Administrative Data Cooperation – Statistics Finland

Ville-Matti Pilviö
Business Registry – Data Collection
Statistics Finland.
24.9.2023

Content

1. Intro: Statistics Finland and myself

2. Cooperation with Administrative Data Providers

3. Quality check of Administrative Data Sources

4. Dataflow

5. Internal and external Users of the SBR

Ville-Matti Pilviö

- Senior Statistician – Business Statistics
- Statistics Finland: 12 years – so a relative newcomer
- Statistical Programming and software development
- Badminton
- Fishing

@VPilvio

/villepilvio firstname.pilvio@stat.fi

+358 29 591 3502

Statistics Finland in figures

Statistics Finland

624
statistical releases

840 experts, of whom

150 statistical interviewers

3.57 personnel job satisfaction index, on a scale of 1-5

150 statistics

4 700 tables in the database

for researchers
423 user licences
204 data deliveries

1 048 service deliveries subject to charge

Funding

EUR 51.5 millio from the state budget

EUR 8.5 million income from charged services

Data needed for statistics

95% from registers
5% with inquiries

95.5% of inquiries could be answered online

90% of Finns know Statistics Finland

90% of Finns consider Statistics Finland impartial

8 000 contacts answered by the information service

3.8 million users on the **stat.fi** website

156 years of data and statistics

112 pcs

Statistical inquiries

168 000 responses from organisations
181 000 responses from persons and households

16 500 followers on Twitter

Read more: stat.fi

Business Statistics:

Statistics under the topic Enterprises:

https://stat.fi/en/topic/enterprises

- Accommodation statistics
- Air emission accounts
- Aquaculture
- Bankruptcies and business restructuring proceedings
- Commercial fellings
- Commercial inland fishery
- Commercial marine fishery
- Crop production statistics
- CVTS, continuing vocational training
- Egg production
- Energy use in manufacturing
- Enterprise openings and closures
- Environmental goods and services sector
- Environmental protection expenditure accounts
- Environmental taxes
- Finnish affiliates abroad
- Fish processing
- Flows of wood material
- Foreign affiliates in Finland
- Foreign direct investments
- Foreign trade in agri-food products

- Foreign trade in fish
- Foreign trade in roundwood and forest industry products
- Foreign trade of goods
- Horticultural statistics
- Index of turnover in industry
- Index of turnover of construction
- Industrial output
- Innovation
- Job vacancy survey
- Labour cost index
- Labour cost survey
- Meat production
- Milk and milkproducts statistics
- New orders in manufacturing
- Number of livestock
- Operating profit in non-industrial private forestry
- Other entrepreneurship in agriculture and horticulture
- Private sector hourly wages
- Private sector monthly salaries
- Producer price indices for services
- Production of electricity and heat

- Providers of education and educational institutions
- Regional statistics on entrepreneurial activity
- Renovation building
- Research and development
- Statistics on business subsidies
- Statistics on cereals and protein and oilseed purchased, used and stockpiled
- Statistics on service industry commodities
- Statistics on the finances of agricultural and forestry enterprises
- Structural business and financial statement statistics
- Structure in agriculture and horticulture businesses
- Turnover of service industries
- Turnover of trade
- Use of crops on farms
- Use of information technology in enterprises
- Volume index of industrial output
- Volumes and prices in industrial roundwood trade
- Wage and salary indices

YTY: An Integrated prodution system for Business Statistics (2013 ->)

YTY is a (very) large system

- It is based on our Statistical Business Registry
    - Used for different business statistics

- Key Concept:
    - Integrate all Business Statistics into one central production system

- Further integration is ongoing:
    - Touristm statistics
    - Education organizers and locations
    - More are being integrated

