



# EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL



EUROPEAN CONFERENCE ON  
QUALITY IN OFFICIAL STATISTICS  
2024 ESTORIL - PORTUGAL

 INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

eurostat 

The conference is partly  
financed by the European Union

# The framework for quality assessment in the Istat Integrated System of Registers: An application to the estimation of the Attained Level of Education

**Romina Filippini, Sara Giavante, Gaia Rocchetti**

# Outline

---

- Introduction
  - The quality framework in the Integrated System of Statistical Registers
  - The context of the Attained Level of Education (ALE)
- Application of the quality framework to ALE estimation process
  - Process description through the metadata model
  - Quality indicators to monitor the process
- Final remarks

# Introduction

Reorganization of the statistical production processes:

→ **Integrated System of Statistical Registers** (ISSR) based on the extensive use of administrative data combined with other types of data sources

Maximize the potential of administrative data:

- reduce costs
- respond more flexibly to information needs
- provide more accurate, timely, and efficient data



**Framework to evaluate the quality** of a system that integrates data from a plurality of sources.

# The quality framework in the ISSR

## Documentation models:

Systematize structural and referential **metadata** guiding the mapping of the process.

Provide a set of quality **indicators** for monitoring and evaluating each sub-process.

Sub-process documentation models are:

(1) Check data availability, (2) Acquire data, (3) Conduct preliminary evaluation, (4) Integrate data, (5) Classify and code, (6) Edit and Impute, (7) Derive new variables and units, (8) Calculate aggregates and (9) Validate outputs.

The template groups the informational objects into three macro-elements:

**Input, Subprocess and Output.**

# The context of the Attained Level of Education

ALE distribution on the Italian resident population is a standard output of the yearly Permanent Italian Census.

- The **Base Register of Individuals (BRI)** forms the foundation of the Permanent Census.

Data on education available from different sources:

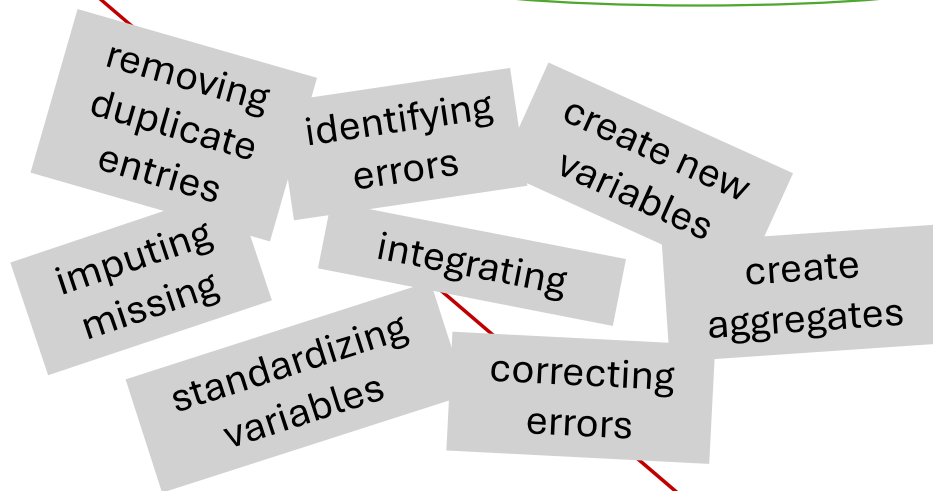
- **Administrative data**: processed and integrated by Istat, leading to the creation of a database on Education and Qualification (**BIT**)
  - Problems with coverage and timelines
- **2011 Population Census (CENS11)**
  - **Sample surveys data (MS)**
  - Auxiliary information from **APR4**



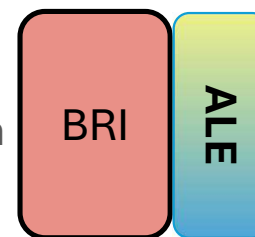
**Application of the quality Framework to ensure the reliability and validity of results**

