



# Modern, process oriented and metadata driven statistical production

**Anna Długosz**  
Deputy Director  
ICT Systems, Geostatistics and Census Department

# Presentation plan

- Model of the Statistical Production Process Introduction
- Data repositories
- Metadata repository
- Data editing
- System Processing of Administrative Data
- Variable Quality System
- Statistical Operations System

# Model of the Statistical Production Process

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update outputs systems	8.1 Gather evaluation inputs
1.1a Prepare Statistical Programme (SSPPS)	2.2 Design variable descriptions	3.2 Reuse or build processing & analysis components	4.1a Geocode frame & sample	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.2 Consult & confirm needs	2.3 Design collection	3.3 Reuse or build dissemination components	4.2 Set up collection	5.3 Review & validate	6.2a Prepare spatial analyses & maps	7.2a Manage spatial analyses & maps using GIS	8.3 Agree an action plan
1.3 Establish output objectives	2.4 Design frame & sample	3.4 Configure workflows	4.3 Run collection	5.4 Edit & input	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	
1.4 Identify concepts	2.5 Design processing & analysis	3.5 Test production system	4.3a Geocode collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5a Design geocoding frame, sample & data collection	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production system & workflow	3.7 Finalise production system	4.5 Verify plans	5.7 Calculate aggregates		7.6 Satisfaction research	
1.7 Prepare a general plan for the implementation of the production process	2.7 Design quality indicators	3.8 Verify projects and construction plans		5.8 Finalise data files		7.7 Prepare an improvement plan for the next research cycle	
1.8 Prepare a detailed plans for subsequent phases	2.8 Verify the general plan and detailed plans						
1.9 Verify the concept to improve satisfaction							