YTY: Currently producing these statistics

Structural business and financial statement statistics (preliminary t+9, final t+12)
Regional statistics on entrepreneurial activity (t+12)
Industrial output (t+11)
Enterprise openings and closures (Q1:t-5, Q2:t-2, Q3:t+1, Q4:t+4,)
Index of turnover in industry (t+75 days)
Index of turnover of construction (t+75 days)
Turnover of trade (preliminary t+22 and t+45 days, final t+75 days)
Turnover of service industries (t+75 days)
Wage and salary indices (t+45 days)
International trade in services (preliminary t+5, final t+12)
Foreign affiliates in Finland (FATS) (t+12)
Finnish affiliates abroad (FATS) (t+16)
Business Services Statistics (t+10)
Business demography (t+12...18)
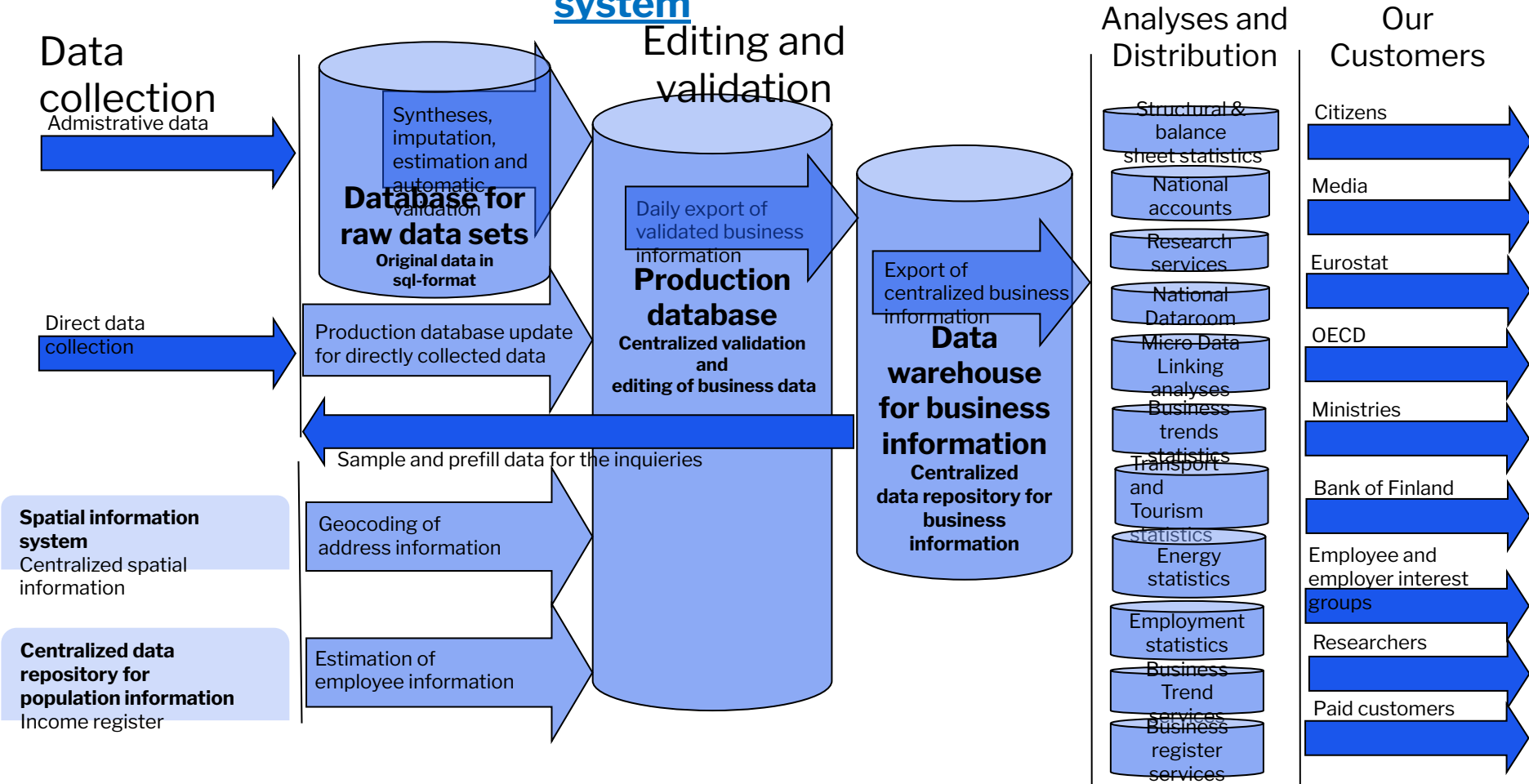Registrations and bankruptcies (t+40 days)
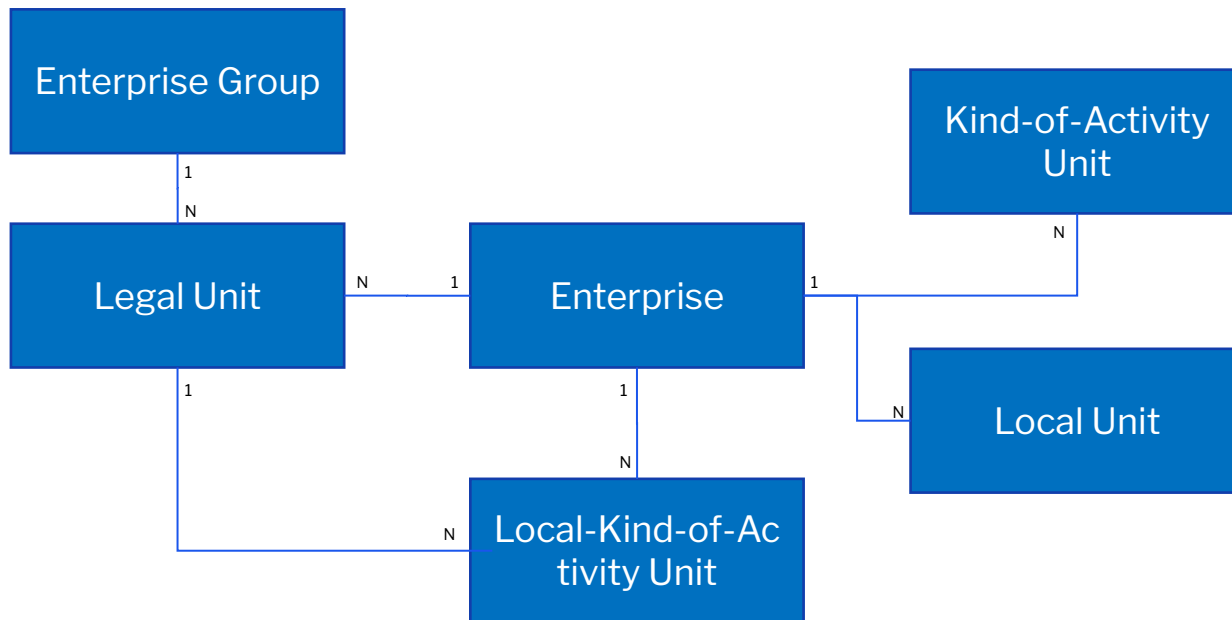+ Chargeable services
+ Research services

