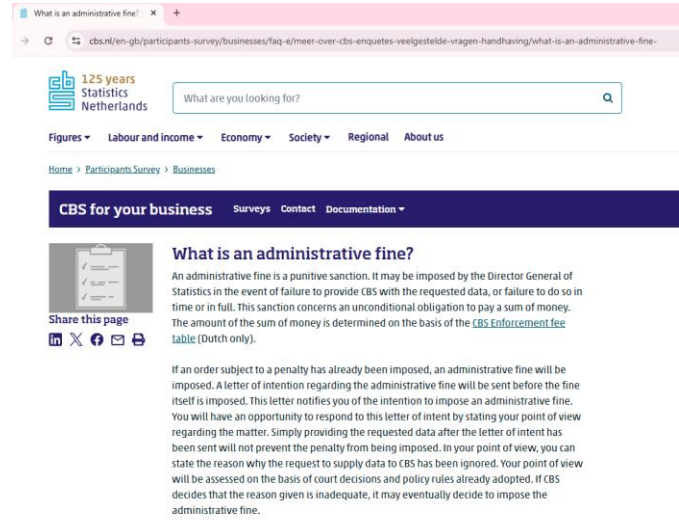


Figure 2.5 Survey life cycle from a quality perspective.

Access to business and population registers
Delineation of units

Mandatory
High reputation

Good input for weighting and calibration



https://zoek.officielebekendmakingen.nl/stcrt-2024-11231.pdf

Groves et al. (2004). *Survey Methodology.* Hoboken: Wiley.

