Geneva (CH) 22-24 May 2024

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS Expert Meeting on Statistical Data Collection and Sources

Conflation of Maps for the Integration of Geospatial Data and Enhancement of Building Registry Quality

Gianluigi Salvucci – Phd Economic Geography
Istat | DCAT

Joint work with D. Abbatini, D.Ichim, S. Lucchetti

Outline

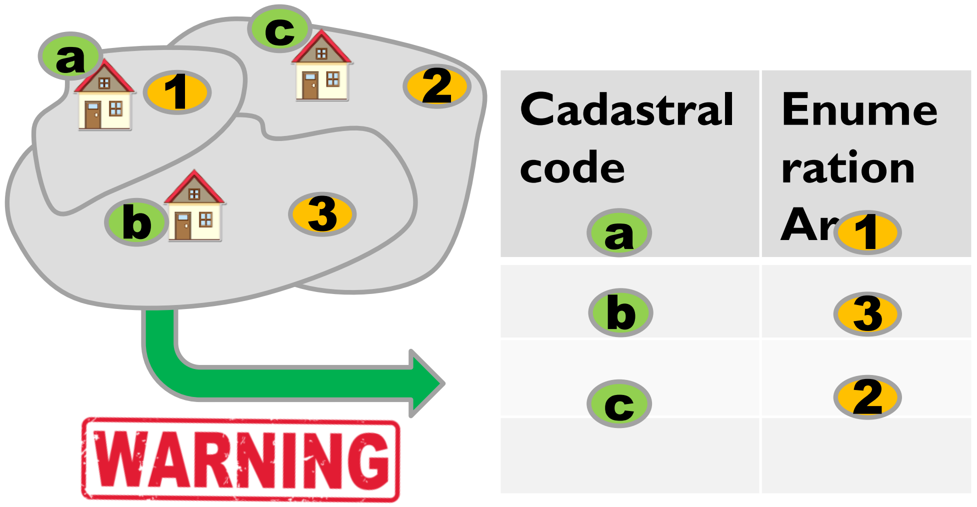
- The Building Census: From Field Survey to Building Registry
- Matching Statistical Units with Census sections
- Geo-Referencing in Cadastre: An Unresolved Issue
- Conflation: Comparing Cadastre Positioning and Integrating Data in Registry
- ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage
- From ASI to ASIR: towards a Relocating Algorithm
- Empirical Findings: Case Study, The Municipality of Cassino (FR)
- Conflation: Comparing Cadastre Positioning and Integrating Data in Registry5
- Conclusions: A New Starting Point

The Building Census: From Field Survey to Building Registry

- Traditional Building Censuses until 2011 required field visits by enumerators.
- Tools such as section maps and address lists were utilized to determine building locations.
- This enabled matching buildings and their characteristics to their corresponding census sections.

Matching Statistical Units (buildings) with Census Sections

- After 2011, the Census of Buildings was modified to rely on administrative registers, particularly the Cadastre.
- The Cadastre is a database of polygons.
- In Italy, the Cadaster does not contain any indication regarding census sections.
- Furthermore, the Cadaster is not georeferenced in a geographic system
- The Cadaster has been repositioned by Istat
- Through a "spatial join," each building was assigned the census section within which the building lies.