# YTY – Business Information System

Business information system

Data collection

Admistrative data →

Direct data collection →

Spatial information system
Centralized spatial information

Centralized data repository for population information
Income register

Syntheses, imputation, estimation and automatic validation

Database for raw data sets
Original data in sql-format

Production database update for directly collected data

Sample and prefill data for the inquieries

Geocoding of address information

Estimation of employee information

Editing and validation

Daily export of validated business information

Production database
Centralized validation and editing of business data

Export of centralized business information

Data warehouse for business information
Centralized data repository for business information

Analyses and Distribution

Structural & balance sheet statistics

National accounts

Research services

National Dataroom

Micro Data Linking analyses

Business trends statistics

Transport and Tourism statistics

Energy statistics

Employment statistics

Business Trend services

Business register services

Our Customers

Citizens →

Media →

Eurostat →

OECD →

Ministries →

Bank of Finland →

Employee and employer interest groups →

Researchers →

Paid customers →

Unit Structures and Relationships

Internal and External Users of the SBR

Stakeholders, coordination and co-operation

Ville-Matti Pilviö
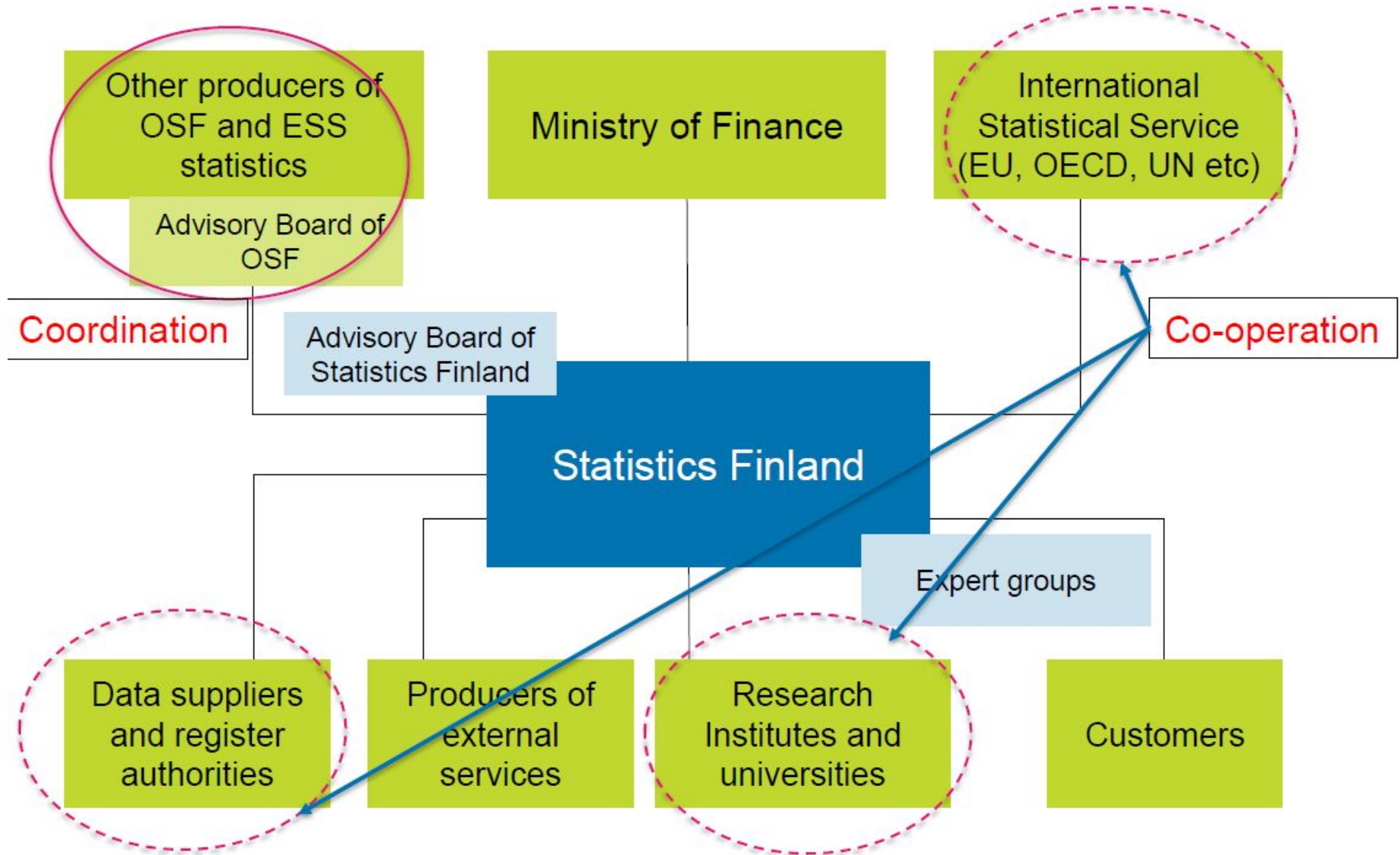
Legal basis for the use of administrative data

- Fundamental principles of official statistics, (1994)
  - **Principle 5.** Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or **administrative records**. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

- EU regulation 223/2009 and its amendment 759/2015, **Article 17a**
  - 1. the NSIs ... shall have the right to **access and use, promptly and free of charge, all administrative records** and to integrate those administrative records with statistics, to the extent necessary for the development, production and dissemination of European statistics
  - 5. The NSIs and owners of administrative records shall establish the necessary **cooperation mechanisms**.

