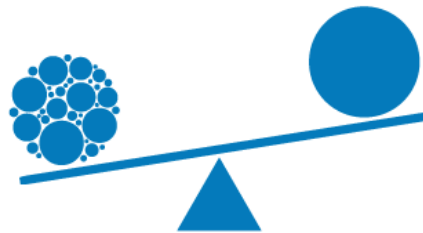# Jean Monnet Chair
Analysis of European Data by Small Area Methods

Lecture 1: European data on poverty and living conditions, policy makers and data needs, the survey data and the small area estimation problem
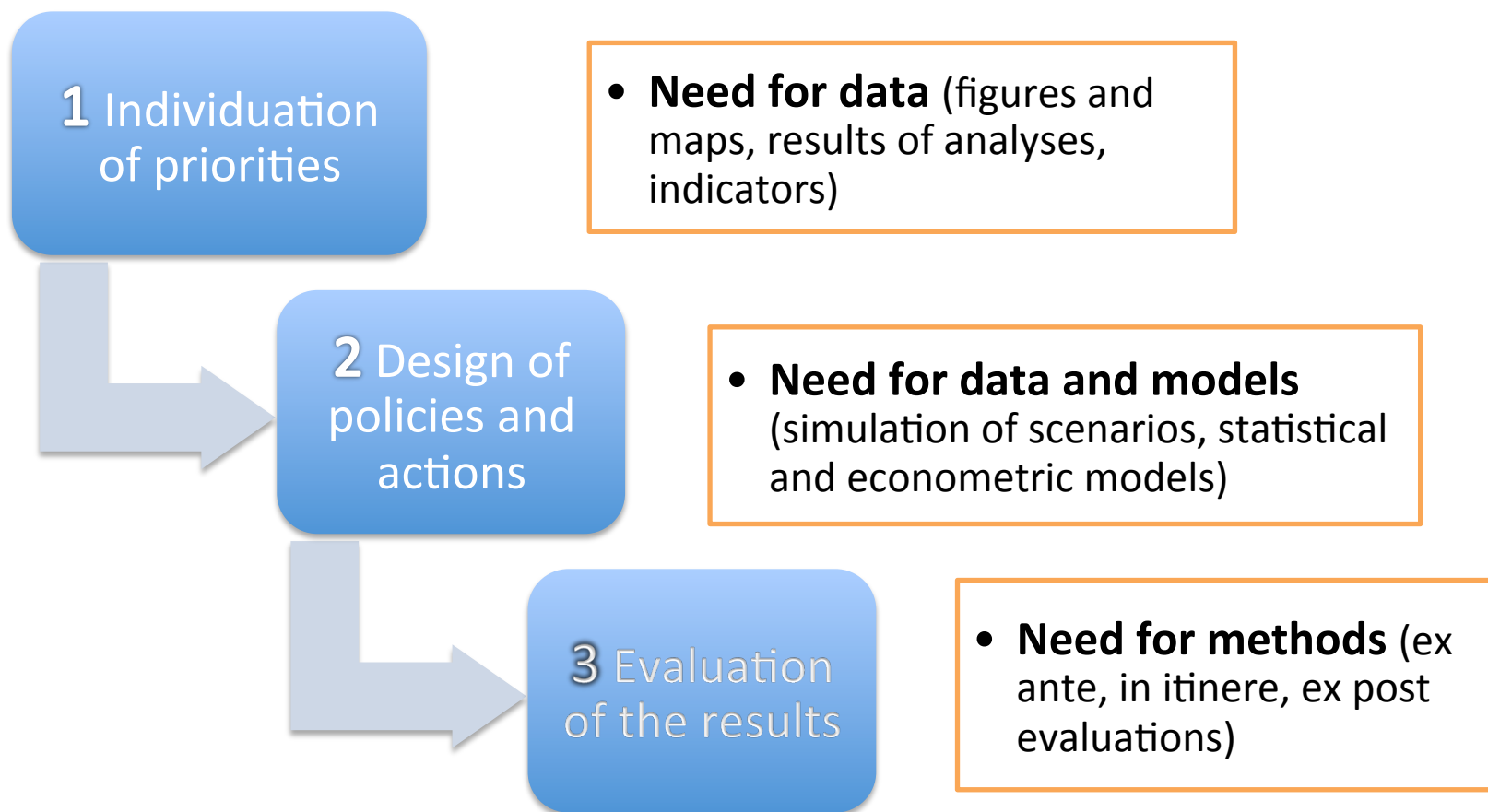
http://sampleu.ec.unipi.it