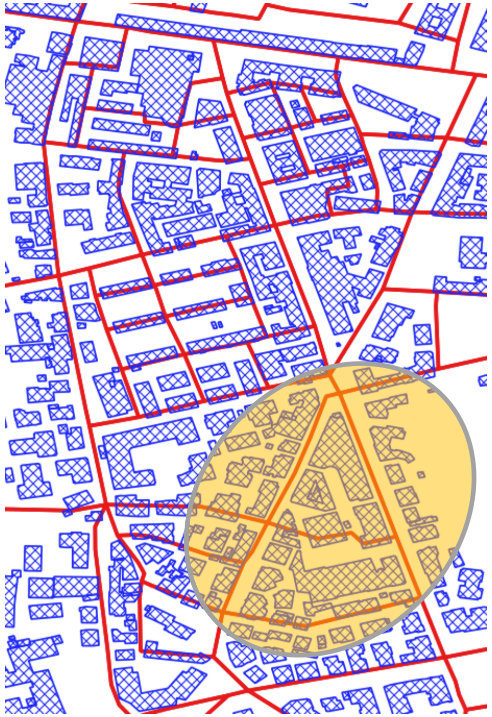
REGISTRO EDIFICI E UNITÀ IMMOBILIARI

Geo-Referencing in Cadastre: An Unresolved Issue

- Georeferencing the Cadastre is a highly time and labor-intensive process.
- The result can lead to wrong, unclear or ambiguous situations.
- The registry system assigns the census section based on various components. In addition to the Cadastre, the assigned census sections are also validated through the building address.
- In Italy, each municipality declares in which census section the addresses are located, allowing us to cross-check."

Cadastrer by Istat on Census sections

DBSN on Cesnsus sections

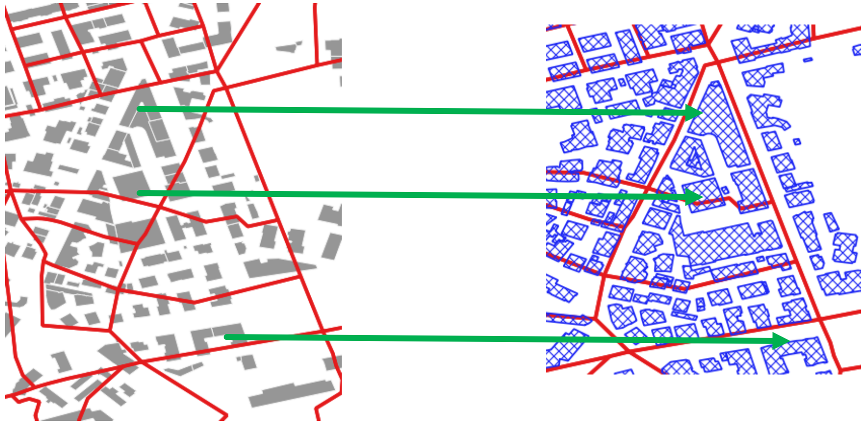


Istat

Conflation: Comparing Cadastre Positioning and Integrating Data in Registry

- To solve inconsistencies, a vertical conflation approach was adopted.
- A reference database was selected, in this case, the polygons database from the Italian Military Geographic Institute (IGM) (a layer of buildings in DBSN geodataset format).
- Since there is no identifier enabling a deterministic record linkage between the georeferenced cadastre and IGM databases, polygon shape recognition is used to integrate the data.

Cadastrer polygons by Istat | DBSN polygons by IGM



Cadaster code	Igm Code	Enumeration Area code

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

- ASI (Aggregate Shape Similarity Index) proposed by Ďuračiová.
- It is a fuzzy indicator that allows classifying the type of polygon correspondence between the two databases.
- This enables us to understand if the buildings have been positioned correctly and therefore proceed with data integration, particularly the assignment of the census section.

isprs International Journal of *Geo-Information*

MDPI

Article

An Aggregated Shape Similarity Index: A Case Study of Comparing the Footprints of OpenStreetMap and INSPIRE Buildings

Renata Ďuračiová

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

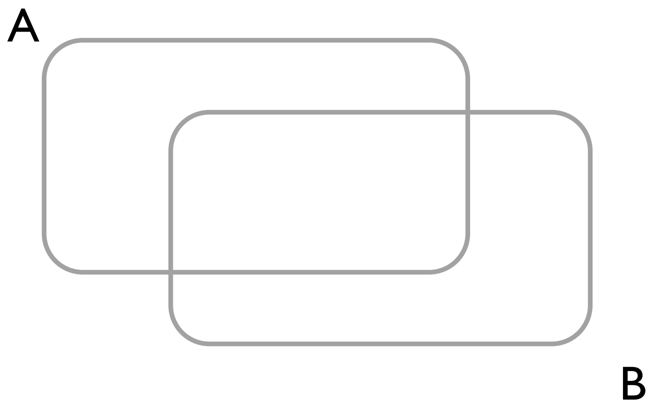
- **Metrics comparing the overlapping surface**