- Statistics Act (280/2004) of Finland
  - Section 4: When data are collected for statistical purposes the **primary exploited sources shall be data accumulated in administering** the tasks of general government
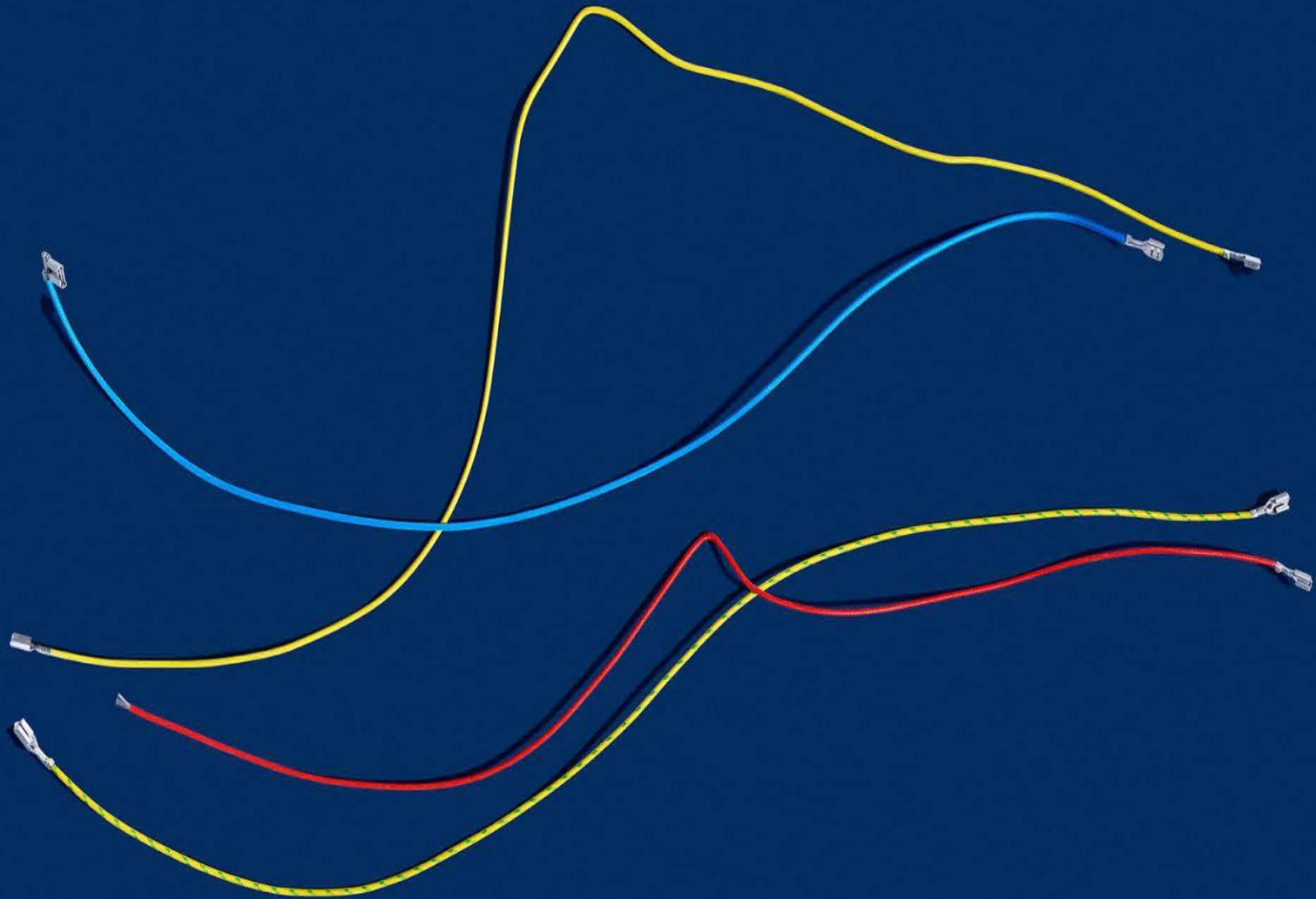
European Statistics Code of Practice (CoP), principles 2 and 8

- Article 11 of EU Statistical Law explains the meaning of CoP

- Priciple 2 : Mandate for data collection
    - 2.1: The mandate of the statistical authorities to collect information for the development, production and dissemination of European Statistics is specified in law.
    - 2.2: **The statistical authorities are allowed by law to use administrative data for statistical purposes.**

- Principle 8: Appropriate Statistical Procedures
    - 8.7: Statistical authorities are involved in the design of administrative data in order to make administrative data more suitable for statistical purposes.
    - 8.8: **Agreements are made with owners of administrative data which set out their shared commitment to the use of these data for statistical purposes.**
    - 8.9: Statistical authorities co-operate with owners of administrative data in assuring data quality.

- https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice

Cooperation with Administrative Data Providers – Statistics Finland

What is administrative data?

- No definition in statistical legislation

- Data collected primarily for administrative purposes, as a result of administrative processes and decision making

- Kept and maintained by various state offices belonging to different administrative sectors

- Administrative registers are regulated by laws (right to collect, store, release, ect.)

Cooperation with data providers

- Bilateral data transmission **agreements** between Statistics Finland and the owners of administrative registers and data

- Discussions at High-level meetings (Director Generals) 1-2 years interval

- Discussions on expert level on monthly or short term intervals

- Identified contact persons for data transmission guarantees continuous and regular co-operation

- **Information request** – when data is needed only once

Content of data transmission agreements

- More specific matters are agreed in the **data transimission agreement:**
  - Data release principles
  - Storage of released data
  - Data secrecy and protection
  - Data delivery (in appendix or here)
  - Compensation of costs
  - Contact Persons
  - Appendices of agreement (deliveries, technical matters, variables)
  - Validity and copies of agreement
  - Signatures

- **Information request** – when data is needed only once or asked first time
  - By email? Delivery, variables, etc

Appendix A: Data deliveries

File names	Delivery frequency	Delivery schedule	Delivery mode	Delivery method	File size category (S-M-L)*	SF: contact persons for data content	XX: Contact persons
						First name Last name, firstname.last-name@stat.fi tel. +358 29 551 xxxx	First name Last name, first-name.last-name@xxxx.xx tel. xxx xxxx

File names are of the form del_data_timeref_part, where

- *1. del: Identifier of the data supplier (e.g. official abbreviation) This part is always the same,*
- *2. data: Name describing the data file or part of it*
- *3. timeref: The identifier for earlier and future data deliveries consisting of*
  - *a) year specifier y20xx (mandatory data) and if necessary*
  - *B) quarter q1-q4 or period p1-px or month m01-m12 or week w01-w53*
- *4. A possible piece indicating the file part used for all large deliveries requiring division of data: o01-o99*

Appex B: Technical matters related to data delivery

- Exporting files –basic data:
    - Data supplier, technical contact person
    - Supply method (server)

- Data files the supplier sends to SF server
    - Server name
    - Trasfer protocol
    - Authentication practice
    - Sender's identifiers (id and password given by SF)

- Data files that SF retrieves from supplier's server
    - Server name /IP-adress
   - SF identifiers
    - Authentication practice

National Board of Administrative Data Suppliers and Users

- This was in active use in Finland in the 1990s and 2000s – during the initial phases of using Administrative Data for Statistical Production

- It allowed a regular channel for suppliers and users to interact

- It was very beneficial to bring about a "cultural change" in developing a collaborative, cooperative mentality between the AD Suppliers and Users