Measurement side

International standards
Master questionnaires

Testing question(naire)s

Good input for checks

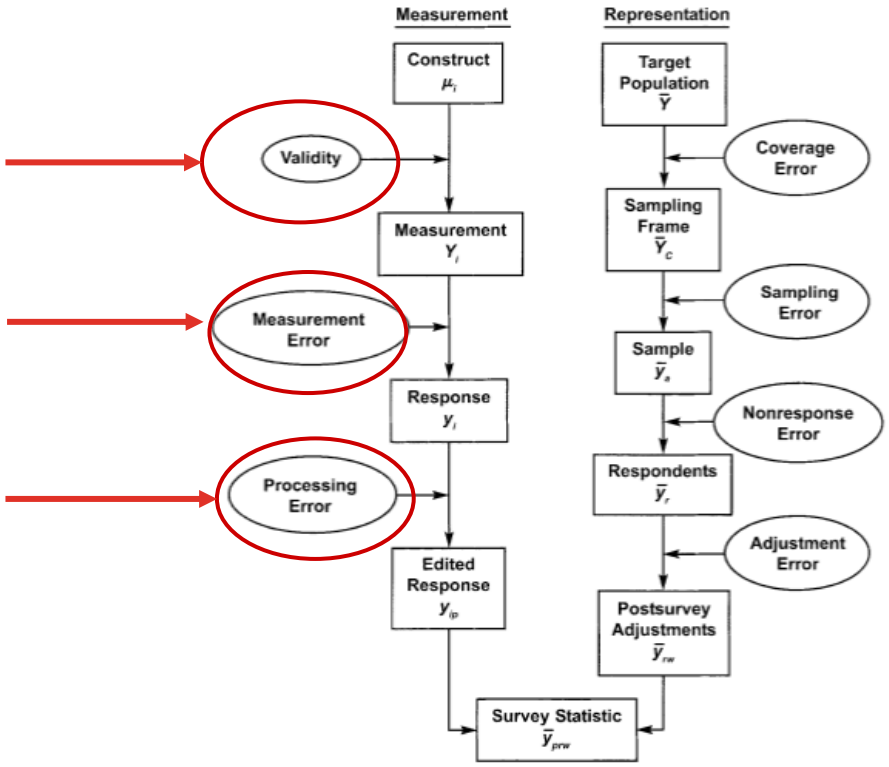


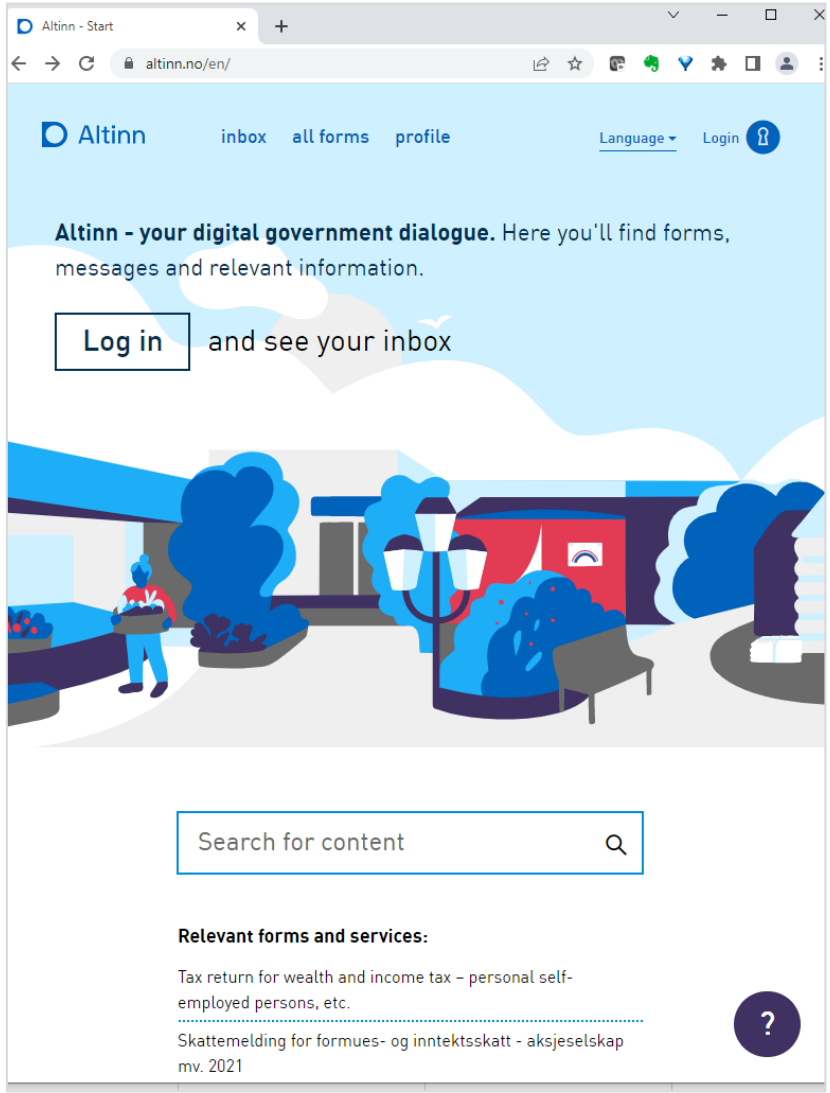
Figure 2.5 Survey life cycle from a quality perspective.

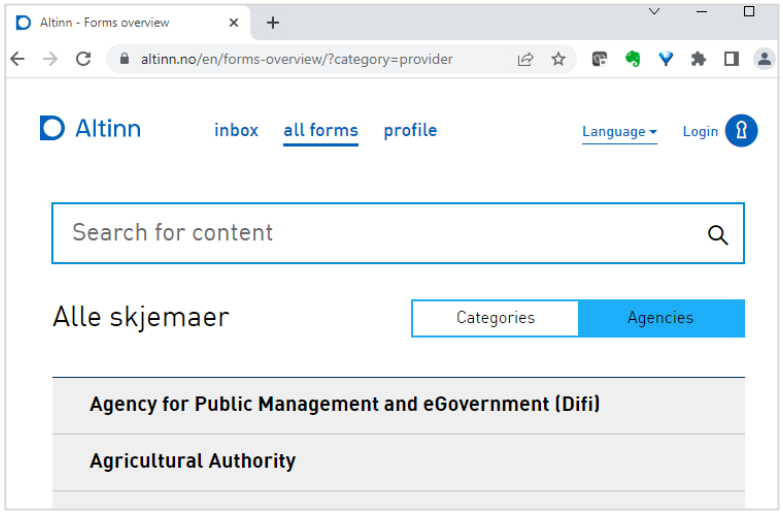
Groves et al. (2004). *Survey Methodology.* Hoboken: Wiley.

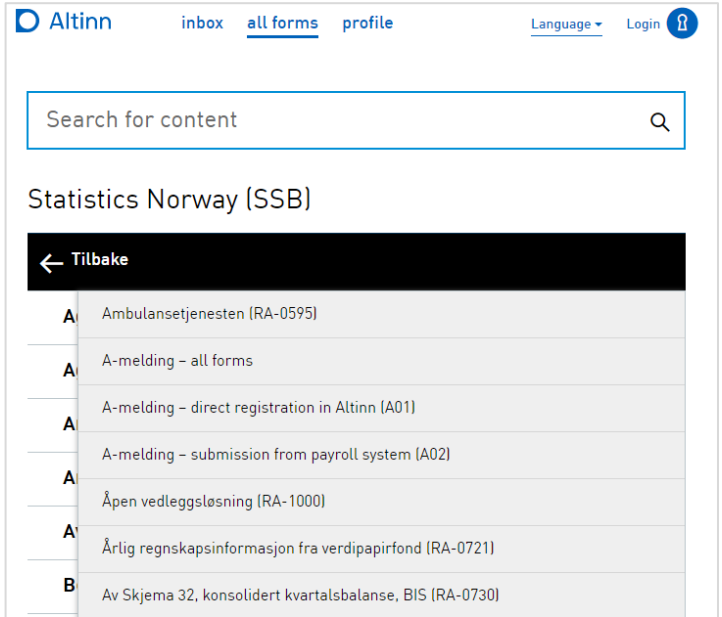
How to test a questionnaire?