Jaccard index in Tanemoto version

$$sim_T = \frac{A \cap B}{A + B - (A \cap B)}$$

Sorensen

$$sim_{SD} = \frac{2(A \cap B)}{A + B}$$

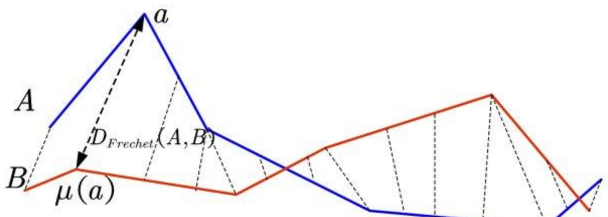


ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

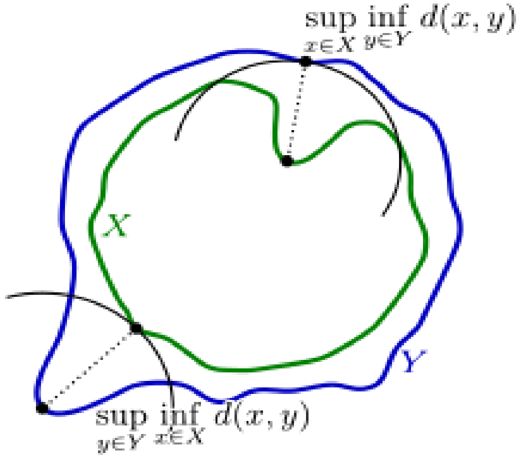
Metrics comparing boundaries:

- These metrics approximate how much similar or different two contours are, thus aiding in polygon recognition.
  - Hausdorff Distance: Measures the maximum distance between the contours of two polygons.
  - Fréchet Distance: Assesses the similarity between the contours of two polygons, considering the order of points along the path.

Frechet distance



Haussdorf distance



Istat

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

Metrics comparing only shapes

- Finally, we have metrics to compare shapes regardless of their position.
- These metrics consider similarity between shapes based on intrinsic characteristics such as area, number of vertices, and perimeter length.
- This means that two polygons with similar shapes will have similar values for these characteristics, regardless of their position on the map.

$$simA = 1 - \frac{|A - B|}{\max(A, B)}$$

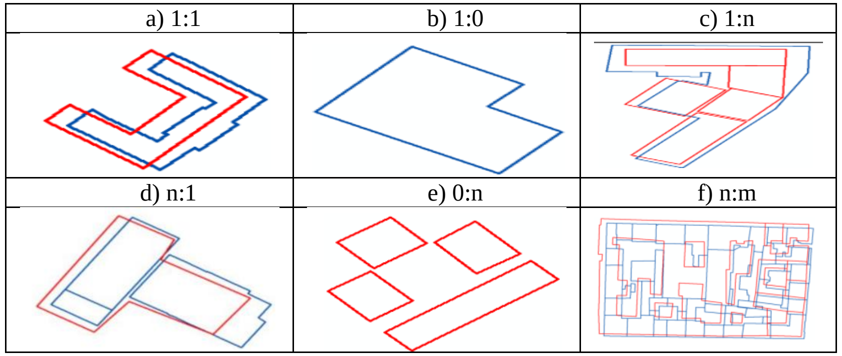
$$simP = 1 - \frac{|A - B|}{\max(A, B)}$$

$$simV = 1 - \frac{|A - B|}{\max(A, B)}$$

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

Cleaning and prepare data

- The main issue in the two databases is the lack of a one-to-one correspondence between elements.
- Situations regarding buildings feature disjointed polygons, leading to discrepancies even if they identify the same buildings or groups of adjacent buildings.
- The approach proposed by Ďuračiová is to merge buildings within a certain distance and then compare the aggregates.