# ALE estimation process

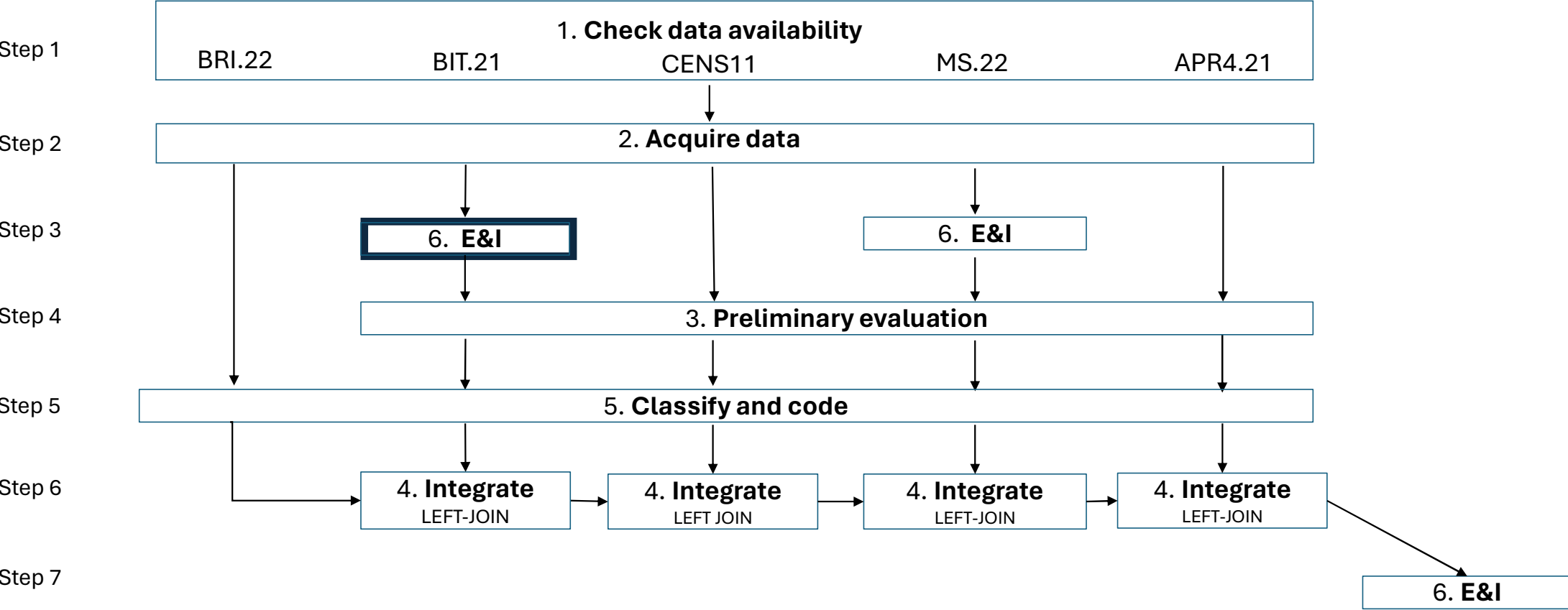
A) Gather data from different sources



B) Application of the model for ALE estimation



# Process description through the metadata model





# “Edit and impute” documentation model

input	Transformable	<b>BIT.21_RawData</b>
	Parameters	Variable to be checked: ALE
	Process support	Validity and consistency edit rules for ALE and school attendance relationship
subprocess	Process function	Impute missing and check records against edits
	Process step	Edit and impute (5.4 in GSBPM)
	Process method	Deductive imputation
	Rule	If ALE=NA and school course =«Primary» then ALE=2 «No ed.» If ALE is inconsistent with school course then «flag»=true
	Software	Oracle
output	Trasformed	<b>BIT.21 (corrected)</b>
	Qual. indicators	...

# Quality indicators to monitor the process

For each sub-process the standard quality framework offers a set of **recommended indicators** to assess quality and monitor the process.

- Compute quality indicators **during the process** (prior to proceeding with the subsequent step)
  - ↳ allows the identification and **rectification** of potential quality issues before they can propagate further in the process.
- Consider the **unique characteristics** of each case when choosing which indicators to use
  - ↳ some indicators presented in the standard version may not directly apply to the specific case: they may either be **declined or reinterpreted** to ensure a more focused and targeted quality assessment process.