- Expert review
- Cognitive interviewing
  - Verbal probing
  - Think-aloud
- Focus groups
- Observation of response process
- Pilot study
- Experiment

Web portals to facilitate response







Electronic data exchange

- Automation of primary data collection
- Requirement: technical standard

Current XML-based solutions:

- XBRL (eXtensible Business Reporting Language)
  ⇒ The Netherlands, maybe Sweden and Canada
- SAF-T (Standard Audit File for Tax)
  ⇒ Portugal
- SDMX (Statistical Data and Metadata eXchange)
  ⇒ Bank for International Settlements, ECB, Eurostat, IMF, OECD, UNSD, World Bank

UNIVERSITY OF L]UBL]ANA | SEB School of Economics and Business

Administrative data

Administrative and register data

General preconditions for extensive use of administrative sources in statistics production:

- legal basis to use these sources
- public approval of the benefits of using administrative data for statistical purposes
- unified identification systems to link sources
- comprehensive and reliable register systems for administrative needs
- cooperation among administrative authorities

UNECE (2007). Register-based statistics in the Nordic countries – Review of best practices with focus on population and social statistics, United Nations Publication

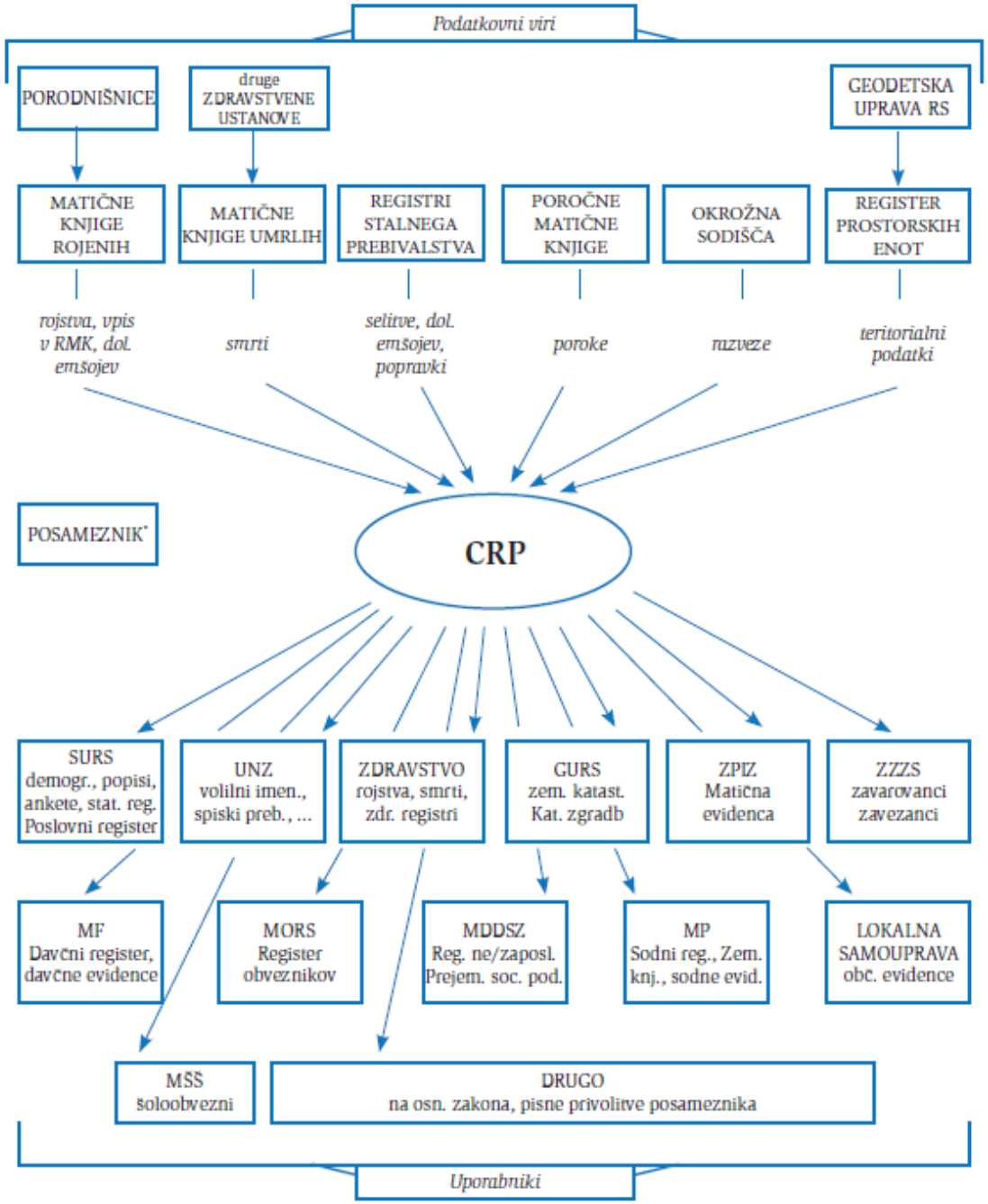
The year of establishing registers/introducing registers in census statistics by type of register and country

	Denmark		Finland		Norway		Sweden	
Type of register	*Estab-lished*	*First used in census*	*Estab-lished*	*First used in census*	*Estab-lished*	*First used in census*	*Estab-lished*	*First used in Census*
Central Population Register	1968	1981	1969	1970	1964	1970	1967	1975
Business Register	1975	1981	1975	1980	1965	1980	1963	1975
Dwellings	1977	1981	1980	1985	2001	*2011*	*2008?*	*2011?*
Housing conditions	1977	1981	1980	1985	2001	*2011*	*2008?*	*2011?*
Education	1971	1981	1970	1975	1970	1980	1985	1990
Employment	1979	1981	1987	1990	1978	2001	1985	1985
Family	1968	1981	1978	1980	1964	1980	1960	1975
Household[a]	1968	1981	1970	1975	2001	*2011*	*2011?*	*2011?*
Income	1970	1981	1969	1970	1967	1980	1968	1975
Totally register-based census		1981		1990		*2011*		*2011?*