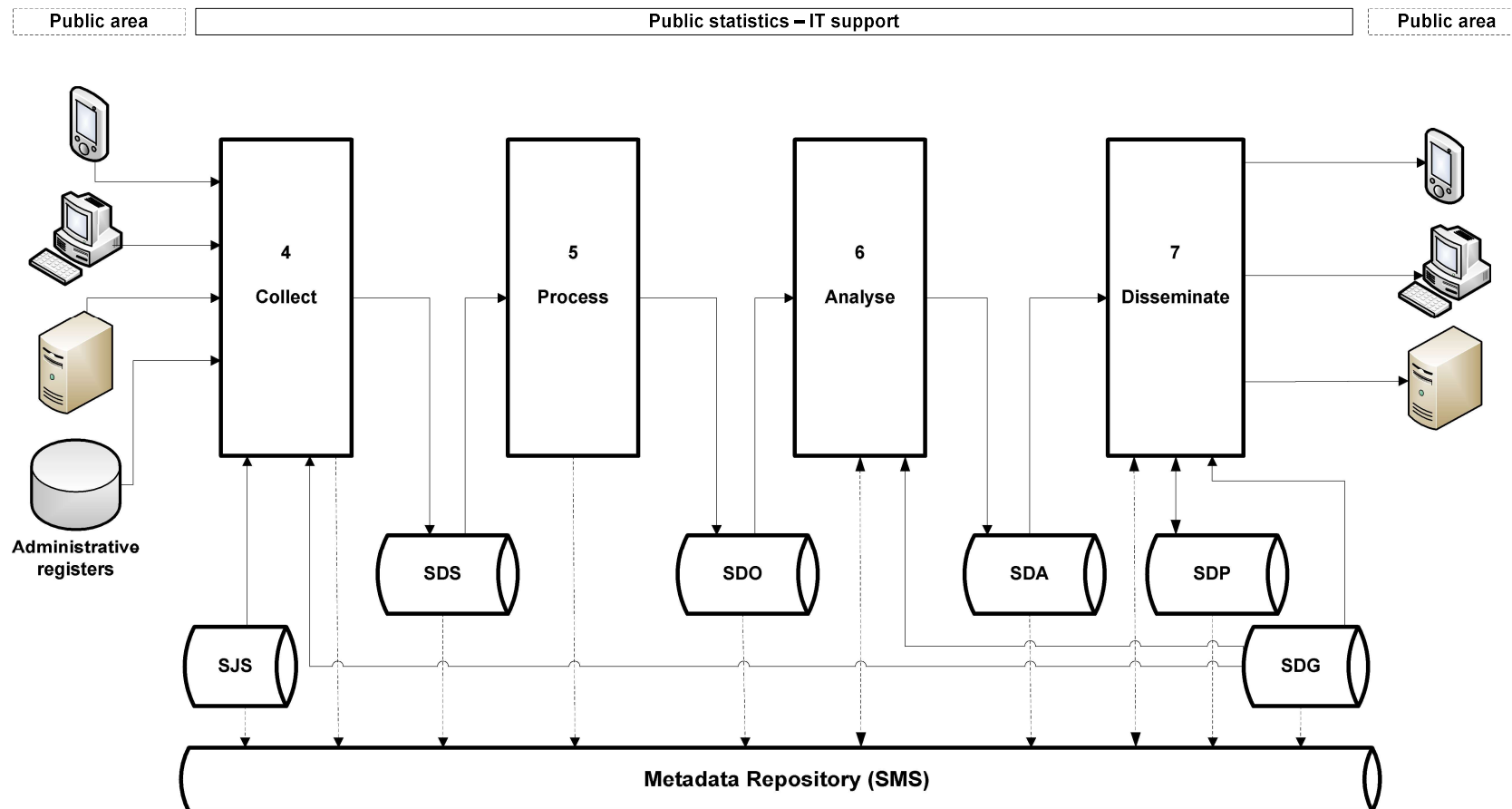
# The concept of repositories

- in order to separate the proposed architectural solutions from the impact of specific products (mainly databases and warehouses) offered by various commercial manufacturers, the concept of "Data Repository" was introduced;
- this approach allows generalization of the description of the functions of database objects in the statistical production process storing and providing statistical data for subsequent "states", resulting from the advancement of their processing in subsequent phases of the statistical production process.

