



Data integration as a tool for benchmarking

Monica Pratesi

Outline

- ❖ A reflection on **Quality** of statistics: Eurostat criteria and benchmarking
- ❖ In concrete: **examples**
 - Integrated System of Registers (SIR) and Census counts: *signs of life*
 - Energy Consumption Survey and Citizen Generated Data

- ❖ **Lessons** learnt
 - revision of the TSE framework
 - Citizen Science experiments

Quality: is it in the eye of the beholder?

Eurostat defines the quality of statistics with reference to **six** criteria

- 1. Relevance:** an inquiry is relevant if it meets users' needs. The identification of users and their expectations is therefore necessary. In the European context, domains for which statistics are available should reflect the needs and priorities expressed by the users of the European Statistical System (completeness).
- 2. Accuracy:** accuracy is defined as the closeness between the estimated value and the (unknown) true value.
- 3. Timeliness** and punctuality in disseminating results: most users want up-to-date figures which are published frequently and on time at pre-established dates.
- 4. Accessibility** and clarity of the information: statistical data have most value when they are easily accessible by users, are available in the forms users desire and are adequately documented
- 5. Comparability:** statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values taken by the characteristic across space and time. The comparability component stresses the comparison of the same statistics between countries in order to evaluate the meaning of aggregated statistics at the European level.
- 6. Coherence:** when originating from a single source, statistics are coherent in that elementary concepts can be combined reliably in more complex ways. When originating from different sources, and in particular from statistical surveys of different frequencies, statistics are coherent in so far as they are based on common definitions, classifications and methodological

Quality: benchmarking

Benchmarking: the act of measuring the quality of something by comparing it with something else of an accepted standard (a level of quality)

1. Benchmarking refers to the case where there are two sources of data for the same target variable, with different frequencies,
2. Benchmarking is a method of using auxiliary information to adjust the sampling weights used in an estimation process
3. Benchmarking is a standard or point of reference by which data can be compared. The process of aligning the sample estimate with the universe estimate is known as benchmarking

Permanent PHC

population counts at the municipal level by age, sex and citizenship

Permanent Population and Housing Census and the Pandemic in 2020

The Permanent Population and Housing Census (PPHC) has been designed according to the modernization programme of the Italian National Institute of Statistics (Istat), which places the integrated system of statistical registers at the core of statistical production. The role of field surveys in this system is to support registers, in the broad sense of assessing their quality and to add information that is missing, incomplete or of insufficient quality. This allows the yearly availability of detailed census statistics

cancelation of the field surveys for the 2020 wave (pandemic)

In order to predict population counts at the municipal level by age, sex and citizenship, a process integrating available data from the past waves and administrative “signs of life” was set up to establish deterministic criteria applied to individual records in RSBI (Basic Statistical Register of Individuals) . This obligatory push towards use of more administrative data has called for a rethinking of the statistical framework for the quality assessment of the estimation processes of the PPHC and, more generally, of the PPHC

Permanent PHC

population counts at the municipal level by age, sex and citizenship

The **RSBI** is a population register updated on an annual basis with reference to the 31 st December of each year through the integration of individual flows of demographic dynamics (deaths, changes of usual residence to and from another municipality or to and from abroad)

Place of residence (municipality) and address (address identifier CUI)

Integration of RBI with the **RSBL** Register of the Places (RSBL – Registro Statistico di Base dei Luoghi)

MIDEA (Micro DEMographic Accounting)

Note by Italian National Institute of Statistics

ECE CES Group of Experts on Population and Housing Censuses

Twenty-fourth Meeting - Geneva, 21–23 September 2022

Permanent PHC

population counts at the municipal level by age, sex and citizenship

The integration of **RSBI** with the **RSBL** and of the **thematic registers**, such as those of Employment and Education, as well as the use of administrative archives held by Public Bodies and Ministries (National Security System archives, Ministry of Economy and Finance archives, Real Estate Register, Pensioners' Register, etc) made it possible to produce a population count for the Census

Each presence of individuals in administrative sources (CGD?)

