Statistics Poland

A modern statistical production process based on administrative registers

ICT Systems, Geostatistics and Census Department
Ewelina Wójcik

03.10.2022, Workshop on Statistical Data Editing

Integrated Microdata System

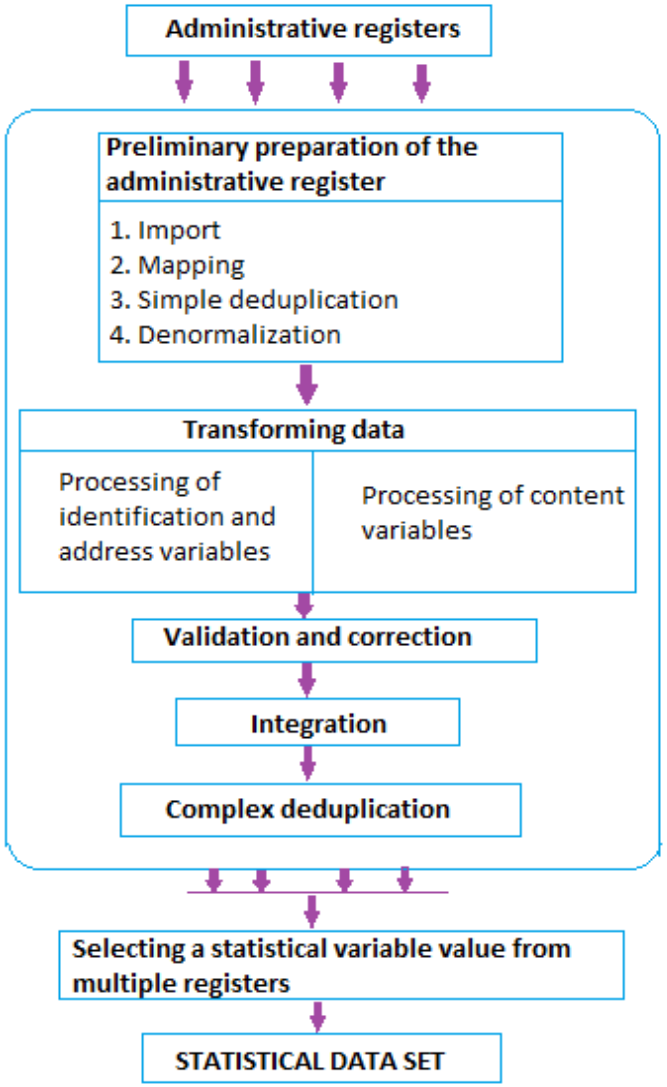
System Processing of Administrative Data (SPAD)

Variable Quality System (VQS)

Statistical Operations System (SOS)

Domain Data Sets (DDS)

System Processing of Administrative Data (SPAD)



System Processing of Administrative Data (SPAD)

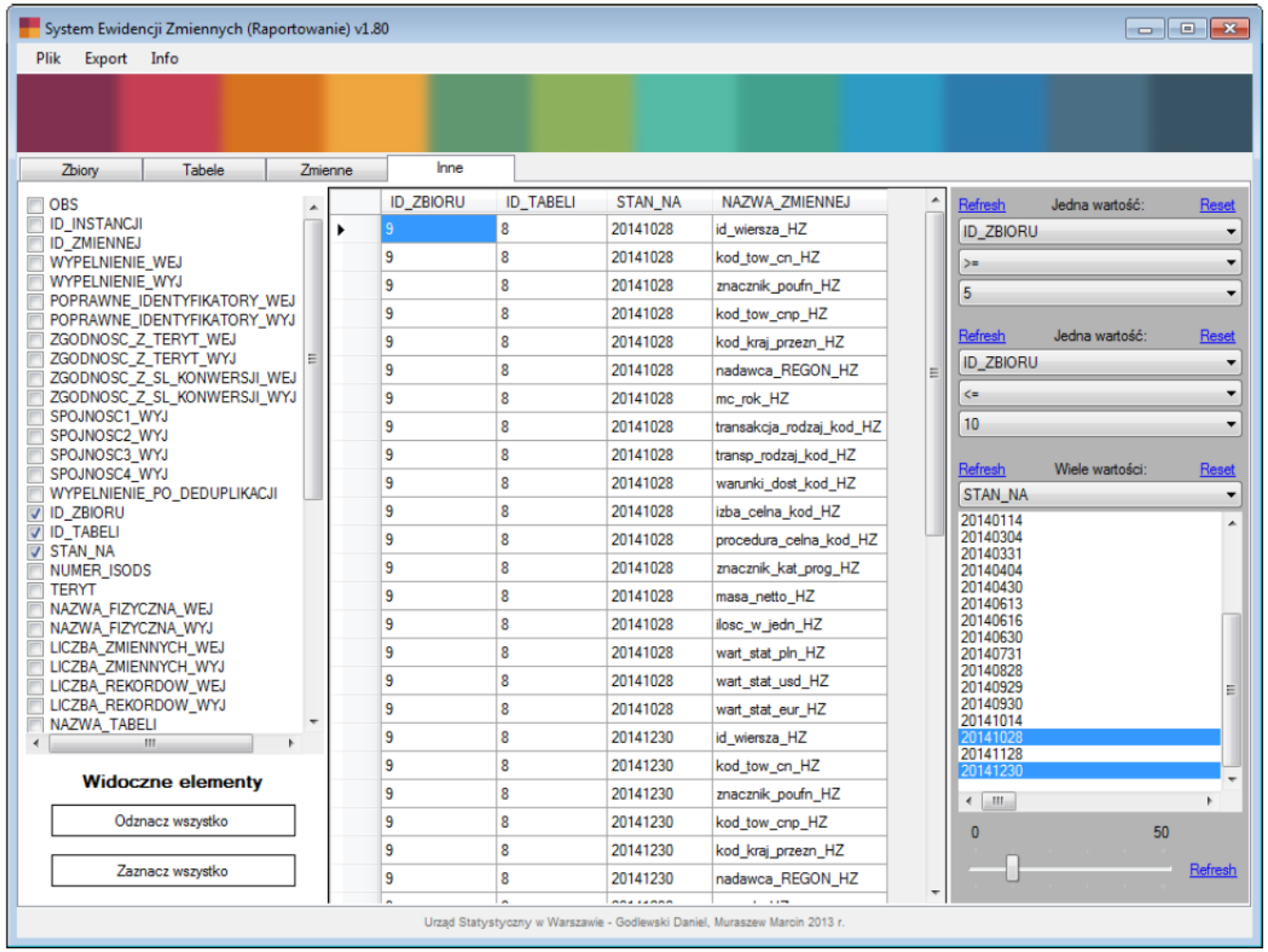
Example of contextual control of address data