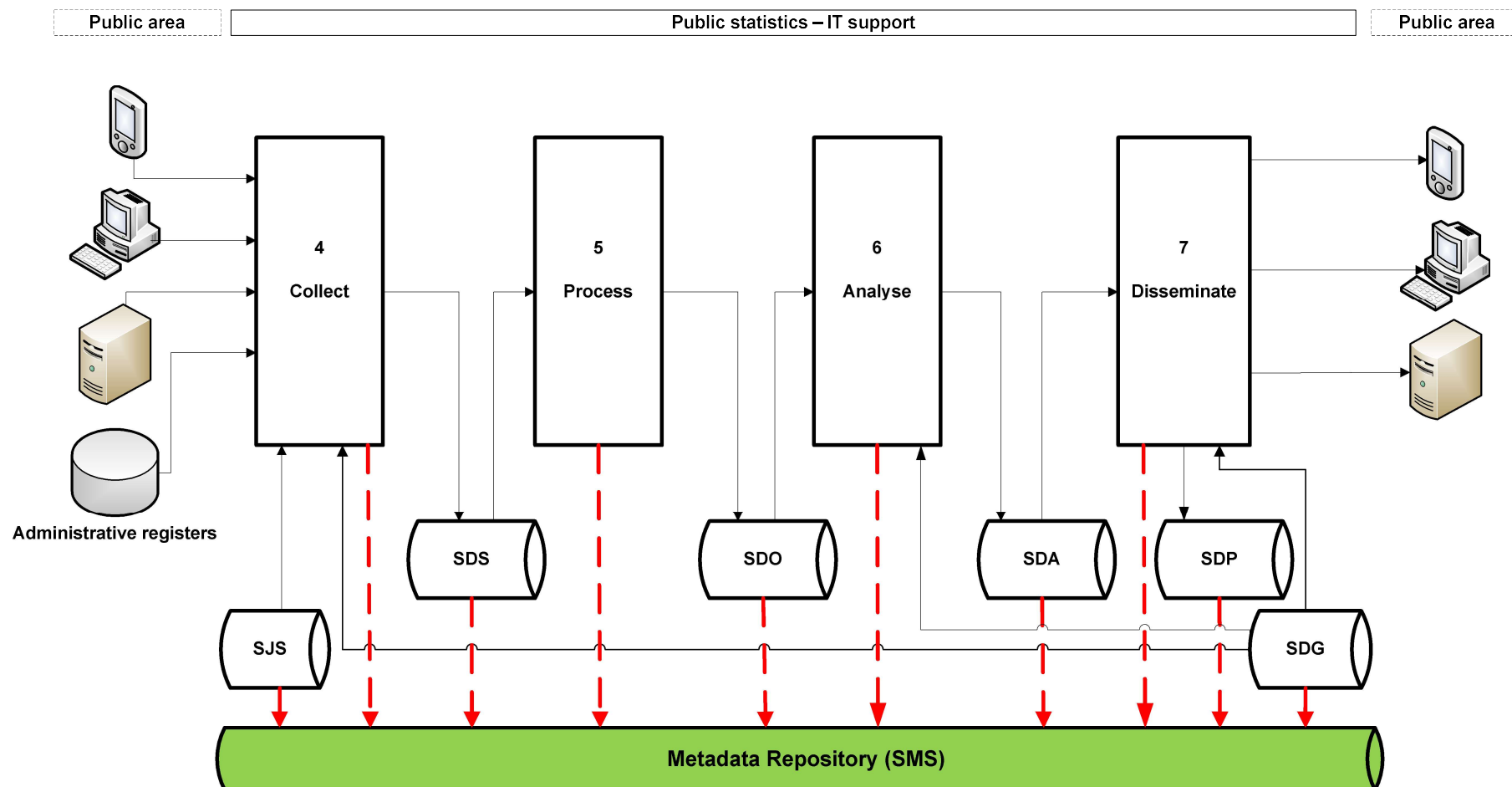
# The Reference Architecture Framework

- a registry data repository (SJS)
- a raw data repository (SDS)
- an operational data repository (SDO)
- an analysis data repository (SDA)
- a geospatial data repository (SDG)
- a publication data repository (SDP)
- a metadata repository (SMS)

# Phases and data repositories according to the MPPS model



# The metadata repository



# The metadata repository

It is a kind of knowledge base on all resources and processes related to the production of statistical data.

It is necessary to monitor and manage the production process.

The metainformation system with the metadata repository are necessary to improve the quality, accuracy and adequacy of the data.



# Data editing

- all the activities on data that we have to do in case of incorrect, incomplete, unreliable or outdated data;
- a process of improving data and data sets;
- special treatment of administrative data - initially collected for a non-statistical purpose;
- sets of data from administrative sources must be processed into a set of data corresponding to the standards adopted in official statistics.

