
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly
financed by the European Union
CASINO ESTORIL
HOTEL PALÁCIO
EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Improving Statistical Registers' Quality through Attribute-Driven Spatial Matching

Damiano Abbatini abbatini@istat.it

Luisa Franconi franconi@istat.it

Daniela Ichim ichim@istat.it

Affiliation: Istat – Italian National Institute of Statistics

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Introduction

- **Context**: Modernization of official statistics production at Istat
- **Model**: Integrated System of Statistical Registers (ISSR)
- **Goal**: Geo-referencing core statistical units

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Registers and Their Importance

- **Population Register**: Demographic data (e.g., age, gender, citizenship)
- **Register of Places**: Multidimensional integration of geographic objects (addresses, buildings, dwellings, census blocks, etc.)
- **Goal**: Accurate geo-referencing of statistical units

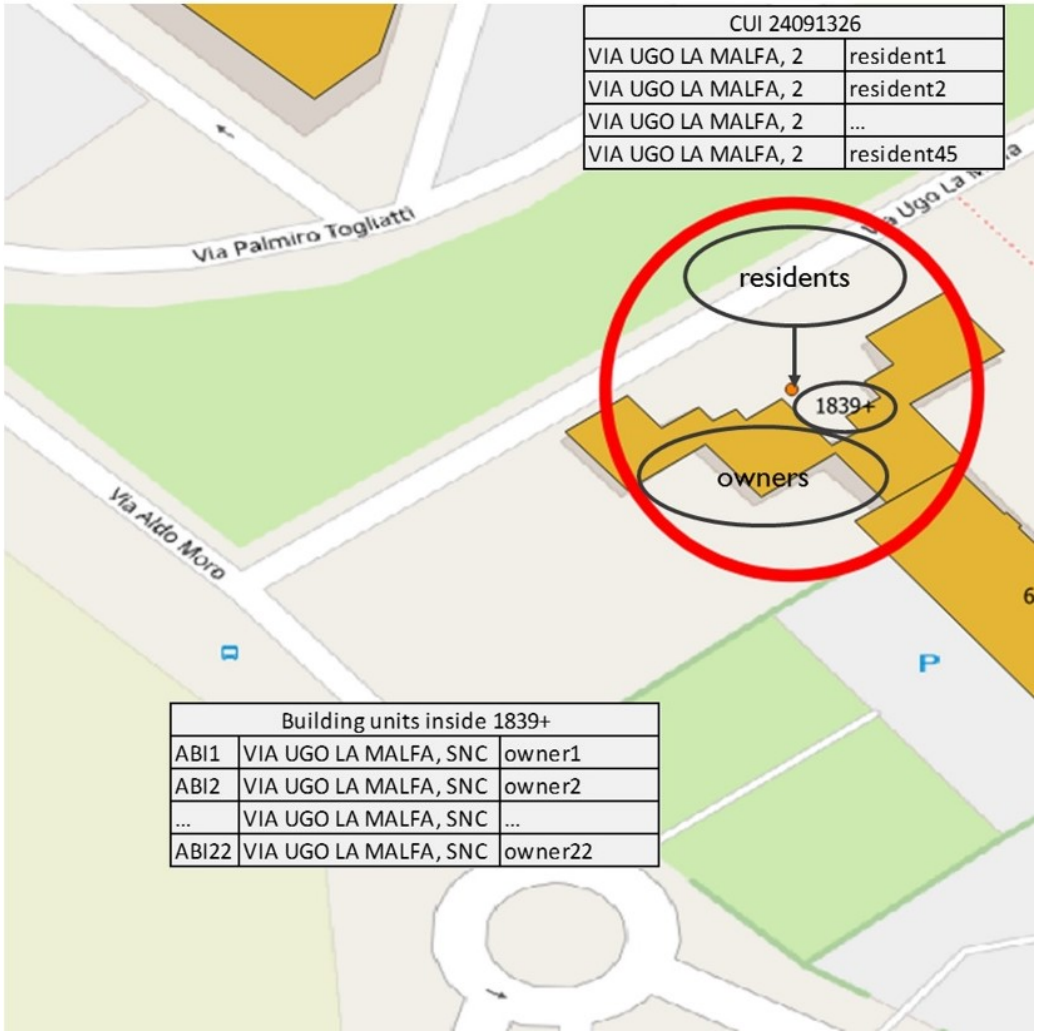
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Challenges

- **Issue:** Incomplete information leading to mismatches (e.g., missing street numbers)
- **Problem:** Difficulty in linking addresses and buildings accurately
- **Example:** Different data sources having slightly different information

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Mismatched units and logic of proposed solution