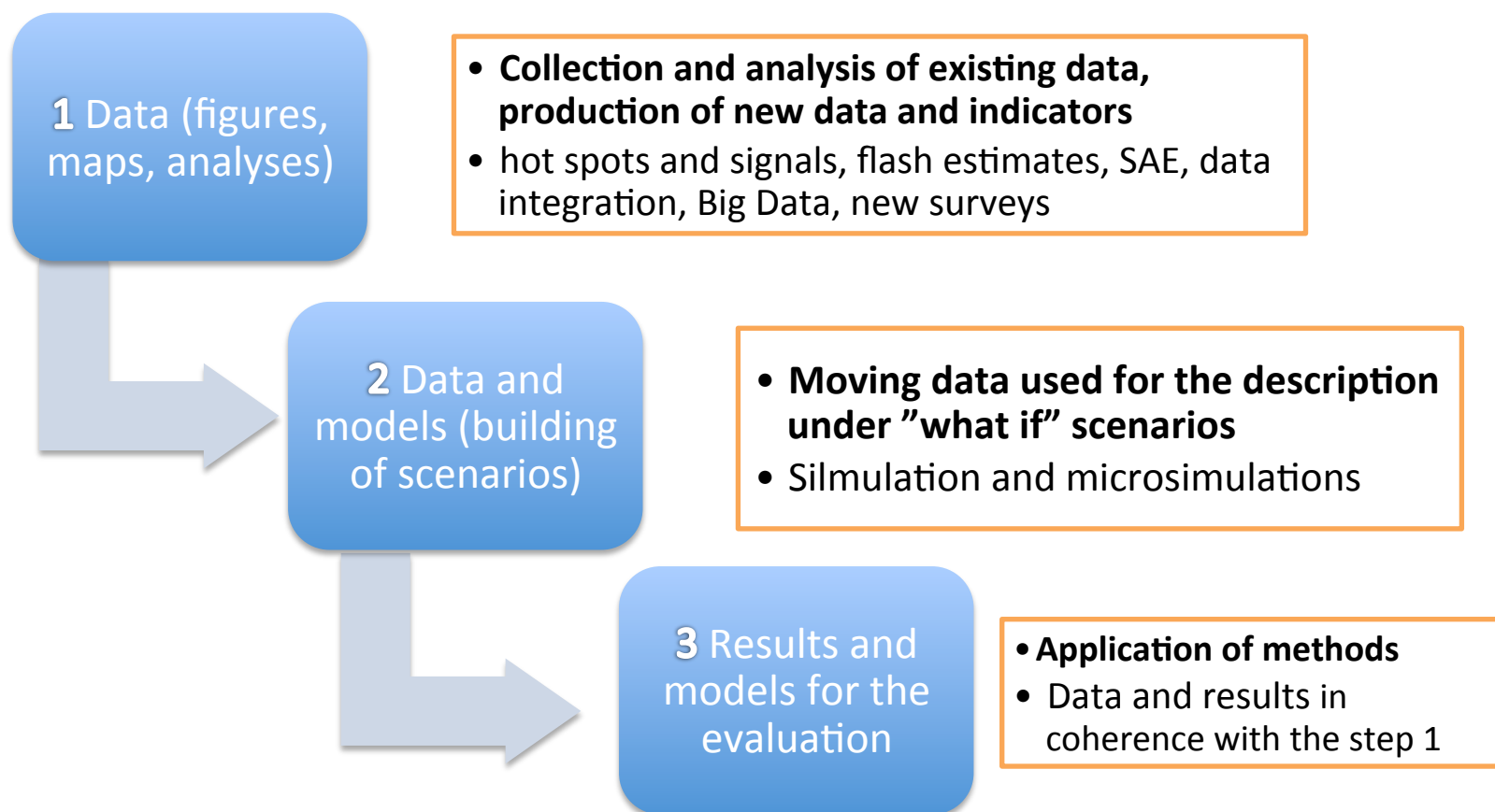# Local statistics for decision-making on well-being and vulnerability