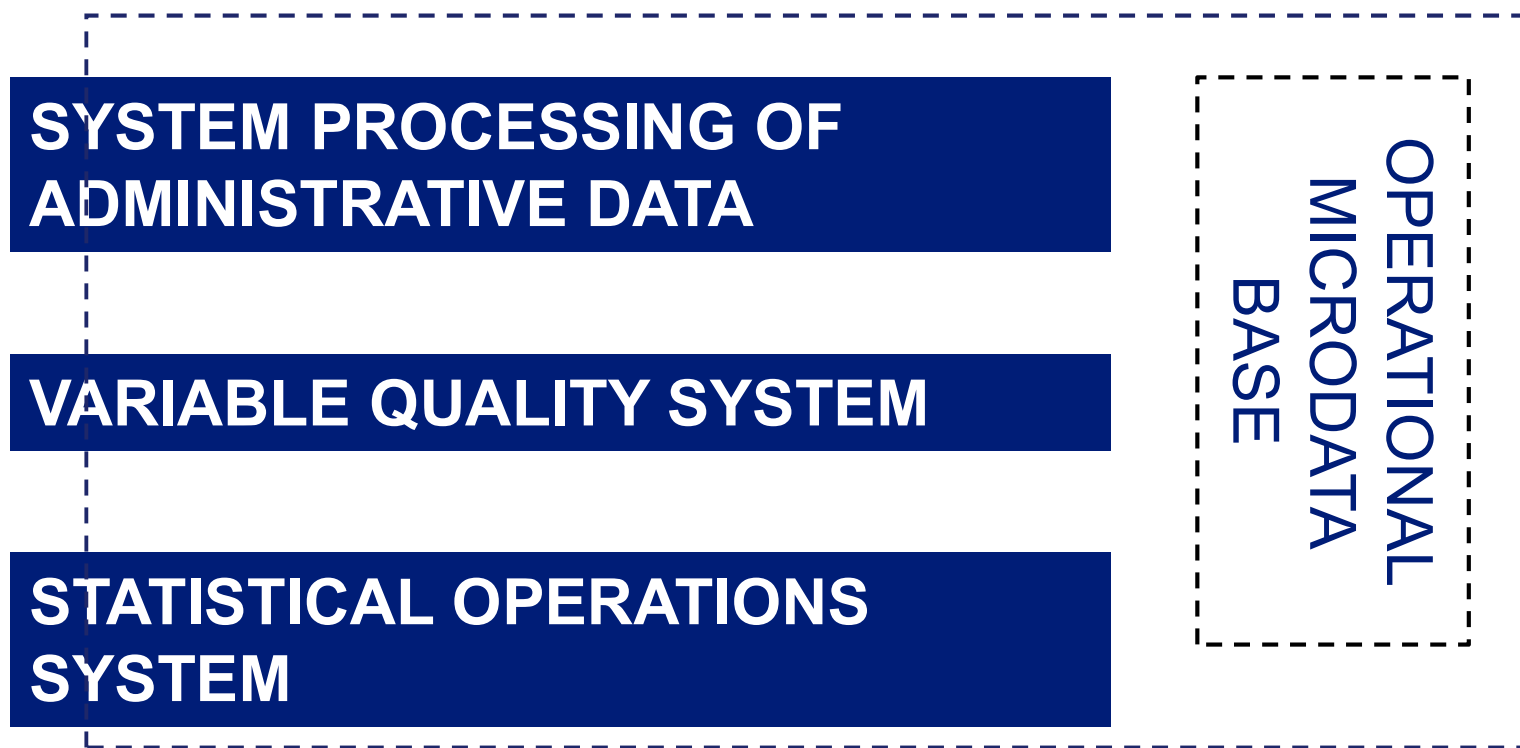
# Why do we use administrative data?

- demand on the part of external recipients for statistical data of high quality and at low levels of territorial aggregation;
- increased number of refusals to participate in statistical surveys;
- economic reasons - significantly lower research costs;
- facilitating business activity - striving to minimize the reporting burden on business entities;

# Why do we use administrative data?

- systematic improvement of the quality of data from administrative sources, which allows for their use on an increasing scale;
- official statistics has the necessary knowledge, experience and tools in the field of knowledge of data from administrative sources, improving their quality, standardization and integration.

# How do we do that?



# System Processing of Administrative Data

SPAD – a separate area in the Operational Microdata Base environment where administrative registers are processed in the form made available from the data owners.

The final product is the statistical register and the recipients of other Statistics Poland units.