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Methodology

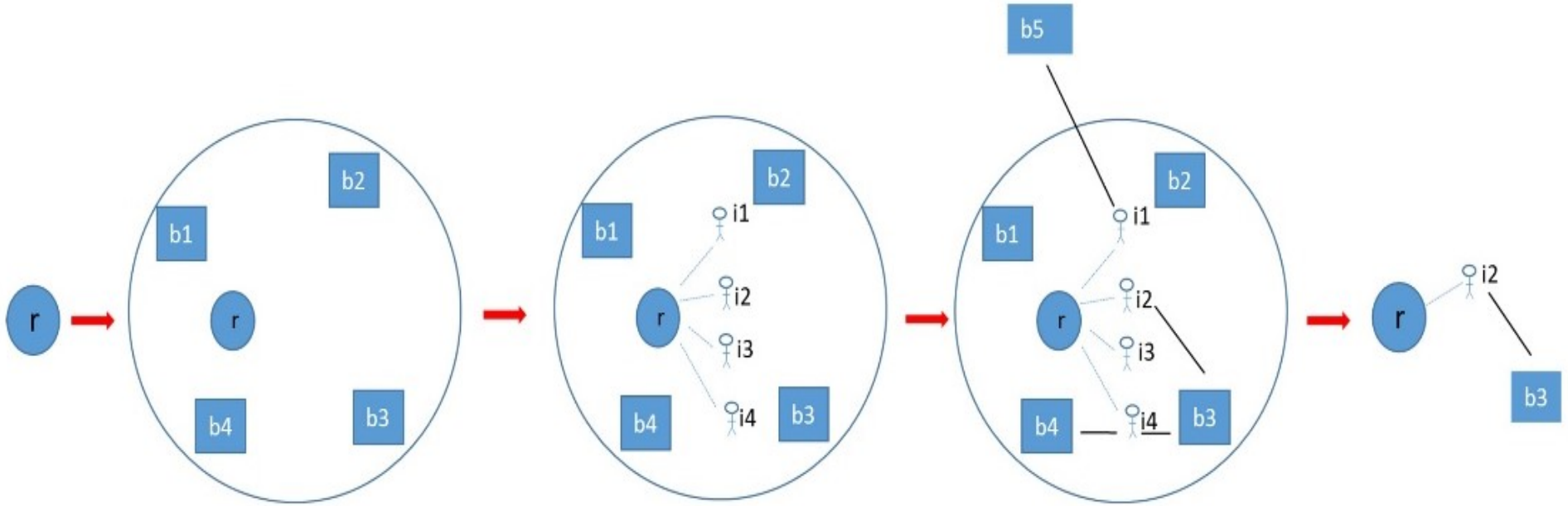
- **Approach:** GIS-based address matching

- **Assumption:** Owners usually live in their own houses (70%)

- **Process:**
  - Measure proximity (Euclidean distance)
  - Identify nearby buildings using a buffer area
  - Match buildings with addresses based on ownership

The conference is partly financed by the European Union

GIS-based address matching (*r-b*) derived from a sequential selection process starting from geometric buffers. Large circle indicates the buffer, blue circles represent residential addresses (*r*), squares represent buildings (*b*), symbols labelled *i* represent individuals, dashed lines indicate a **residential** relationship, full lines indicate **ownership** relationships



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Study Area and Data

- **Data Sources:**
  - Register of residential buildings: 743 thousand buildings
  - Register of addresses: 2.16 million addresses
  - Population register: 4.38 million people
- **Focus:** Residential addresses and buildings
- **Region:** Emilia-Romagna, Italy (NUTS2 region)

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Study Area

NUTS3 regions of Italy (red) and of the Emilia-Romagna NUTS2 region (yellow)

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Results

- **Initial Situation:** 25.11% of population unallocated to buildings
- **Matching Success:** Varied across provinces, highest in Bologna (51.25%), lowest in Parma (23.13%)
- **General Outcome:** Improved linkage and data reconciliation

The conference is partly financed by the European Union

Results

Resident Population subject to spatial matching and success rate by Province

NUTS3	Population on Addresses not associated with Buildings* (%)	Success rate** (%)
Piacenza	38.00	28.22
Parma	24.02	23.13
Reggio nell'Emilia	37.64	43.19
Modena	26.51	51.15
Bologna	18.11	51.25
Ferrara	21.80	44.53
Ravenna	18.48	26.23
Forlì-Cesena	21.75	32.60
Rimini	29.12	45.75
Total	**25.11**	**40.87**

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Conclusions and Future Work

- **Summary**: Encouraging results from spatially-based matching

- **Future Directions**:
  - Explore different buffer sizes
  - Analyze territorial differences
  - Optimize geographical position of buildings

- **Potential**: Enhance and refine methodology for better data integration

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Thank you for your attention!

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

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union