Statistics Poland

Data imputation for the purposes of statistical research with the use data from administrative registers

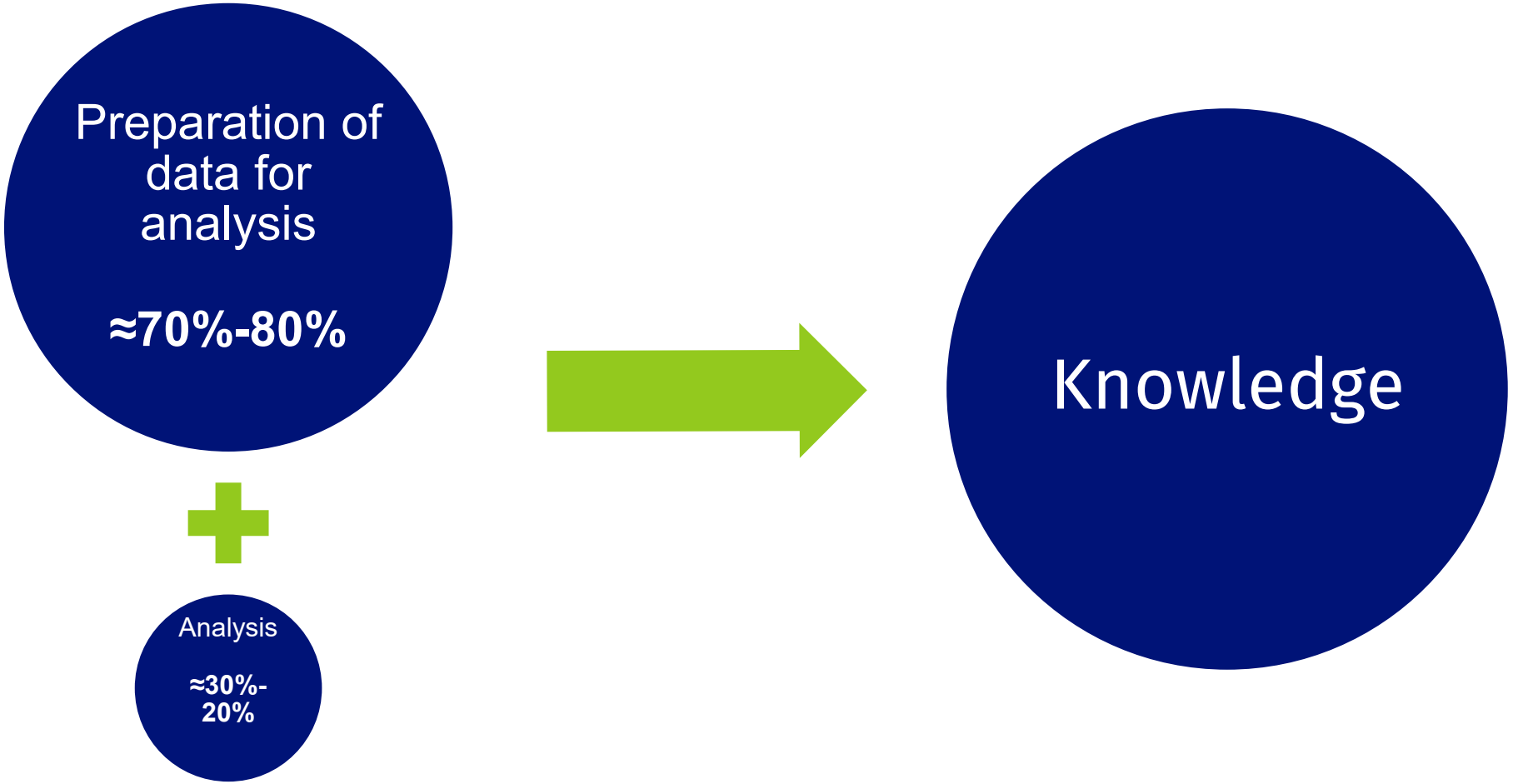
Paweł Murawski
Consultant

October 5, 2022

Outline

1. Introduction
2. Administrative registers
3. Practical application – Research of the use of information and communication technologies – Information Society in Poland (SSI)
4. Conclusions