[a] Household-dwelling unit, i.e. all the persons living in one dwelling

UNECE (2007). Register-based statistics in the Nordic countries – Review of best practices with focus on population and social statistics, United Nations Publication

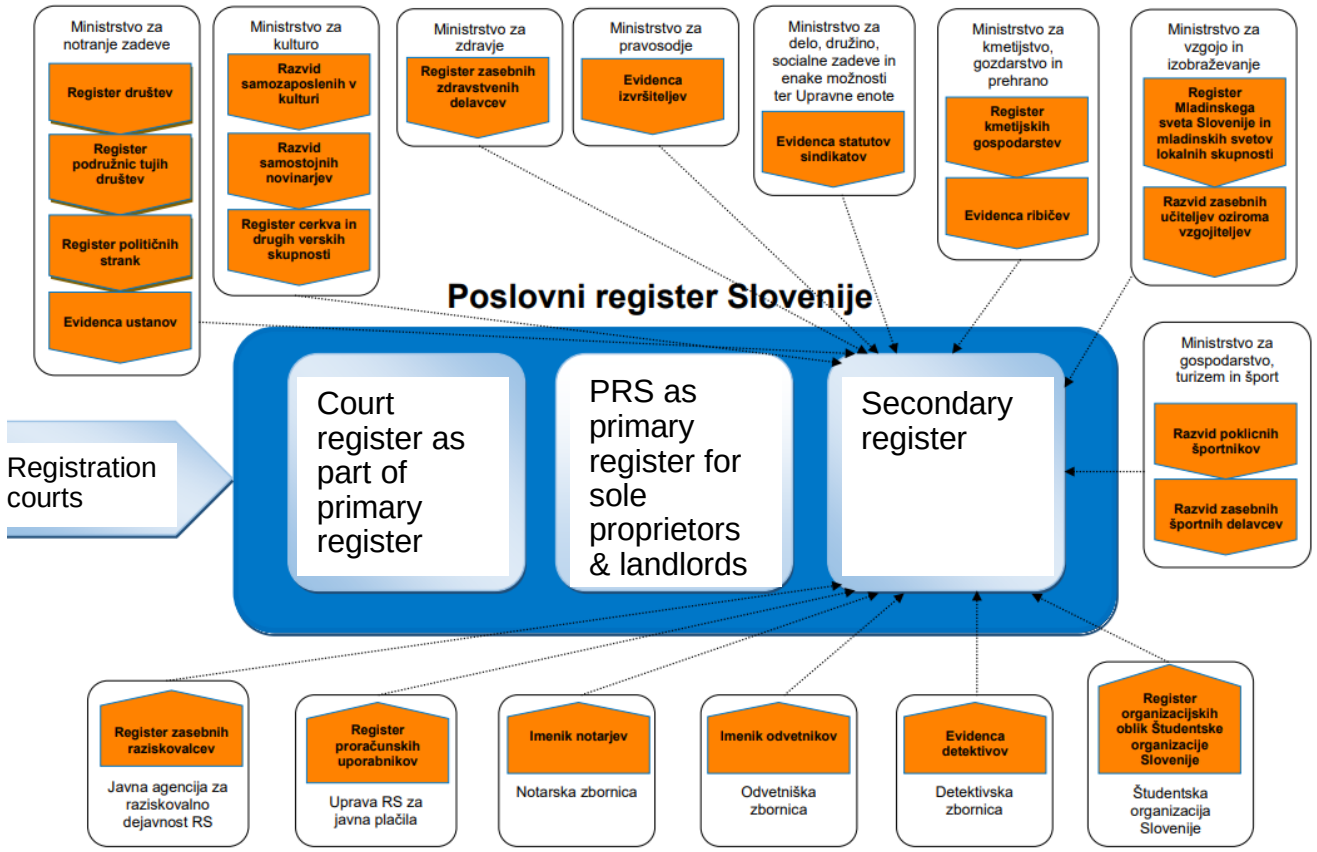
CRP - VOZLIŠČE PODATKOV O PREBIVALSTVU



Population census, Slovenia, 1 Jan 2011

- Register-based census (usual residence).
- Advantages and disadvantages.
- Statistical linking of cca 20 administrative & statistical sources:
  - Central Population Register + Household Register (Ministry of the Interior)
  - Real Estate Register + Register of Spatial Units (Surveying and Mapping Authority of RS)
  - Business Register of Slovenia (Agency for Public Legal Records and Related Services)
  - Statistical Register of Employment; Statistical surveys based on complete coverage, e.g. births, migration, tertiary education, recipients of scholarships (SURS)
  - Other databases, e.g. unemployed persons, graduates, national examinations, recipients of pensions and social transfers, insured persons, income tax.