Labor and Education archive,
Tax Returns archive,
Earnings,
Retired and Non Pension
Benefits archive,
Permits to Stay archive

signs of life

classified by duration patterns, type and reliability of the specific source, association with other individual records (e.g. household relations)

Household criteria for evaluation of national over/under coverage of RBI

Type of household	Criteria for individuals in RBI eligible to National over-coverage or confirmed like usually residents of RBI	Outcome with respect to RBI		
One person household	One-person household with no signs of live over the past 5 years who don't live in border municipalities	National Over-coverage	Individuals not in RBI eligible to under-coverage	
Individuals in a multi-person household	All household members with no signs of life		Confirmed like usually residents	Signs of work/study for at least 12 months over the relevant period
	Individual < 14 years of age attending school in the same municipality of RBI	Owner of a dwelling (or holder of a rent contract)		
	If at least one member owns a dwelling (or holds a rent contract) in the same municipality of RBI	No-border municipality		
	Husband/wife (even if not physically dependent) of worker/student/pension perceiver			

Identification of population groups for the 2020 population census counts

Description of outcomes	Type of register or Archive	Total population counts	Population census counts	
Population correctly placed in RBI	RBI vs AIDA	58,713,660	Yes	Usual resident population at December 31 th 2021
Under-coverage at national level	Only in AIDA	324,932	Yes	
<i>Uncertain units</i>	<i>Only in RBI</i>	<i>197,621</i>	<i>Yes</i>	
Over-coverage at national level	Only in RBI	1,005,908	No	
<i>Uncertain units</i>	<i>Only in AIDA</i>	<i>288,211</i>	<i>No</i>	
Under/over-coverage at local level	AIDA vs RBI	20,423	No	
Population not entered in the count	AIDA with unusable signs	1,410,497	No	
Usual resident population	AIDA vs RBI	59,236,231	Yes	
Total population	AIDA	61,961,252		

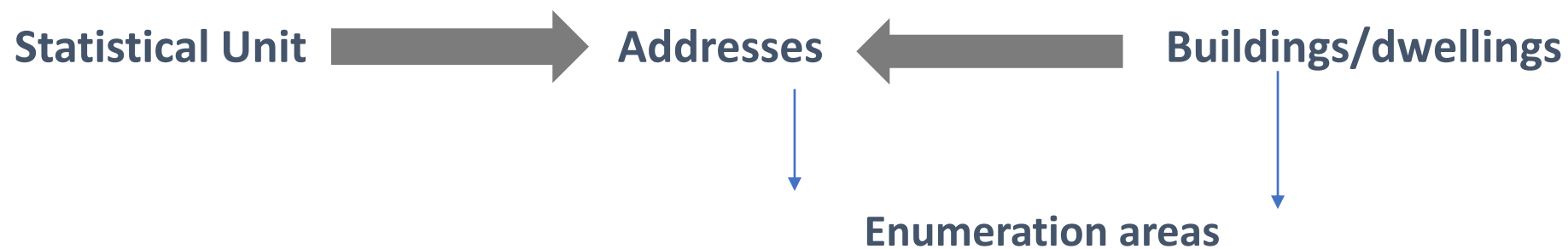
Some analyses are underway to verify *Uncertain units* (foreigners recorded in the RBI and only with permit to stay sign) to better assess whether or not to confirm them as usual residents

Benchmarking: measuring the quality of something by comparing it with something else of an accepted standard (a level of quality)