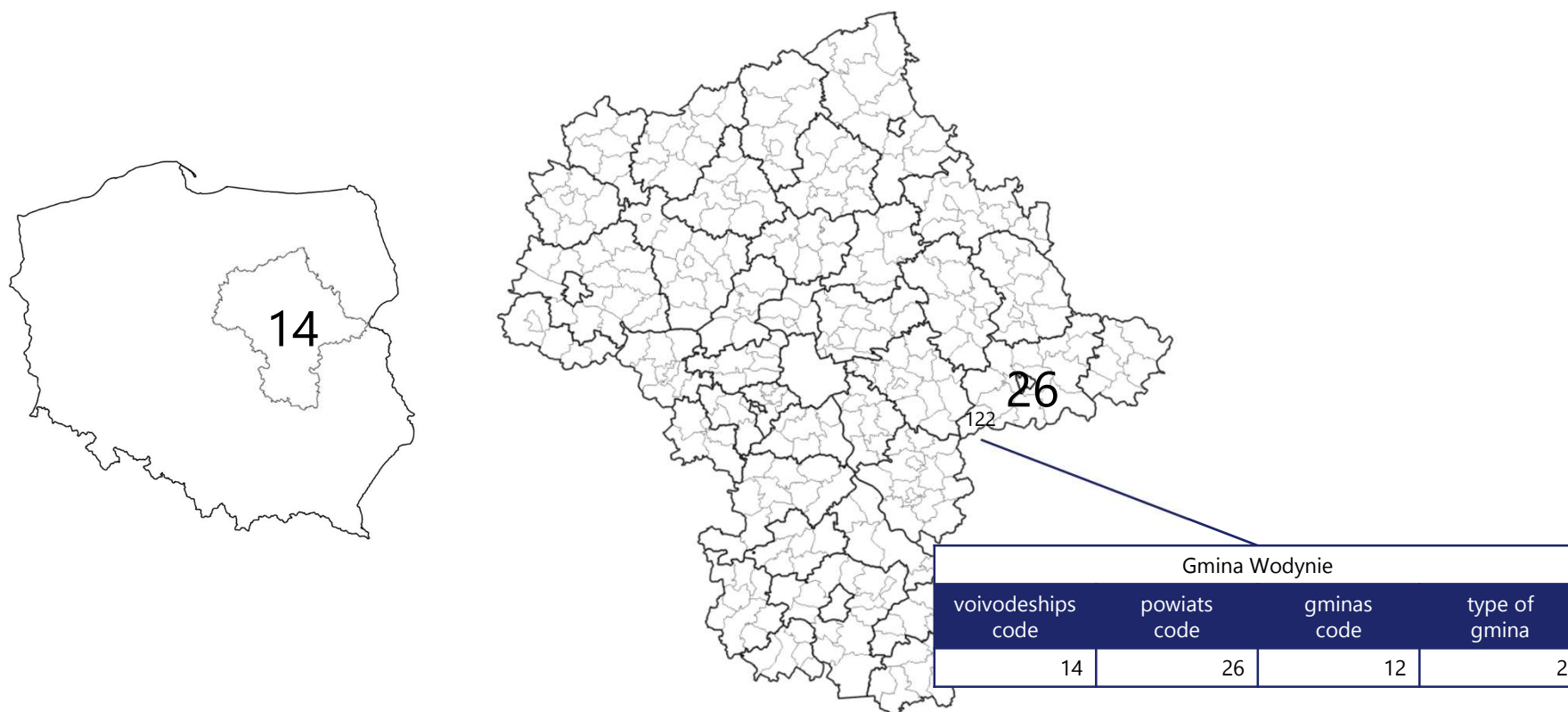
# SPAD – the main processes

- development of set conversion rules – individually for each data set;
- conversion of sets to flat tables + deduplication and transposition;
- elimination of incorrect records in the data;
- improving the quality of data – complementing data on the basis of other features in the set;
- complementing data as a result of integration with other data sets;

# SPAD – the main processes

- cleaning data using defined dictionaries;
- correction of data on the basis of logical rules;
- conversion of values;
- the development and calculation of ad hoc quality statistics;
- an inspection of the operations carried out.

# System Processing of Administrative Data — coding with TERYT identifiers





# Variable Quality System

VQS – an automated system for calculating basic quality indicators and creating reports.