Composition of the Slovenian Business Register (PRS)



Adapted from AJPES (2023). PRS. https://www.ajpes.si/Doc/Registri/PRS/Shema_PRS.pdf

UNIVERSITY OF LJUBLJANA | SEB School of Economics and Business

Big data

Examples

- Social networks (human-sourced information)

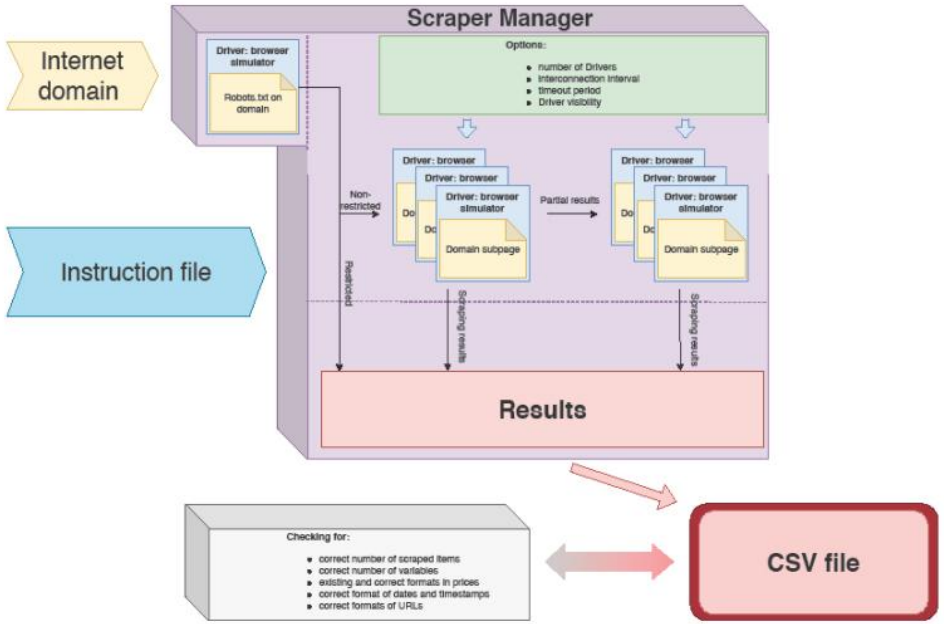
- Traditional (business) systems (process-mediated data), e.g. commercial transactions, stock records, e-commerce, credit cards, medical records...

- Internet of Things (machine-generated data), e.g. fixed sensors (home, weather/pollution, traffic, security), mobile sensors (cars, mobile phones, satellites), data from computer systems (logs)

Web scraping example from Statistics Slovenia (SURS)

At SURS we have developed our custom internet scraping tool which we use to scrape prices from internet shops. This document presents an example of how the tool and process work in the scope of scraping the data for products in ECOICOP classes Information processing equipment and Package holidays. The tool consists of 3 components: the Driver (a human-like behavior simulation that runs on browsers), the Scraper Manager (organizes multiple Drivers, inputs, and outputs to work concurrently), and the Instruction file (specific instructions for work on internet domain).

The general scraping process looks like this:



Eurostat (2020). Practical guidelines on web scraping for the HICP. Retrieved from https://ec.europa.eu/eurostat/documents/272892/12032198/Guidelines-web-scraping-HICP-11-2020.pdf/

Through use of HTML tags we can export item names, prices, Internet addresses.

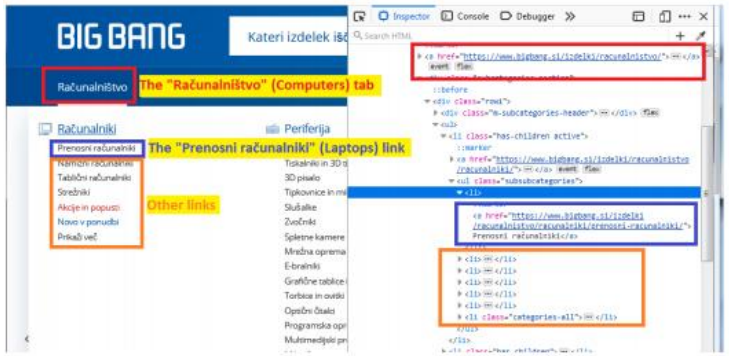


Figure 3: Understanding the structure of HTML

Big data at Statistics Netherlands (2021 & 2023)

cbs.nl/en-gb/over-ons/innovation

Beta products in development

An innovative way to estimate solar energy yields

22/04/2021 12:20

More insight into mobility with the doughnut map

31/03/2021 13:56

Identifying population movements using anonymised telephone data

05/03/2021 10:16

Deep learning for solar panel detection

10/02/2021 11:30

Using Website texts to detect Innovative Companies

10/02/2021 09:55

Extracting data on road network transportation from sensor data without sample design

19/01/2021 15:22

cbs.nl/en-gb/about-us/innovation

Beta products in development

Automatically detect solar panels with aerial photos

25/02/2023 01:36

What is synthetic data?

25/02/2023 01:36

Pregnancy as an indicator of economic recessions?