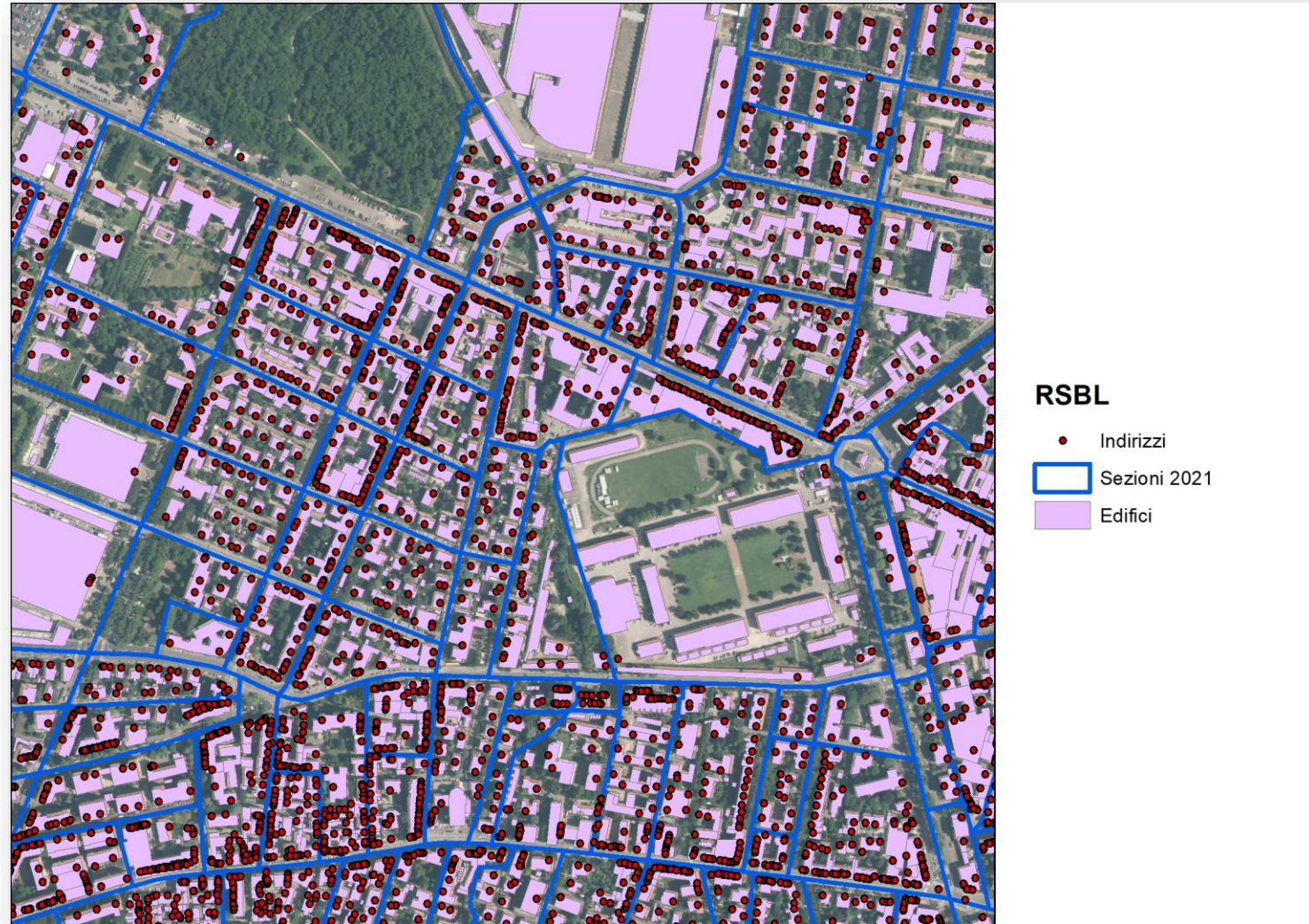
Register of Place - RSBL

RSBL

- Administrative territorial units and functional areas
- Enumeration areas and microzones
- Addresses and geographic coordinates
- Buildings and dwellings