Themes

- **Policies** against poverty and social exclusion
- Poverty **Indicators** and **Measures**
- **Sources** : Integration of administrative data, Small Area Estimates, Big data
- Policy **Evaluation**

Decision making

1 Individuation of priorities

- **Need for data** (figures and maps, results of analyses, indicators)

2 Design of policies and actions

- **Need for data and models** (simulation of scenarios, statistical and econometric models)

3 Evaluation of the results

- **Need for methods** (ex ante, in itinere, ex post evaluations)

Need for a system to produce *local*, *meaningful*, *flash* data and indicators on poverty and vulnerabilities, which are *understandable* and *useful* to policy making**.**

The decision on "what" data and "how" collected is not neutral, but it is a map of reality to be defined given the goal of the policy maker….

…sometimes it is important to have a signal and not an "error-free" estimate!

see: https://inclusivegrowth.be/about-ingrid

Data on poverty by Eurostat

There are many sources of data

Eurostat collects data from an harmonized set of current surveys

- European Survey on Income and living conditions

- Household Budget survey

- Labour Force survey

 sample surveys conducted yearly (LFS every trimester) in Member States

iPad 08:59 100%

ec.europa.eu

Impossibile aprire pagina | IL BENESSERE EQUO... | Statistica RM - issuu | ⊗ Overview - Eurostat | EU statistics on income... | ec.europa.eu/eurostat/...

Your key to European statistics

Type a keyword, a code, a title...

European Commission > Eurostat > Income and Living Conditions > Overview

| News | Data | Publications | About Eurostat | Help |

OVERVIEW

Statistics on **income, social inclusion and living conditions** cover objective and subjective aspects of these themes in both monetary and non-monetary terms for both households and individuals. They are used to monitor the Europe 2020 strategy in particular through its poverty reduction headline target.

The main source for the compilation of statistics on income, social inclusion and living conditions is the EU-Statistics on Income and Living Conditions (EU-SILC) instrument. It collects comparable multidimensional micro-data on:

- income
- poverty
- social exclusion
- housing
- labour (see also Labour market)
- education (see also Education and training)
- health (see also Health)

Additional data on living conditions can be found in other domains, notably population statistics, health statistics, education and training statistics and labour market statistics.

What is EU-SILC?

More

EU-SILC sample size

More

Target population

More

ec.europa.eu

Sign In | Register 👤

eurostat
Your key to European statistics

Legal notice | 📶 RSS | Cookies | Links | Contact   English ⌄

Type a keyword, a code, a title... 🔍

European Commission > Eurostat > Income and Living Conditions > Data > Main tables

News	Data	Publications	About Eurostat	Help

INCOME AND LIVING CONDITIONS

Statistics illustrated

Overview

– Data

    MAIN TABLES ❯

    Database

    Ad-hoc modules

– Methodology

    Main concepts and definitions

    Classifications

    List of variables

Legislation

– Quality

    EU quality reports

    National quality reports

    Questionnaires

– Publications

    Statistical books

    Statistics in focus

MAIN TABLES

⊟ 📂 Income and living conditions (t_ilc) Ⓜ

    ⊞ 📁 People at risk of poverty or social exclusion (Europe 2020 strategy) (t_ilc_pe)

    ⊞ 📁 Income distribution and monetary poverty (t_ilc_ip)

    ⊞ 📁 Living conditions (t_ilc_lv)

    ⊞ 📁 Material deprivation (t_ilc_md)

eurostat
Your key to European statistics

English ∨

Type a keyword, a code, a title...