The application was also created in response to increasing interest in data from administrative registers.

VQS is used to study quality improvement and report creation for all data from administrative registers that are processed by the Statistics Poland.

# Variable Quality System – presentation results

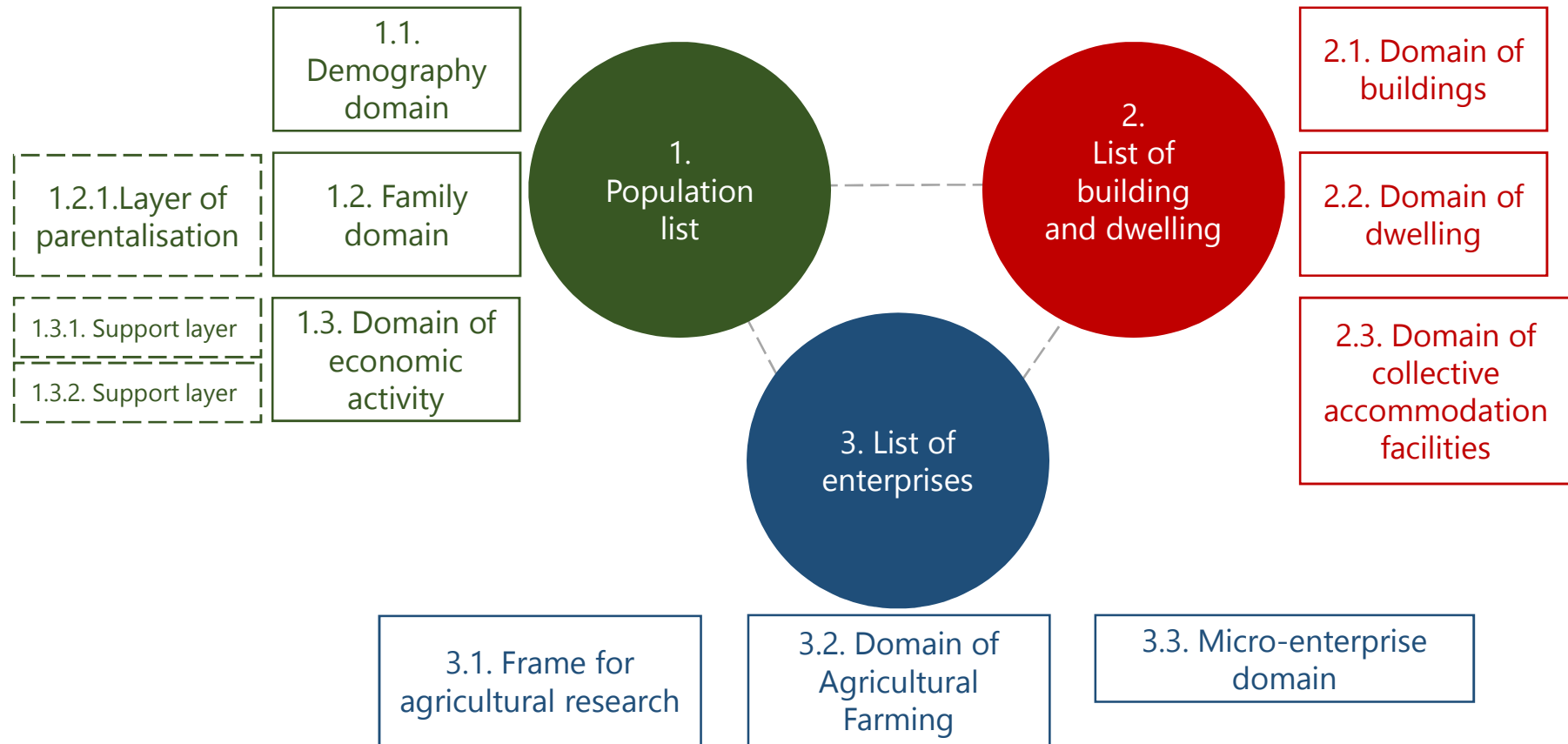
- main information (data provider, name, ISODS number, status, etc.);
- information on tables (name, number of variables, number of records);
- information on variables:
  - descriptive (name, type, category);
  - enumeration (quality indicators).