# “Edit and impute” quality indicators

## Indicators proposed in the framework

6.1. Records with at least one missing

6.2. Records failing at least one edit

6.3. Variables failing at least one edit

6.4. Non-response rate per variable

6.5. Imp. rate per var.

6.6. Modification rate per variable

.....

# “Edit and impute” quality indicators

Indicators proposed in the framework	Indicators for ALE estimation: Step 7 - Integrated dataset
6.1. Records with at least one missing	<b>6.1.Int.</b> Records with missing ALE (Not in MS)
6.2. Records failing at least one edit	<b>6.2.Int.a.</b> Records with conflicting ALE and age <b>6.2.Int.b.</b> Records with ALE from MS < ALE from BIT
6.3. Variables failing at least one edit	Not relevant
6.4. Non-response rate per variable	Not relevant
6.5. Imp. rate per var.	Not relevant
6.6. Modification rate per variable .....	<b>6.6.Int.</b> Records with estimated ALE different from ALE in BIT (referred to t-1)

# “Edit and impute” quality indicators

Indicators proposed in the framework	Indicators for ALE estimation: Step 7 - Integrated dataset		value	
6.1. Records with at least one missing	<b>6.1.Int.</b> Records with missing ALE (Not in MS)		53,2mln/ 55,6mln	<b>=95.7%</b>
6.2. Records failing at least one edit	<b>6.2.Int.a.</b> Records with conflicting ALE and age		223/ 55,6mln	<b>=0.0%</b>
	<b>6.2.Int.b.</b> Records with ALE from MS < ALE from BIT		13.350/ 626,549	<b>=2.1%</b>
6.3. Variables failing at least one edit	Not relevant		-	
6.4. Non-response rate per variable	Not relevant		-	
6.5. Imp. rate per var.	Not relevant		-	
6.6. Modification rate per variable .....	<b>6.6.Int.</b> Records with estimated ALE different from ALE in BIT (referred to t-1)		3.1mln/ 14.7mln	<b>=20.2%</b>

# “Edit and impute” quality indicators

Indicators proposed in the framework	Indicators for ALE estimation: Step 7 - Integrated dataset		value	
6.1. Records with at least one missing	<b>6.1.Int.</b> Records with missing ALE (Not in MS)	53,2mln/ 55,6mln	=95.7%	
6.2. Records failing at least one edit	<b>6.2.Int.a.</b> Records with conflicting ALE and age <b>6.2.Int.b.</b> Records with ALE from MS < ALE from BIT	223/ 55,6mln 13.350/	=0.0%	
6.3. Variables failing at least one edit	<p style="text-align: center;"><u>At the end of the ALE estimation process</u></p> <p style="text-align: center;"><b>36 quality indicators</b> are computed, pertaining to <b>4 different documentation models</b>, carried out through <b>7 process steps</b>.</p> <p style="text-align: center;">To facilitate comparisons, all the indicators are computed using a <b>uniform structure</b>.</p>			
6.4. Non-response rate per variable				
6.5. Imp. rate per var.				
6.6. Modification rate per variable .....		<b>6.6.Int.</b> Re		

# Final remarks

- The standard template for documenting the overall process enables **efficient organization of information**, facilitates collaboration among team members, thus enhancing **transparency** and facilitating meaningful analysis and interpretation of the results.
- Integrating quality assessment into **every phase of the process** encourages a proactive approach, enhancing process efficiency and higher quality outcomes.
- The **flexibility of the framework** is essential to select indicators that best align with the goals of the project.

This approach leads to **more efficient and effective processes**, contributing to the **continuous improvement** of the system.



EUROPEAN CONFERENCE ON  
QUALITY IN OFFICIAL STATISTICS  
2024 ESTORIL - PORTUGAL

# Thank you

Romina Filippini [filippini@istat.it](mailto:filippini@istat.it)

Sara Giavante [giavante@istat.it](mailto:giavante@istat.it)

Gaia Rocchetti [grocchetti@istat.it](mailto:grocchetti@istat.it)



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

eurostat 

The conference is partly  
financed by the European Union





# EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL