

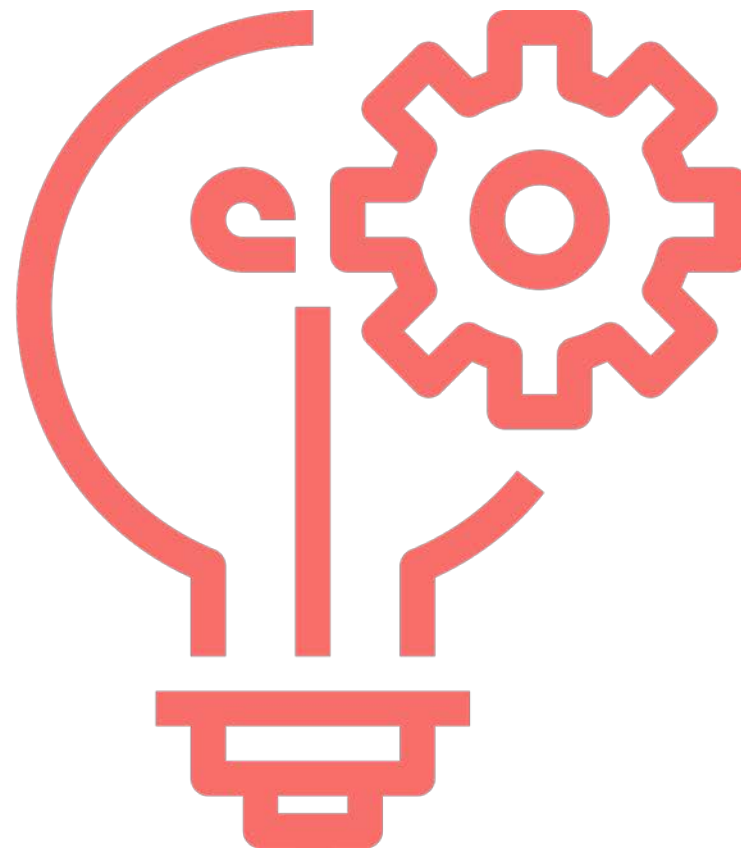
Technical aspects
regarding the application of
statistical methods for
estimation based on
census data

Ionela-Roxana Glăvan

Ana-Maria Ciuhu

Department of Social Statistics,

National Institute of Statistics, Romania



- Combining datasets
- Steps for data pre-processing
- Statistical methods
- Limitations of using multiple data sources for estimations

Census
Data
RECIPIENT

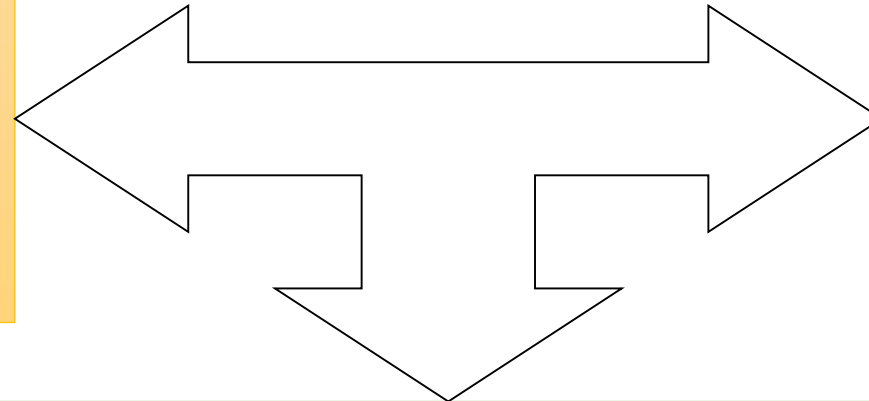
Needs to be improved

Administrative
Data
DONOR

Will be explored and used to
impute the units of the Recipient

MATCHED
Dataset

Final dataset containing data from Recipient and Donor



1. Identify the need for enhance data quality in the Recipient

2. Establish data sources for Donor

3. Harmonization of donor file(s)

3.1. Evaluate donor variables

3.2. Reconciliation of donor file(s)

3.3. Identify imputation variables

3.4. Create Donor final version

Relationship with recipient variables

Accuracy of data (quality)