Fan et al. (2014)

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

- ASI a fuzzy indicator

Sim_D = Max between:

sim_H—Hausdorff distance similarity index,
sim_F—Fréchet distance similarity index,

Sim_S = Min between :

sim_T—Tanimoto (Jackard) similarity index,
sim_SD—Sørensen–Dice similarity index,

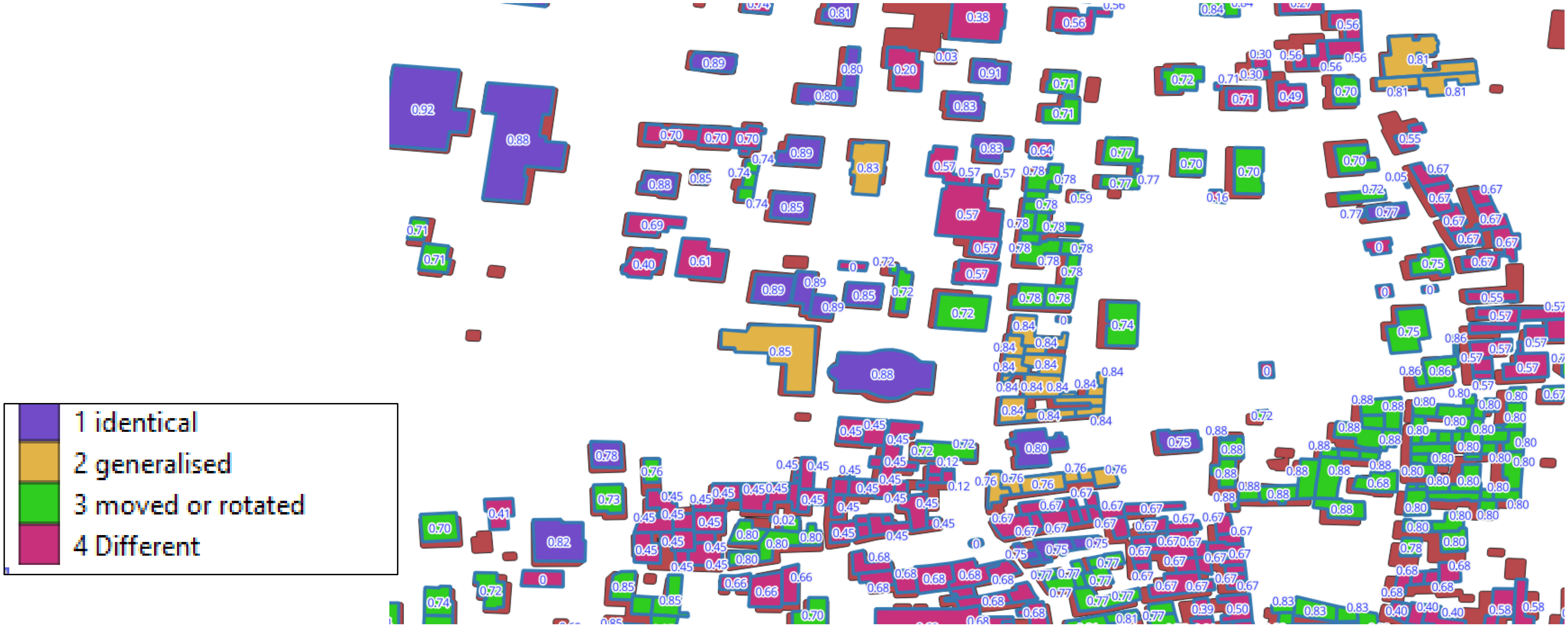
Sim_Sh= Min between :

sim_A—area similarity index,
sim_P—perimeter similarity index,
sim_V—vertices similarity index.

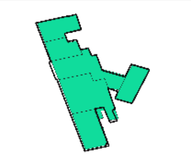
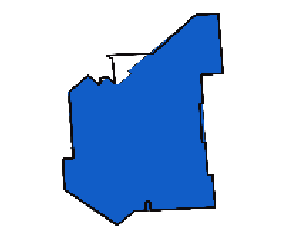
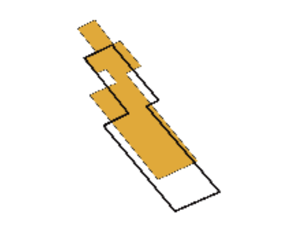
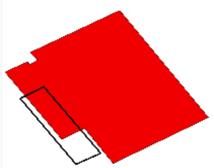
Max between