News	Data	Publications	About Eurostat	Help

INCOME AND LIVING CONDITIONS

Statistics illustrated

Overview

− Data

   MAIN TABLES ❯

   Database

   Ad-hoc modules

− Methodology

   Main concepts and definitions

   Classifications

   List of variables

Legislation

− Quality

   EU quality reports

   National quality reports

   Questionnaires

− Publications

   Statistical books

   Statistics in focus

MAIN TABLES

⊟ 📂 Income and living conditions (t_ilc) M

   ⊟ 📂 People at risk of poverty or social exclusion (Europe 2020 strategy) (t_ilc_pe)

      📊 ZIP People at risk of poverty or social exclusion (t2020_50) M ⓘ

      📊 ZIP People at risk of poverty or social exclusion by NUTS 2 regions (tgs00107) M ⓘ

   ⊞ 📁 Income distribution and monetary poverty (t_ilc_ip)

   ⊞ 📁 Living conditions (t_ilc_lv)

   ⊞ 📁 Material deprivation (t_ilc_md)

Pad 🔋 09:05 🌙 98% 🔋

ec.europa.eu

Impossibile… | IL BENESS… | Statistica R… | ⊗ Main tab… | ec.europa.… | eur-lex.eur… | ec.europa.… | ec.europa.… | EU statistic… | ec.europa.…

Sign In | Register 👤

eurostat
Your key to European statistics

Legal notice | 📶 RSS | Cookies | Links | Contact   English ∨

Type a keyword, a code, a title...   🔍

European Commission > Eurostat > Income and Living Conditions > Data > Main tables

News	Data	Publications	About Eurostat	Help

INCOME AND LIVING CONDITIONS

Statistics illustrated

Overview

– Data

   MAIN TABLES   ❯

   Database

   Ad-hoc modules

– Methodology

   Main concepts and definitions

   Classifications

   List of variables

Legislation

– Quality

   EU quality reports

   National quality reports

   Questionnaires

– Publications

   Statistical books

   Statistics in focus

MAIN TABLES

⊟ 📂 Income and living conditions (t_ilc) Ⓜ

  ⊟ 📂 People at risk of poverty or social exclusion (Europe 2020 strategy) (t_ilc_pe)

    ❤ ZIP People at risk of poverty or social exclusion (t2020_50) Ⓜ ⓘ

    ❤ ZIP People at risk of poverty or social exclusion by NUTS 2 regions (tgs00107) Ⓜ ⓘ

  ⊟ 📂 Income distribution and monetary poverty (t_ilc_ip)

    ⊞ 📁 Monetary poverty (t_ilc_li)

    ⊞ 📁 Monetary poverty for elderly people (t_ilc_pn)

    ⊞ 📁 In-work poverty (t_ilc_iw)

    ⊞ 📁 Distribution of income (t_ilc_di)

  ⊟ 📂 Living conditions (t_ilc_lv)

    ⊞ 📁 Population structure (t_ilc_lvps)

    ⊞ 📁 Health and labour conditions (t_ilc_lvhl)

    ⊞ 📁 Housing conditions (t_ilc_lvho)

    ⊞ 📁 Childcare arrangements (t_ilc_ca)

  ⊟ 📂 Material deprivation (t_ilc_md)

    ⊞ 📁 Material deprivation by dimension (t_ilc_mddd)

    ⊞ 📁 Housing deprivation (t_ilc_mdho)

    ⊞ 📁 Environment of the dwelling (t_ilc_mddw)

Data quality

- Quality is in the eye of the beholder!

- What does statistical quality mean?

Statistical quality

Goal:

how to assure *quality* through effective and appropriate design or redesign of a *statistical project* or program from inception through to data evaluation, dissemination and documentation.

Statistical quality

Concept paper:

no *standard definition* of quality for Official Statistics: quality embodies a broad notion of "fitness for use"

see: Statistics Canada, ABS, ONS, Eurostat quality guidelines

Statistical quality

Dimensions of quality:
relevance,
accuracy,
timeliness,
accessibility,
interpretability,
coherence
"ghost" dimension: validness

Statistical quality

The *validity* of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure.

Statistical quality

The *relevance* of statistical information reflects the degree to which it meets the real needs of clients.

Statistical quality

The *accuracy* of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure (given the measurement is valid).

It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components.

Official Local data in Europe

- What does local mean? We need to measure poverty where it matters, in the places where people live

- "local" is an attribute not defined once for ever!