Alternative data sources for donor

Data cleaning

Standardization - harmonization of variables

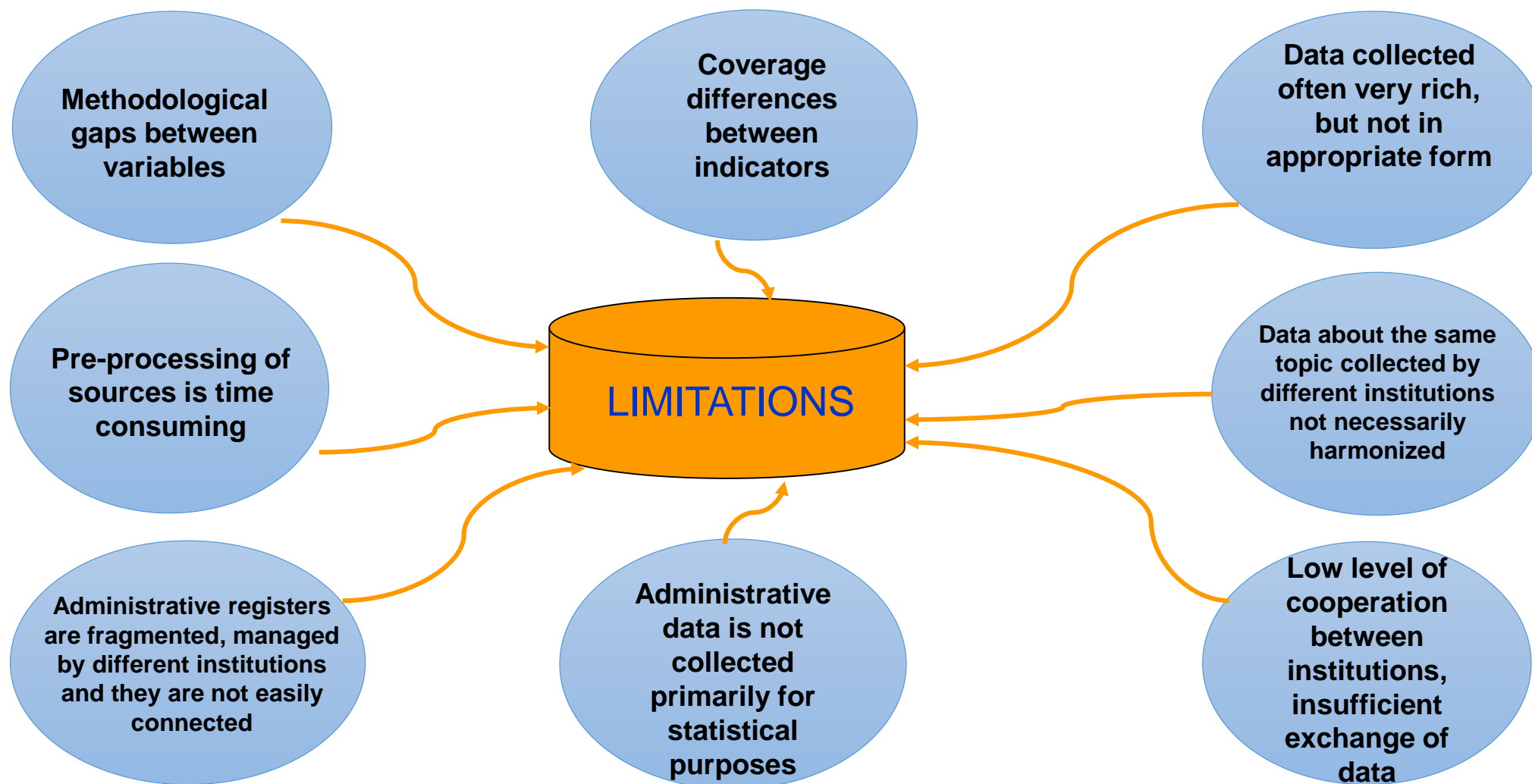
Classification - harmonization of classifications

Imputation

Matching

- Imputation is the process used to determine and assign replacement values for missing, invalid or inconsistent data.
- Create profiles
- Statistical techniques for imputation:
 - ❑ **Unit imputation** - is typically used to treat unit nonresponse
 - ❑ **Item imputation** - item imputation is used when some but not all the desired information is collected
 - ✓ **Single imputation** - the missing value is replaced by a **single** value
 - **Hot-deck method** - missing value is imputed from a randomly selected similar record; the information donors come from the same dataset as the recipients
 - **Cold-deck method** - by contrast, selects donors from another dataset
 - **Mean substitution** - replacing any missing value with the mean of that variable for all other cases
 - **Regression based methods** - available information for complete and incomplete cases is used to predict the value of a specific variable.
 - ✓ **Multiple imputation** - the method imputes **several** values for each missing value

- Condition: common or very similar characteristics
- Process of bringing together data (represented in records) relating to units and spread over two data sets, based on common or very similar characteristics in the form of primary or object characteristic values.
- The purpose of matching is to find objects that have a certain type of relation to one another. Those objects can be found within the same data set or in different data sets concerning different object types.
- Matching Census data to administrative sources could provide new insights, both empirical and theoretical, into the relationship between different administrative datasets.



Thank you for your attention!