Min between

ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

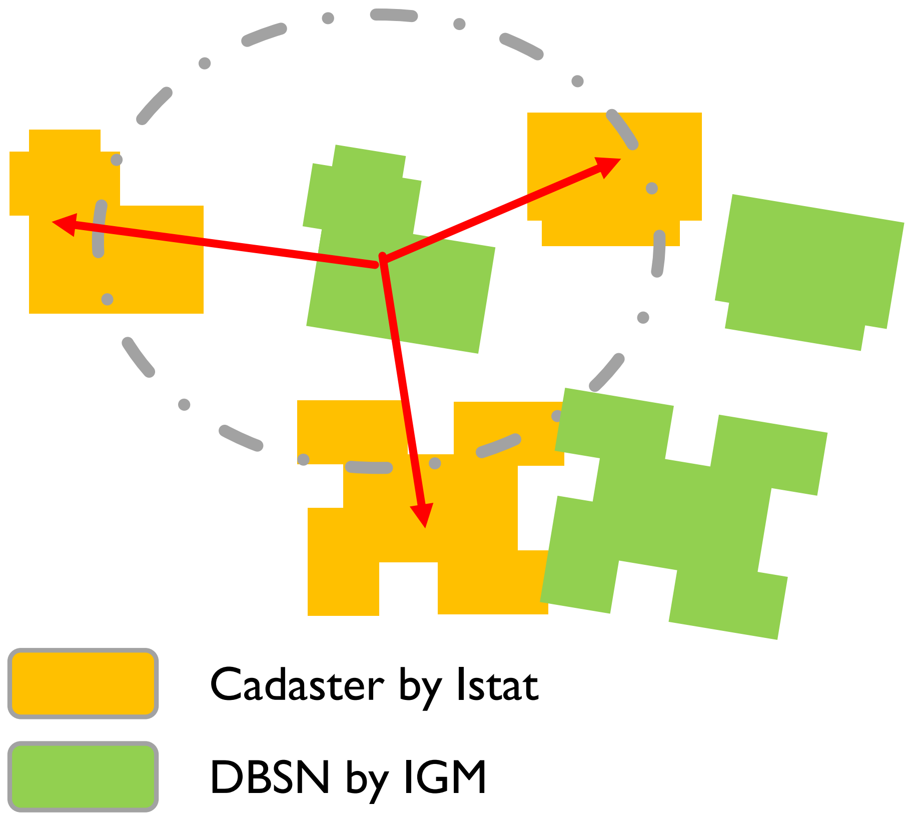


ASI (Aggregate Shape Similarity Index): A Tool for Geodatabase Record Linkage

Type	Description	parameters	Example
1	**Identical**	ASI > 0.85	
2	**Generalised**	sim_d > 0.75 sim_s > 0.75 sim_sh > 0.75	
3	**Moved**	sim_v > 0.75 sim_s < 0.75 sim_sh > 0.75	
4	**Different**		

From ASI to ASIR: towards a Relocating Algorithm

- ASI measures the correspondence between two building geodatabases, assuming they are accurately positioned.
- The comparison is typically done by overlaying corresponding pairs.
- In our case, the accuracy of the Cadastre positions is uncertain, requiring us to expand the comparison to all buildings within a certain distance.
- ASIR (Aggregate Shape Similarity Index for Relocation) is an automated procedure which relocates Cadastre buildings to match IGM ones.

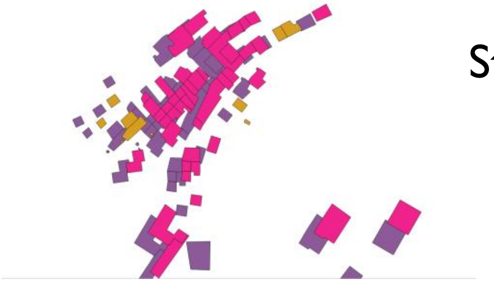


Liu, Lingjia, Xiaohui Ding, Xinyan Zhu, Liang Fan, e Jun Gong. 2020. «An Iterative Approach Based on Contextual Information for Matching Multi-scale Polygonal Object Datasets». *Transactions in GIS* 24 (4): 1047–72. https://doi.org/10.1111/tgis.12625.

From ASI to ASIR: towards a Relocating Algorithm

- Sequential procedure: on map sheet basis (meaning a subset of buildings designed respecting their relative distances).
- ASIR evaluates the proportion of unrecognized buildings, observing those that do not match between the two databases.
- The comparison is typically done by overlaying corresponding pairs. In our case, for each cadastral building, pairs are created with IGM buildings within 500 meters using Geopandas' geodataframe.
- Once the map sheet is repositioned, the ASI is recalculated to assess the improvement compared to the initial situation.
- ASIR integrates Duročiová's qualitative approach with Liu's matrix comparison and adds geographical displacements to enhance quality.