The **Classification of Territorial Units for Statistics (NUTS)**;
French: *Nomenclature des unités territoriales statistiques*) is a geocode standard for referencing the subdivisions of countries for statistical purposes

There are three levels of NUTS defined, with two level of LAUs (Local Administrative Units) below.

 Note that not all countries have every level of division, depending on their size. One of the most extreme cases is Luxembourg, which has only LAUs; the three NUTS divisions each correspond to the entire country itself.
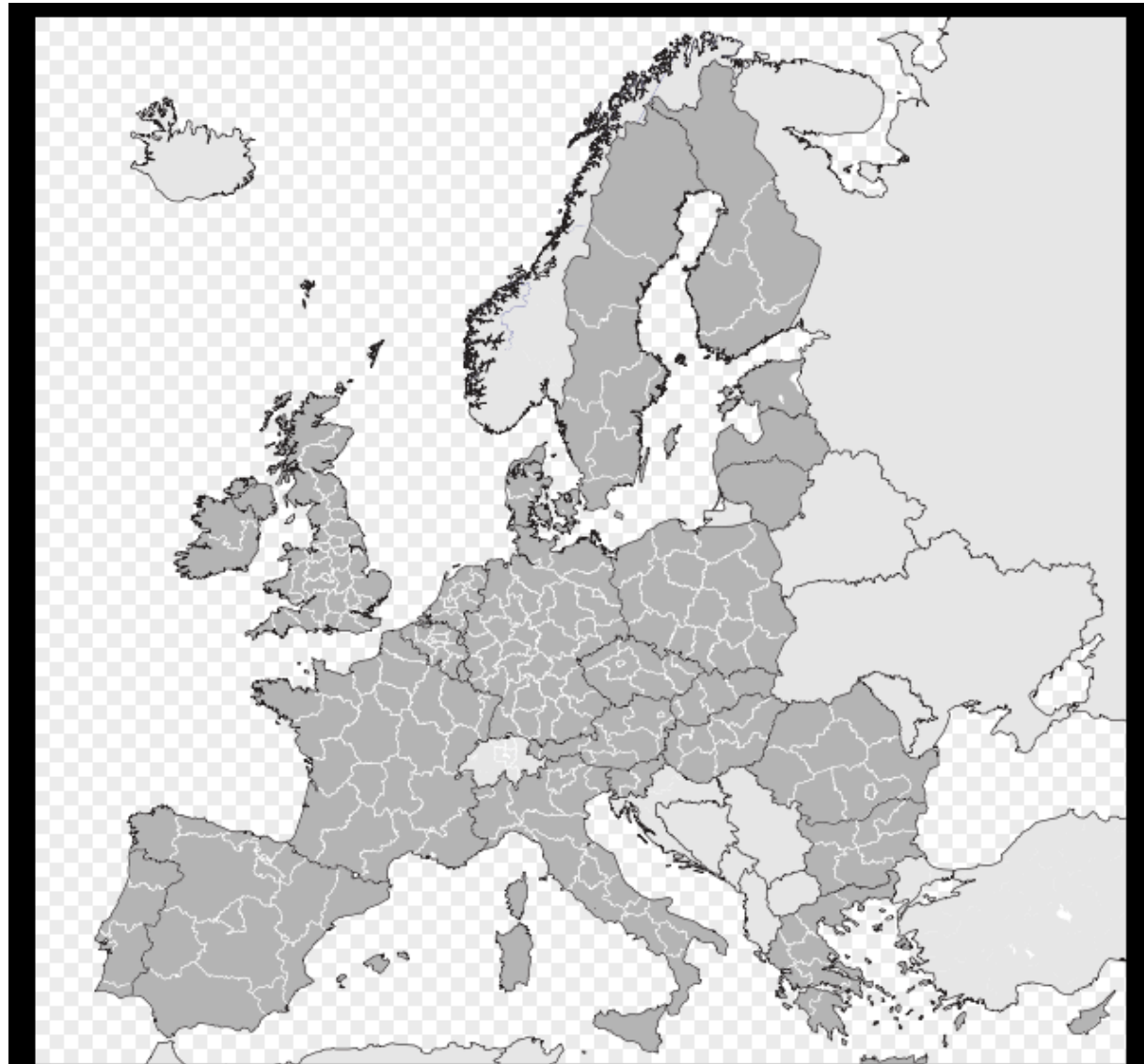
NUTS1