VOIVODSHIPS	POWIATS	GMINA	LOCALITY	STREET	COHESION	FLAG
DOLNOŚLĄSKIE	LEGNICA ✓	LEGNICA ✓	LEGNICA ✓	3 MAJA ✓	to the street ✓	1
ŁÓDZKIE	ZGIERSKI ✓	ZGIERZ ✓	ZGIERZ ✓	_+|(*|+_ ✓	to the locality ✗	2
ŁÓDZKIE	ZGIERSKI ✓	ZGIERZ ✓	!@%$%#@ ✓	)*(&_+ ✗	to the gmina ✗	3
<>?:"%^*)^& 12	„:{}m_^ \=- ✗	><?:{ _ ✗	~!@# 23 ✗	<>?:{! /., ✗	none coherence ✗	4

contextual control - analysis of the consistency of address data from the voivodeship level to the street level in relation to the TERYT register, it consists to check the degree of integration of all address data with the TERYT register and assigning the appropriate value to the consistency flag.

The level of consistency is understood as the highest compliance of the sequence of variables: *voivodship, powiat, gmina, locality, street* with the entry standard in force in the national official register of the territorial division of the country (TERYT).

Variable Quality System (VQS)



- A unique system in the scale of Statistics Poland used to store information on collections obtained from administrative registers.
- It stores metadata describing administrative information systems in a multidimensional layout.
- It is used to calculate qualitative statistics for each variable obtained from administrative registers.
- Creates reports based on calculated qualitative statistics for variables and data sets, thanks to it's possible to study the degree of quality improvement obtained during the data transforming process.

Statistical Operations System (SOS)

The aim of the Statistical Operations System is to create a population register by integrating selected administrative sources. It is assumed that the target population consists of people - residents of Polish, i.e. the population (permanent residents and people coming from another place, including foreigners), who at a certain moment of time, it shall indicate in the administrative registers the place of residence in Poland.

Stages of construction a population register :

- collection and merging all PESEL numbers from the selected administrative registers- single selection of the identifier according to the set hierarchy defined by the algorithm,
- exclusion of units that do not meet the definition of the target population,
- transfer of values from selected, most reliable administrative sources (first name, middle name, surname, citizenship and address data).

Domain Data Sets (DDS)

Domain Data Sets are thematic blocks, which structure is based on the created population register, thematic blocks and shared variables.

Domain Data Sets are characterized by:

- uniqueness of thematic blocks,
- subjective scope, which covers the entire population or subpopulation,
- objective scope, which covers the subject matter that is the basis for the observation of social, economic and spatial phenomena,
- high complexity of the algorithms generating the resulting statistical variables,
- possibility of sharing result variables between domains.

Domain Data Sets (DDS)

The algorithms generating the resulting statistical variables are characterized by high complexity.

The most commonly used algorithms assume:

- moving data between tables using the linking key,
- processing of a given key group of many records in order to determine the best value (in several sources),
- using reference dictionaries for data (according to the linking key) to verify the existing data or their conversion,
- using reference dictionaries for data (according to the linking key) in order to provide more detailed information (e.g. adding names to the citizenship codes),
- processing within one table and one record (calculating features from several others for a given unit),
- transposing data from many records into one record with many fields (possible depending on the scope of data and the efficiency of the environment),
- loop processing (looking for the best value until a set threshold is reached or no data is available for the next iteration),
- one-time transmission of artificial keys to facilitate subsequent multiple connection.

What are the benefits of a modern statistical production process?

1. Data can be used as a direct source for research, imputation, estimation and supplementation of data in the study, can be the basis for the creation and update of frames for surveys.
2. Efficiency - quick presentation of results, direct access to the resulting data.
3. Increasing the use of administrative data in experimental and statistical research through shortened and easy access.
4. Production of a wide range of data from various thematic areas, at low levels of territorial aggregation, including the one-kilometer grid.
5. Greater flexibility to the growing information needs of citizens and providing information adequate to the needs by creating complex, multi-domain and multi-faceted statistical studies describing in a comprehensive manner socio-economic phenomena.

Statistics Poland

Thank you for your attention!