Register of Place - RSBL



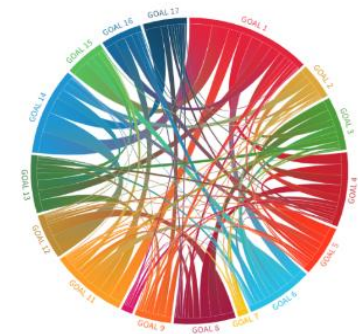
Register of Place - RSBL



- In **green** geographic coordinates of addresses
- In **red** microzones

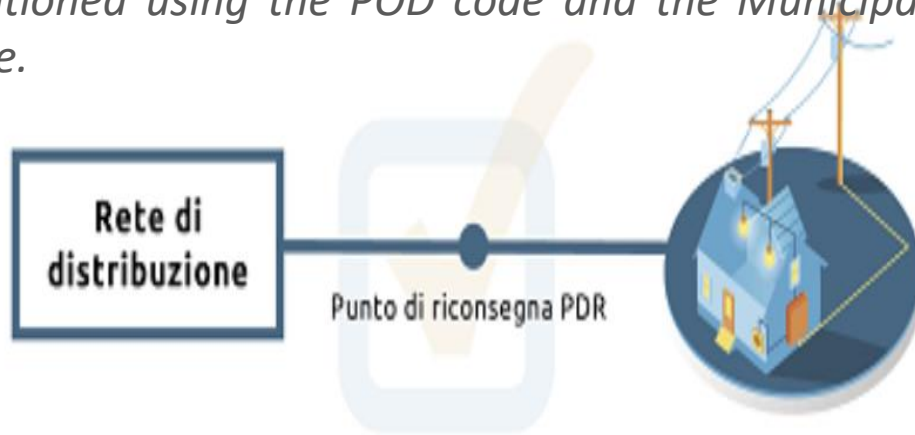
Administrative energy data file (CGD) and Energy Consumption Survey

The energy sector plays a crucial role today because of the needs of the energy transition driven by climate change and the current geopolitical and socioeconomic implications. Statistical information on energy sector is therefore of particular relevance for the objectives and monitoring the National Sustainable Development Strategy, the National Recovery and Resilience Plan and the Ecological Transition Plan.



Administrative energy data file (CGD) and Energy Consumption Survey

The Energy Administrative Register (**EAR_CGD**) provides data on energy statistics, with respect to the residential and industrial sector. The archive collects information on total consumption of electricity and on gas withdrawals on a yearly and monthly basis - service delivery agreements of customers, geo-positioned using the POD code and the Municipality code.



POD point of delivery



Place of residence (municipality) and address (address identifier CUI)

The 2021 Istat Household Energy Consumption Survey (**HECS**) provides a contribution to the national framework on energy statistics, with respect to the residential sector. The survey collects information on household energy equipment and their uses, drawing the overall picture of energy consumption and related expenditures in the reference year (2020).

Sample:

54 thousands households, nationally and regionally representative



Interview:

mixed **CAWI** (Computer-Assisted Web Interviewing) and **CATI** (Computer-Assisted Telephone Interviewing) technique

Administrative energy data file (CGD) and Energy Consumption Survey

HECS main themes investigated:

- characteristics of the dwellings
- systems in the dwelling for heating, production of domestic hot water and air conditioning
- their source of supply and technical characteristics
- frequency of use
- lighting systems and household appliances
- **expenses incurred in the previous year for energy consumption**
(**electricity, methane**, LPG, diesel, biomass) (*results in next dissemination Nov. 2022*)



*Note by Italian National Statistical Institute
International Conference on Sustainability Analysis, Sapienza
University of Rome
14-15 July 2022*

Administrative energy data file (CGD) and Energy Consumption Survey

- Expenses incurred in the previous year for energy consumption (electricity, methane, LPG, diesel, biomass)

HECS (partially) becoming 'Trusted Smart' Survey !
Energy questions replaced progressively by the CGD
Meter data instead of data from CATI - CAWI questionnaires



Benchmarking is a standard or point of reference by which data can be compared

- The process of aligning the sample estimate (**HECS**) with the universe estimate (**EAR_CGD**) is known as **benchmarking**.
- The national Energy Consumption Survey would benchmark the energy consumption level to the level of the Energy Administrative Register (EAR_CGD). This process of aligning the sample estimate (ECS) with the universe estimate (EAR_CDG) is known as benchmarking.