Italy: groups
of Regions 5

Germany:
States 16
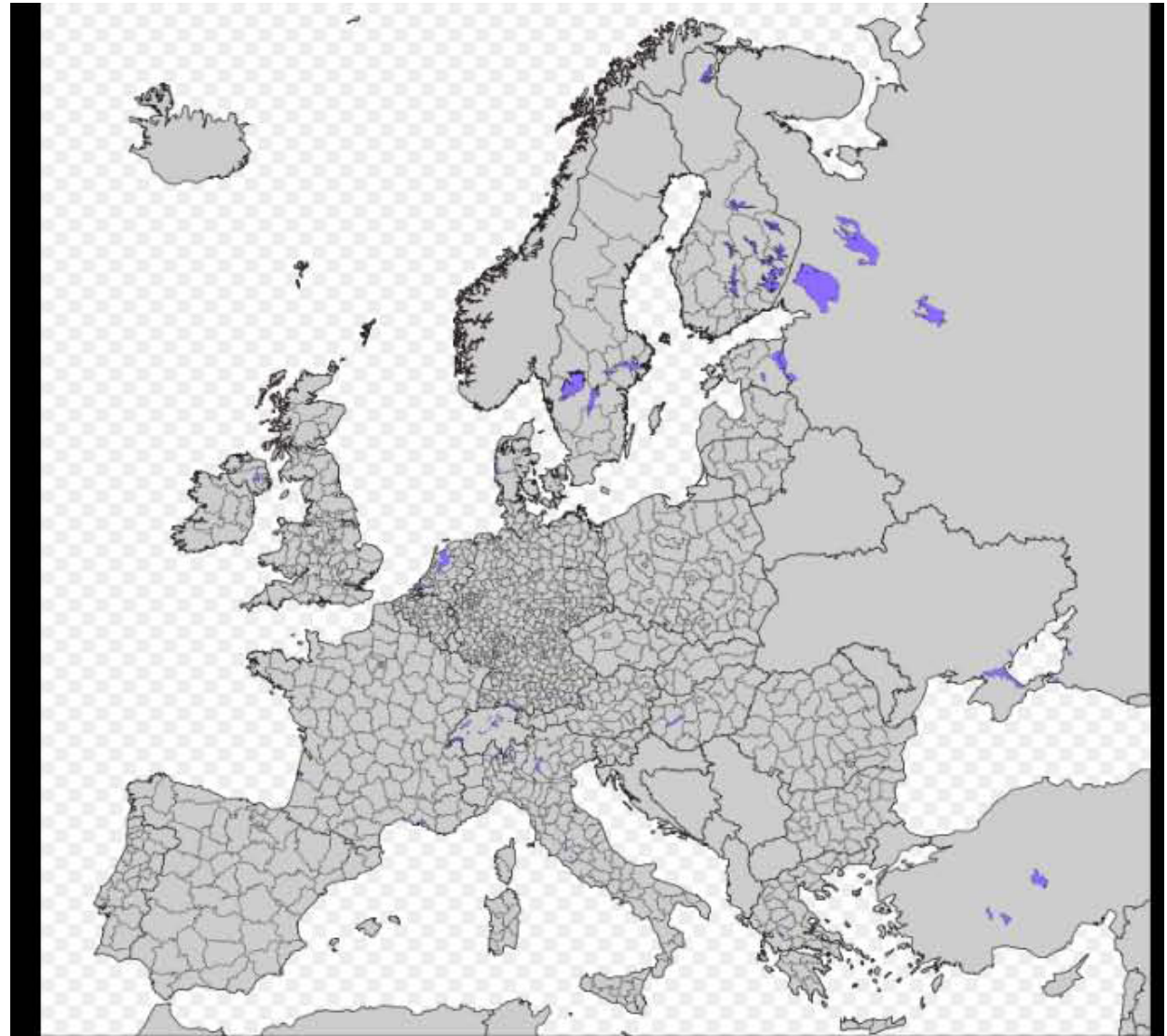(Bundesland)