- Personal interaction

- Understanding why Users wanted access to the data

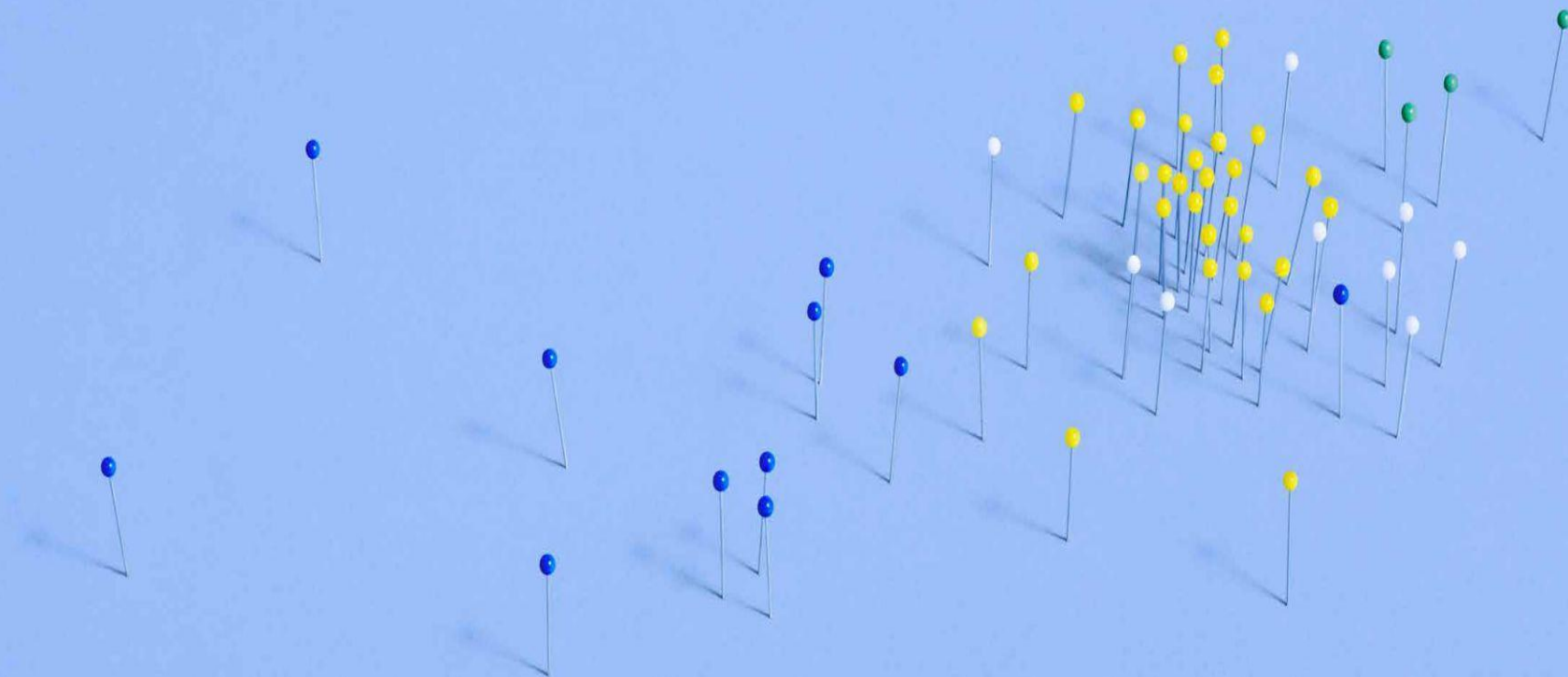- Building trust that the data will not be misused