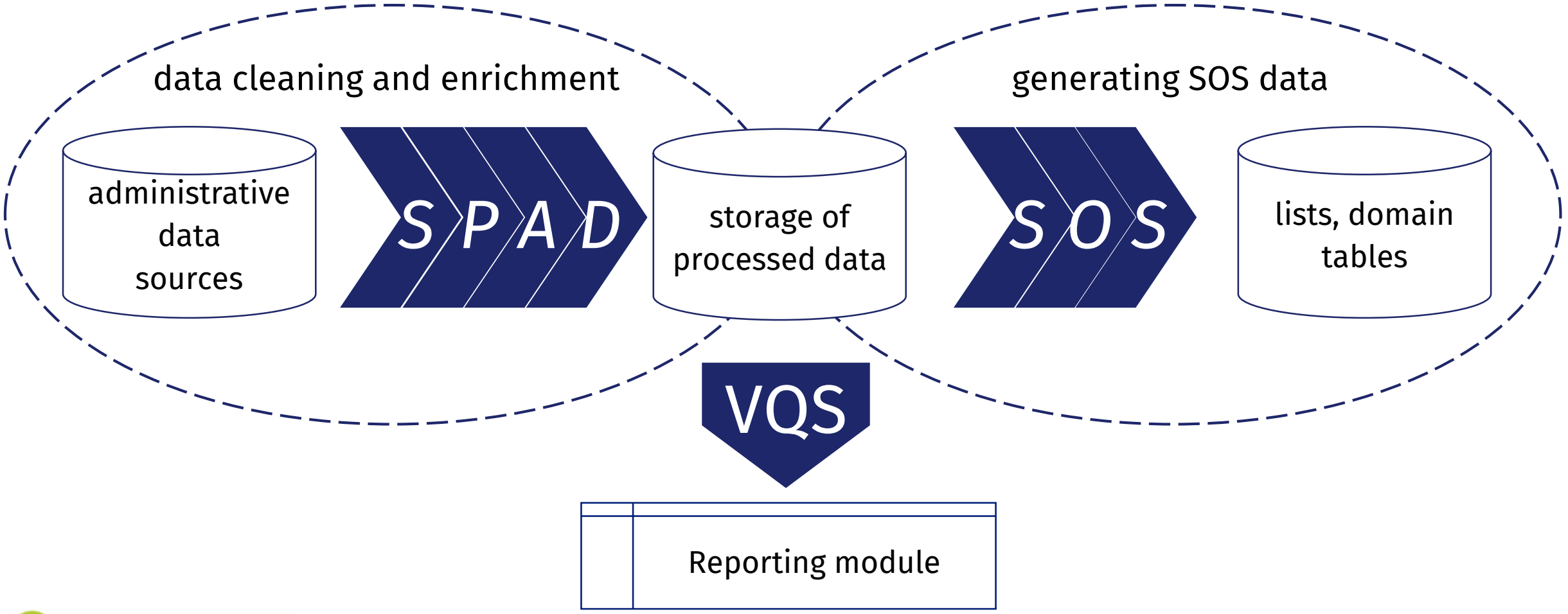
Introduction



Administrative registers

- Data of varying quality
- Data in various formats
- Data not in line with statistical standards
- Data created for various purposes
- Information for different levels of territorial aggregation
- Information describing different populations

Statistical Operations System — model of functioning



Administrative registers

- Quality of registers
- Data integration
- Data availability

Practical applicatation

Research on the use of information and communication technologies – Information Society in Poland 2019 (SSI)

The aim of the survey is to collect information on the use of the Internet by households

Practical applicatation

The stages of the research:

- Import data
- Generating the integration key
- Standarization of address data
- Generating a list unique values
- Include relevant information features

Practical applicatation – imputation data

- The hot deck imputation based on the nearest neighborhood method was used for the given criteria for distinguishing imputation classes - variables WOJ and KLM. Generating the integration key

- In this method, the imputed value for the feature y for the kth object is

$$\hat{yk} = yl(k)$$

Statistics Poland

Practical applicatation – results

- According to the given guidelines, the possibility of connecting to the Internet was examined - the possibilities of Internet service providers, not the actual connection used.
- It was assumed that due to the incompleteness of the set with UKE and the imputation of non-integrated records, all address points will be shown as "with the possibility of access to the Internet". Therefore, there will be no addresses "without the possibility of access to the Internet".
- The number of surveyed households amounted to 12,635,252 and is consistent with the number of households with people aged 16-74 from the SSI 2017 survey.
- There were no households with the possibility of access to the Internet through a permanent connection, only a narrowband one.
- There were no households with the possibility of accessing the Internet through a narrowband only mobile connection.

Practical applicatation – results

- There were no households without broadband Internet access.
- It was found that the number of households with the possibility of access to the Internet through a permanent connection is equal to the number of households with the possibility of accessing the Internet through a permanent broadband connection.
- It was found that the number of households with the possibility of access to the Internet through a mobile connection is equal to the number of households with the possibility of accessing the Internet through a mobile broadband connection.
- It was found that the number of households with the possibility of access to broadband Internet is equal to the number of all surveyed households.
- Internet access in households. Table in layout according to the class of the place of residence

Conclusion

- A similar pattern of action can be applied to other similar studies. Thanks to this, both the process of calculating the results and imputation of data will be standardized, which contributes to increasing the effectiveness of the activities carried out, accelerating the implementation of processes and improving the quality of the resulting data. It also lowers the costs of carrying out the research due to the fact that the test sample does not have to be increased in order to maintain the quality of the results at a sufficiently high level.
- Due to the availability of high-quality data from administrative sources, data imputation may be more precise, which will enable output data to be generated at the lowest aggregation levels.

Statistics Poland

Thank you.

Email:

Paweł Murawski: p.murawski@stat.gov.pl