# Statistical Operations System

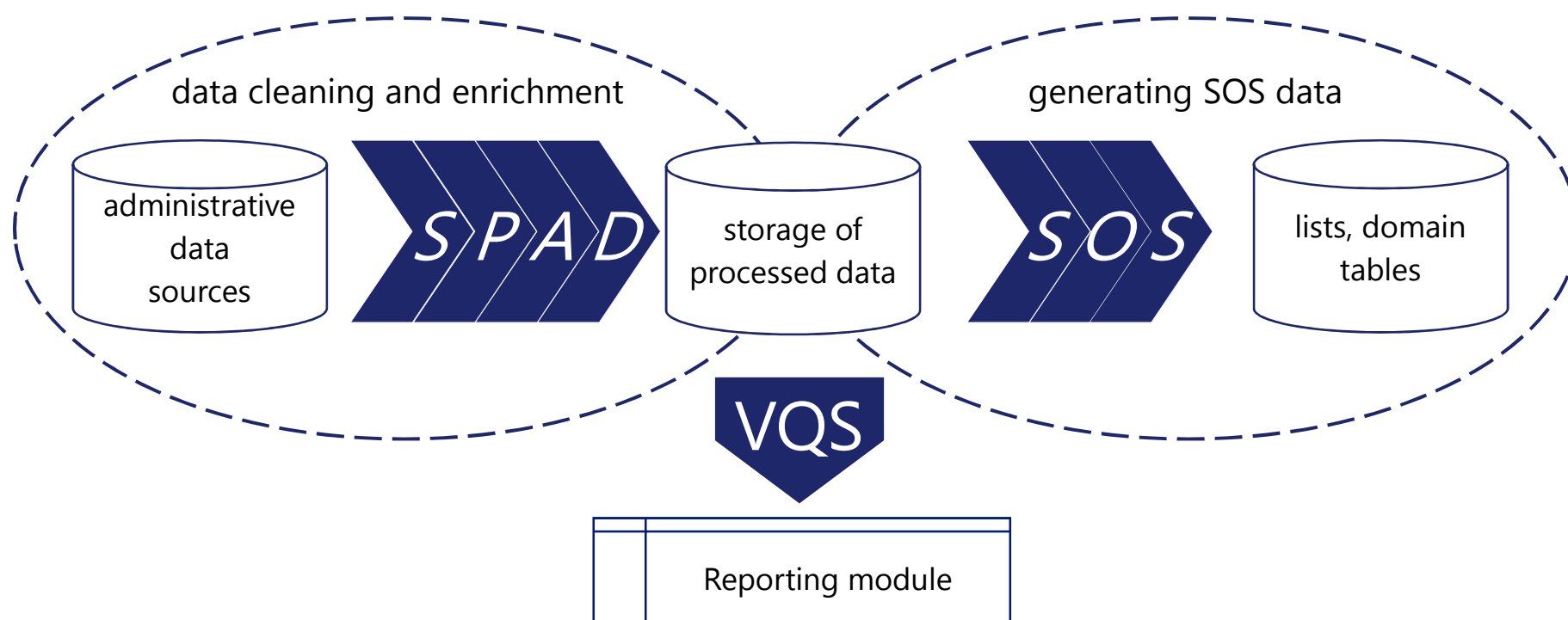
The Statistical Operating System is a comprehensive system, which – in an automated and dynamic way – based on input data and algorithms describing the method of their processing, allows to generate variables for a specific population, concentrated in separate thematic blocks.

The modularity of the final products assumed. The final form of the system will be obtained evolutionarily.

# Statistical Operations System – products



# Statistical Operations System — model of functioning





**Thank you for your attention!**