25/02/2023 01:36

Use machine learning to estimate chance of moving

25/02/2023 01:36

Follow-up study on detection of solar panels from earth observation

25/02/2023 01:36

Using social media to measure intentions to move house

25/02/2023 01:36

Traffic data

Total number of vehicles per minute at one road location (60 pts/h * 24h = 1440 pts)

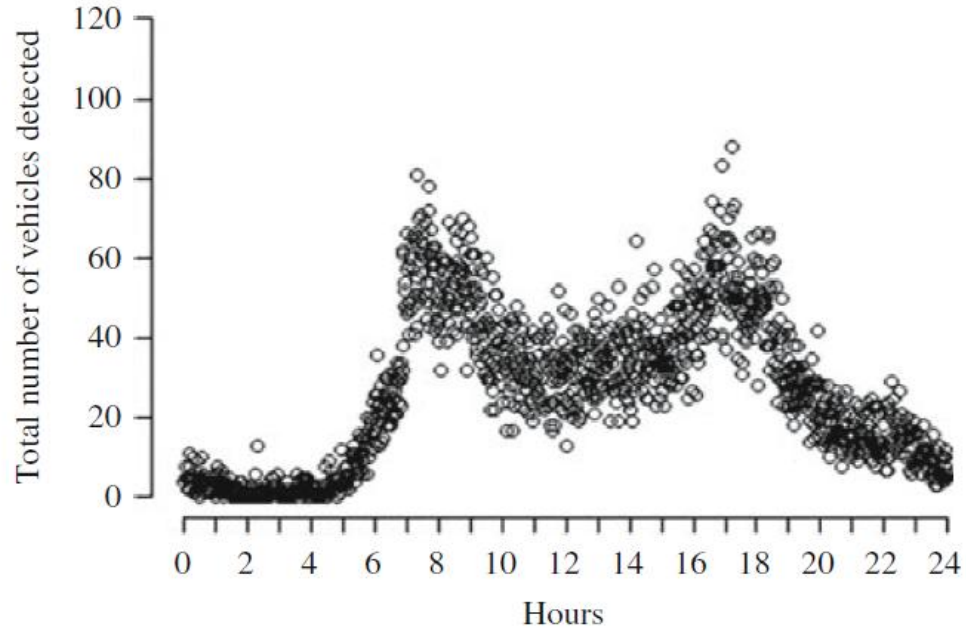


Fig. 3. Total number of vehicles counted by a detection location on highway A4 near Bergen op Zoom.

Daas, P., Puts, M., Buelens, B., et al. (2015). Big Data as a Source for Official Statistics. *Journal of Official Statistics*, 31(2), pp. 249-262. Retrieved 13 May. 2017, from doi:10.1515/jos-2015-0016

Big data and official statistics

Open issues:

- strategy, partnerships, legislation
- privacy, data protection, ethics
- skills, collaboration
- methodology, quality, IT

UNIVERSITY OF L]UBL]ANA | SEB School of Economics and Business

Challenges

Data sources compared

Data source	Sample survey	Register	Big data
Volume	Small	Large	Big
Velocity	Slow	Slow	Fast
Variety	Narrow	Narrow	Wide
Records	Units	Units	Events or units
Generating mechanism	Sample	Administration	Various
Fraction of population	Small	Large, complete	Large, incomplete

+ new topics
+ timeliness
+ detail
+ response burden reduced
– noise
– low information content
– veracity

Buelens, B. et al. (2014). Selectivity of Big data. Discussion Paper. Statistics Netherlands. Retrieved from http://www.cbs.nl/NR/rdonlyres/457A097A-DA43-4006-AFE0-A8E8316CFEF0/0/201411x10pub.pdf

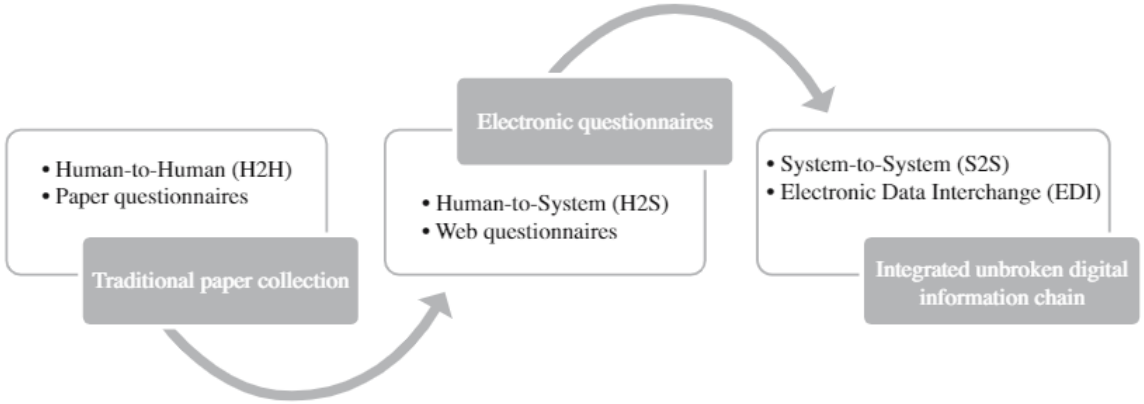


Fig. 1. Evolution of computerisation in primary business data collection (adapted from Erikson et al. 2016).