NUTS2

Italy:
Regions 21

Germany:
Government
regions 39
(*Regierungbezirk*)

NUTS3

Italy:
Provinces
110

Germany:
Districts 429
(*Kreis*)

Local areas - NUTS3 - "small areas"

- Eurostat publishes estimates at NUTS1 and NUTS2 level

- These are direct estimate: estimate coming directly from a sample survey – design-based estimate from sample data

- Small area = domain of interest, for which the sample size is not adequate to produce reliable direct estimates – in EU lower than NUTS2 level

Estimation: scope and purpose

Estimation is a process that approximates unknown population parameters using only that part of the population that is included in a sample.

Examples of parameters

- simple descriptive statistics: totals, means, ratios and percentiles,

- complex statistics: **poverty indicators**

- analytical statistics: regression coefficients

Estimation: scope and purpose

Measures of *precision* are usually computed to evaluate the quality of a population parameter estimate and to obtain valid inferences.

SAE: choice of estimation method is crucial

Estimation: direct estimators

The estimation method and the sampling design determine the properties of the sampling error.

SAE is based on direct estimates

Estimation: direct estimators

- The basic design-consistent *Horvitz-Thompson estimator* is the most natural estimator to use if there is no auxiliary information available at the estimation stage.

SAE can be based on direct estimates where sampling weights, calibration, re-weighting had their effect

Estimation: direct estimators

- Proper estimation conforms to the sampling design. Sampling weights are incorporated in the estimation process (stratification, clustering, and multi-phase or multi-stage information)

SAE vs direct estimates: with or without sampling weights?

Fabrizi et al Biometrical Journal (2012)

Estimation: direct estimators

Use auxiliary data whenever possible to improve the reliability of the estimates.

Evaluate the use of the auxiliary data.

SAE vs direct estimates: simple HT estimator or ratio/regression estimators?

Estimation: direct estimators

Stat Canada suggestion 1:

"Whenever auxiliary data are available for sample units, together with known population totals for such data, consider using calibration estimation so that the weighted auxiliary data add up to these known totals. This may result in improved precision and lead to greater consistency between estimates from various sources."

Estimation: domain estimators

Stat Canada suggestion 2:

"When the original classification of sampling units has changed between the time of sample selection and estimation, consider *domain estimation* so that the new classification is reflected in the estimates. Domain estimation refers to estimation for specified subsets of the population (or *domains*) of interest."

Estimation: small area estimators

Stat Canada suggestion 3:

"Incorporate the requirements of small domains of interest at the sampling design and sample allocation stages (Singh, Gambino and Mantel, 1994). If this is not possible at the design stage, or if the domains are only specified at a later stage, consider special estimation methods (small area estimators) at the estimation stage. These methods "borrow strength" from related areas (or domains) to minimize the mean square error of the resulting estimator (Platek et al., 1987; Ghosh and Rao, 1994; Rao, 1999)."

Statistical quality

- major sources of error that potentially cause inaccuracy: coverage, sampling, nonresponse, response.

  SAE design-based (direct estimators)

- other sources: previous ones + specification of the model that *borrows strength* from related areas

  SAE model-assisted, model-based   (predictors)

Statistical quality

The *timeliness* of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available.

SAE must be timely and cost-effective *flash estimates*!!! *Giusti et al Survey Research Methods (2012)*

Statistical quality

The *accessibility* of statistical information refers to the ease with which it can be obtained from the Agency.

SAE it often offered through maps *(poverty mapping)*: questionable medium of accessibility

Statistical quality

The *interpretability* of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately.

  SAE requires special metadata (*model fitting, geography used*)

Statistical quality

The *coherence* of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time

SAE requires special attention on this dimension
(*calibration, benchmarking, harmonization with other estimates*)

Statistical quality

Two leading ideas:

1) These dimensions of quality are overlapping and interrelated.

2) There is no general model that brings them together to optimize or to prescribe a level of quality.