Phases of administrative data collection system

- The system consists of three phases: automatic phase, centralised phase and statistical production process.
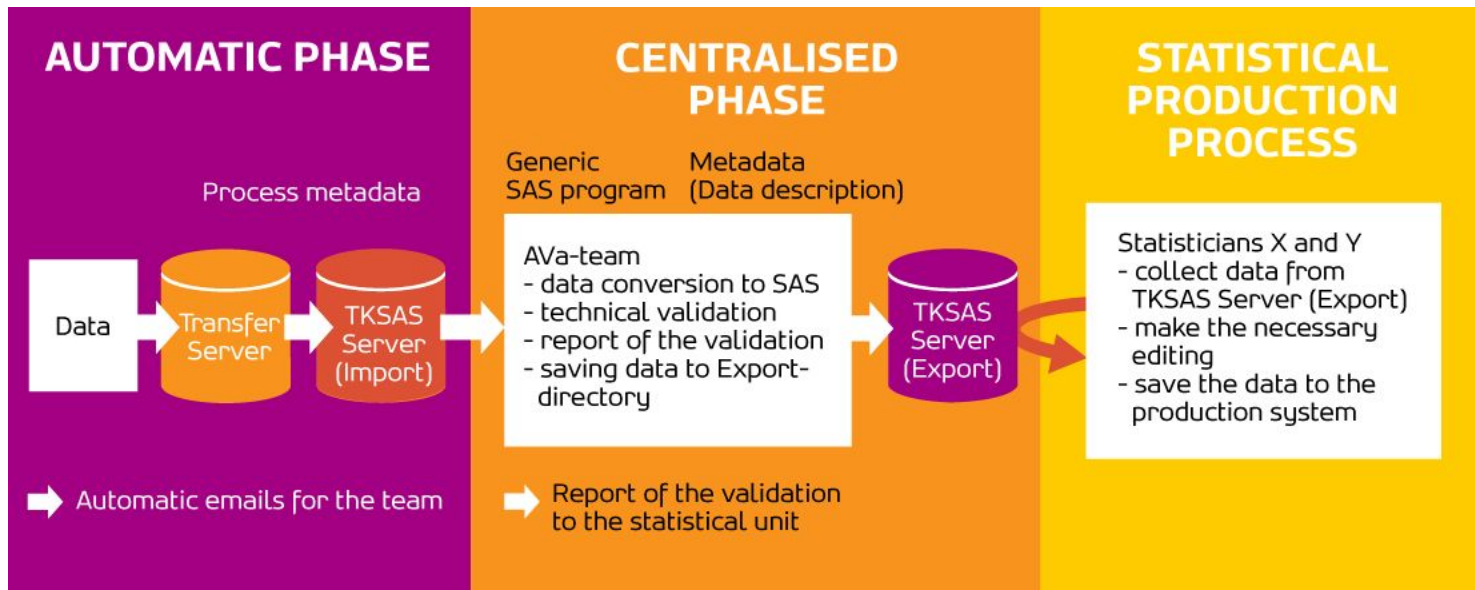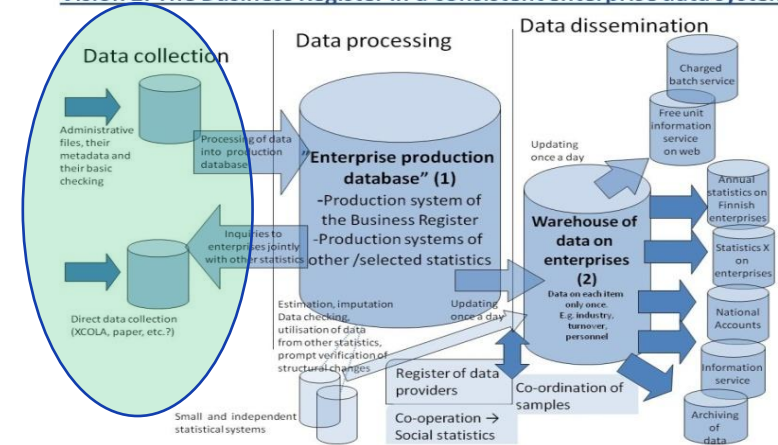
Quality check of Administrative Data

ETL1

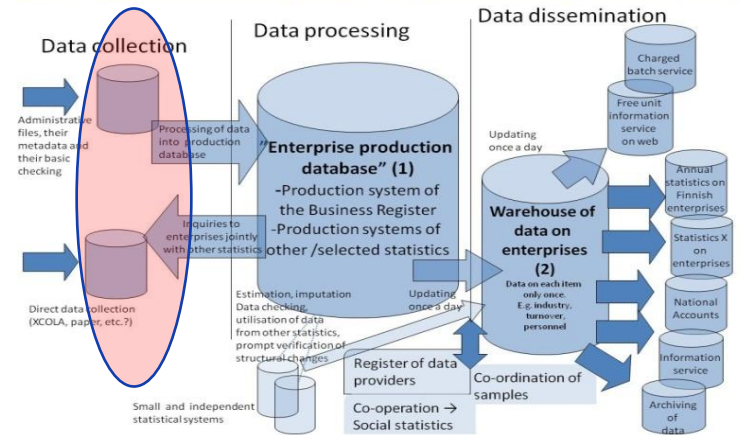Vision 2: The Business Register in a consistent enterprise data system

Data Arriving into YTY:

- **Administrative data**

- **Data from direct collections (Surveys)**

- These are implemented by standardised procedures

- ETL: Extract – Transfer – Load.

Raw Data Collection Warehouses

Vision 2: The Business Register in a consistent enterprise data system

- All data entering StatFin enters through one "door" – AVA (standard entry point)

- Is now in safe keeping and in a usable format

- No corrections are made to this data contents

- Data are technically corrected (formats)

- Validity of data has been established

- Source data must be described in the **metadata** system

- Standardised Quality controls (meta-controlled) of data entering by "type"

Thank You - Kiitos!

Ville-Matti Pilviö / Statistics Finland
ville-matti.pilvio@stat.fi
Twitter @VPilvio