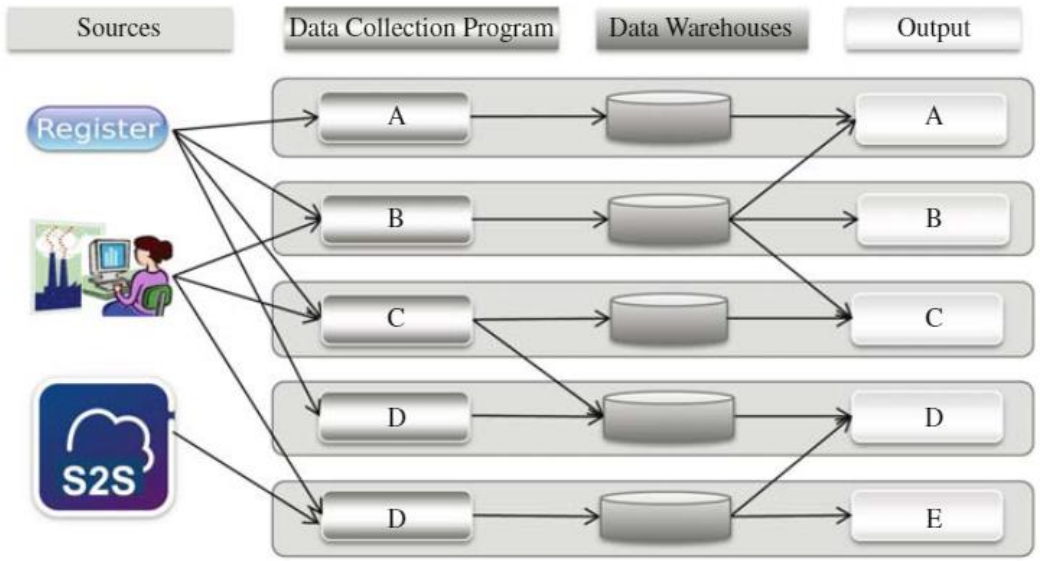


Fig. 2. Coordinating data collection and dissemination – the current situation (adapted from Erikson et al. 2016).

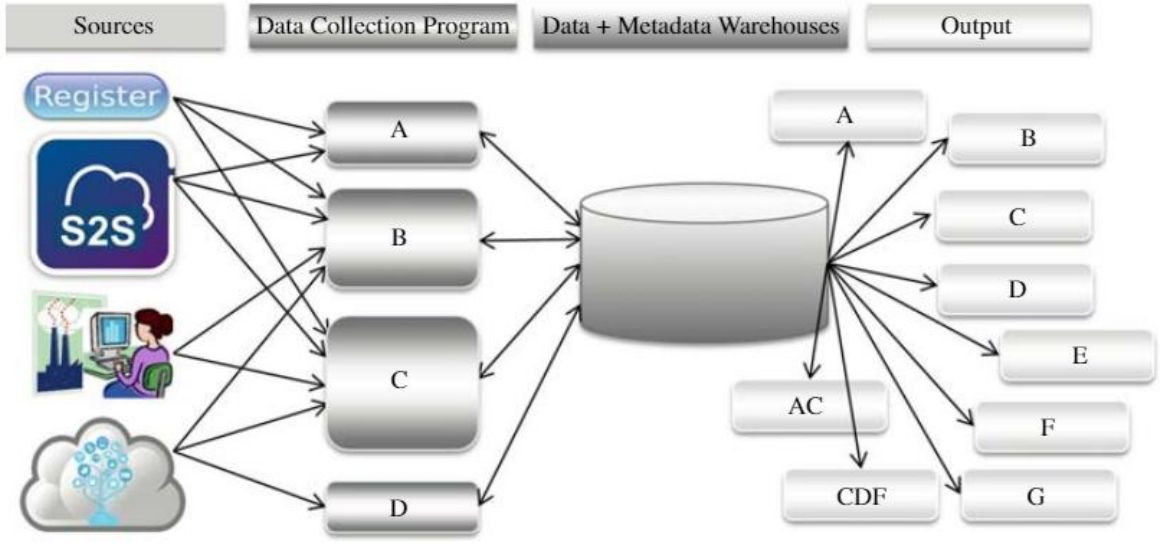


Fig. 3. Coordinating data collection and dissemination – the future (adapted from Eriksen et al. 2016).

Single-source official statistics are rare

Multisource statistics

- Statistics on businesses: surveys, business register, administrative data
- Population statistics: surveys, population register, mobile phone data
- Farm statistics: surveys, administrative data, satellite images, smart meters

Multipurpose sources

- Population register: vital statistics, migration, methodology (sampling frame, weights)
- Mobile phone data: commuting, migration, traffic, population, tourism

Change of **paradigm**?