Lessons learnt

New TSE model

Signs of life benchmarking

- over coverage – under coverage: a priori remedies, they are proposed by design

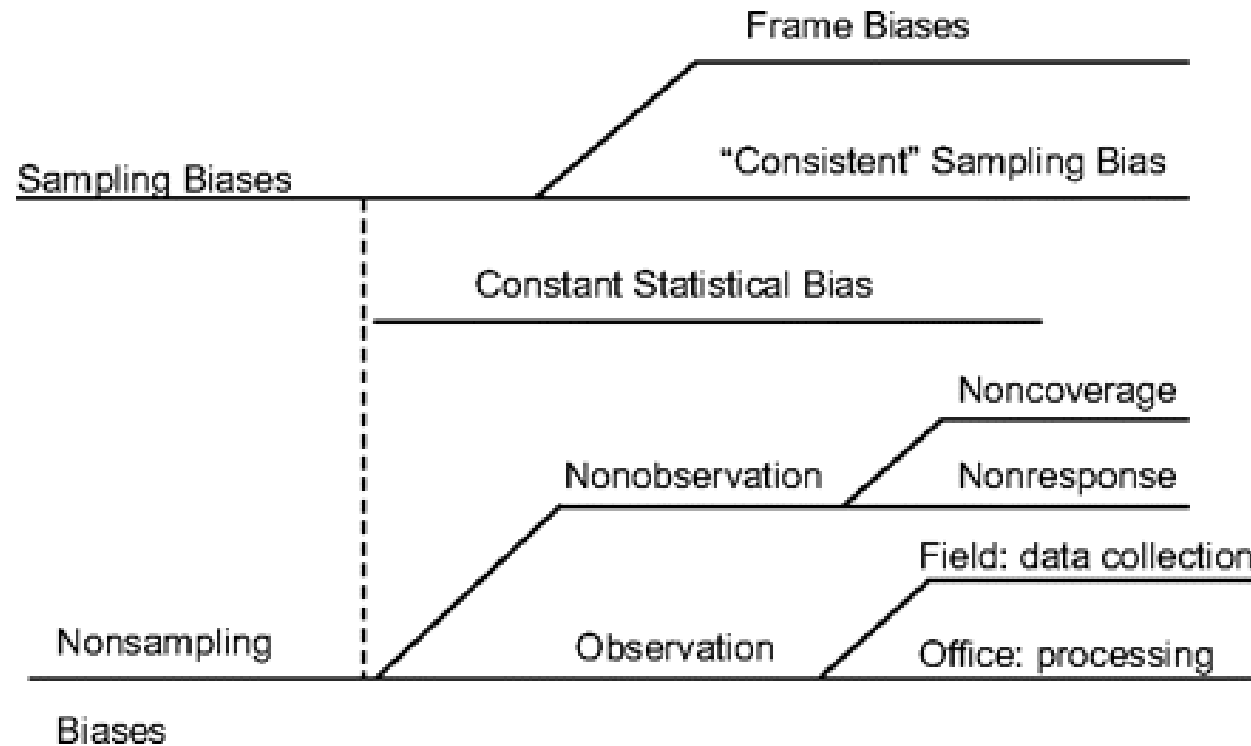
Reinventing the wheel?

- aligning sampling estimate with the universe estimate until we replace it with “true values”
- are meter data “true values”?

New biases in surveys



Figure 2. Schematic Presentation in Kish (1965) of Biases in Surveys.



Public Opin Q, Volume 74, Issue 5, 2010, Pages 849–879, <https://doi.org/10.1093/poq/nfq065>

The content of this slide may be subject to copyright: please see the slide notes for details.



Lessons learnt

Need for paradata

Metadata: Formalized data about statistical data needed to search for, display and analyze those data.

Paradata: Formalized data on methodologies, processes and quality associated with the production and assembly of statistical data

Thank you

Monica Pratesi monica.pratesi@istat.it