Data sources alternative to surveys: more detail, faster and more often, new topics, cheaper, less burden, but

- quality issues with data in some alternative data sources
- some data not covered in alternative data sources

From finite population sampling methodology and design-based inference **to** a combination of design-based and model-based inference, machine learning, AI

An expanded toolbox:

- designing data collection processes
- checking quality of secondary data
- modelling at various stages
- designing statistical products

Privacy issues

Quality framework for combined survey, administrative and big data

Relevance
Population coverage
Population representativity
Variable validity
Concept stability
Correctability
Recentness
Processing timing
Accessibility
Meta-data
Comparability

Source: Gootzen, Y.A.P.M, Daas, P.J.H., van Delden, A. (2023). Quality framework for combining survey, administrative and big data for official statistics.Statistical Journal of the IAOS, 39, 439–446. DOI 10.3233/SJI-220110

IAOS

INTERNATIONAL ASSOCIATION FOR OFFICIAL STATISTICS

Improving Official Statistics

Home | About us | Statistical Journal | YSP | Conferences | Links | FAQ | Contact us | Resources

ISI AND IAOS EXPRESS CALL FOR AN END TO THE PERSECUTING OF ANDREAS GEORGIOU

The ISI and IAOS have followed with growing concern the continued persecution through the justice system in Greece of Andreas Georgiou, former Head of the Greek Statistical Office, ELSTAT.

Also see the previous ISI Statements and Letters Concerning Statistical Ethics.

A Greek appeals court has found Andreas Georgiou liable for slander. The slander charges relate to a public statement Georgiou made in his official role as President of ELSTAT while fulfilling his responsibility to defend the official deficit and debt statistics for Greece produced under his leadership. These statistics have been fully validated by Eurostat — the statistical office of the EU. The UN's Fundamental Principles of Official Statistics give statistical offices the right to comment publicly on criticisms and misuse of statistics.

Greece's legal actions and decisions against Georgiou on several closely related cases is now in its tenth year with the initial legal proceedings against him having begun in September 2011.

We restate our grave concern that these continued prosecutions have damaged the scientific integrity of highly regarded work addressing Greece's problematic fiscal statistical reporting from the 2000s. This case has implications for the international statistical system, and the rights

FINANCIAL TIMES

myFT

Your guide to a disrupted world

Start a 4-week trial

Turkish economy + Add to myFT

Turkish president sacks statistics chief as inflation tension escalates

Sait Erdal Dincer lasts just 10 months in the role amid reports of friction over direction of policy

Goldman Sachs forecasts annual consumer price inflation will hit 48 per cent in January © REUTERS

Save

Laura Pitel in Ankara JANUARY 29 2022

333

Feedback

SIGN IN | SUBSCRIBE | SUPPORT US

The Slovenia TIMES

DAILY NEWS | BEST 2021 STORIES | BUSINESS | POLITICS | SOCIETY | TRAVELS | EVENTS | GIFTS FROM SLOVENIA

PM asked to explain dismissal of Statistics Office head

11. June, 2020

The letter was sent to Janša yesterday with the aim to provide complete compliance with the principles of impartiality and professional independence of national statistics offices, the press service said.

The move comes after the government dismissed in late May director general of the Statistics Office Bojan Nastav and appointed Tomaž Smrekar acting director general. The latter will serve until a full-fledged director is appointed but no longer than six months.

Earlier this month, the Statistics Council, an expert advisory body, asked the Constitutional Court to review the dismissal of Nastav.

The council is not sure which law applies in this case – the one on public sector employees, which allows the government to dismiss a top public sector employee a year after the employee started their job, or the national statistics act.

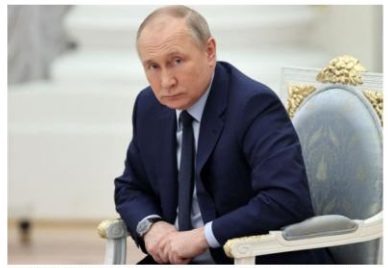
Janša said in late May that the dismissal of Nastav was necessary "due to responsiveness". "This is about a body functioning in a professional fashion, being responsive, so that we can rely on getting data tomorrow if we need it."

THE WALL STREET JOURNAL.

WORLD | EUROPE | RUSSIA

Russia Blocks Economic Data, Hiding Effect of Western Sanctions

Authorities have stopped publishing data on banks, oil and debt



Under President Vladimir Putin, Russia has stopped disseminating many economic statistics, making it challenging to assess the effect of Western sanctions imposed following Moscow's invasion of Ukraine.

PHOTO: MIKHAIL TERESHCHENKO/AGENCE FRANCE-PRESSE/GETTY IMAGES

By Georgi Kantchev Follow

April 23, 2022 5:30 am ET

SHARE | TEXT | 52

Listen to article (7 minutes)

The West's window into the Russian economy is closing.

Roeland Beerten · 1st

Chief Statistician, National Bank of Belgium

3w · Edited ·

Given the confusion and speed at which things develop in the USA 🇺🇸 this is a good - but highly worrying - overview of the threats to the system of trustworthy independent high quality #officialstatistics - by Steve Pierson and colleagues at the American Statistical Association - ASA :

👉 cuts in statistical programs

👉 delays, reduction of detail, or cancellations of data products

👉 decreases in budget or staffing

👉 politicizing leadership and undermining staff security

👉 accessing and using statistical data for nonstatistical purposes

Five Situations to Watch to Ensure Trusted, Quality Federal Statistics

Steve Pierson

UL | SEB

Jessica Espey · 2nd

Associate Professor, School of Geog...

1d ·

+ Follow

Census response rates are declining worldwide, whilst large scale survey programmes like DHS are being cut.
But when people vanish from data they vanish from policy.
Read Andy Tatem and my latest piece on this at TheConversation:

Global population data is in crisis – here's why that matters

theconversation.com