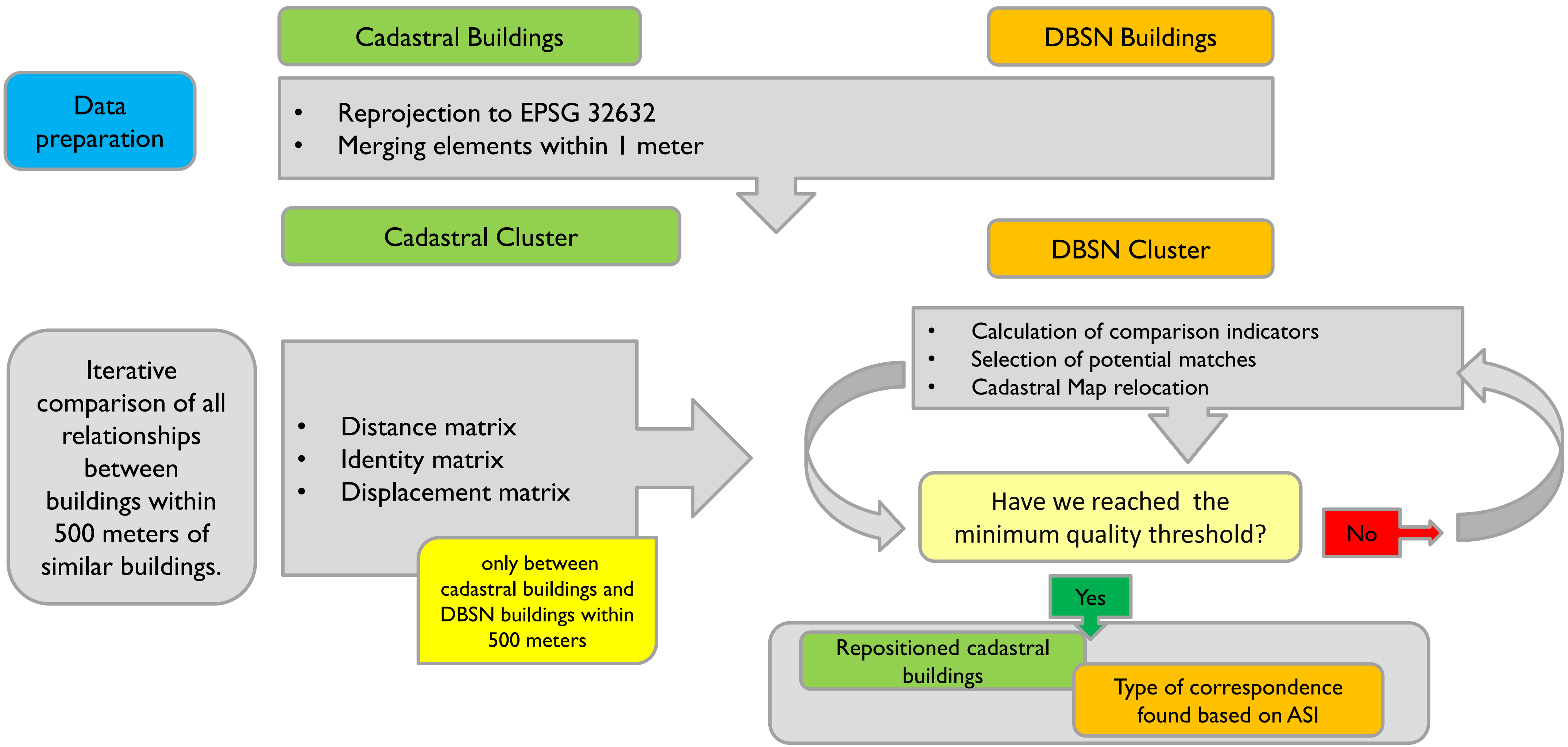
Step 1- DBSN data



Step 2 – ASI comparison

Step 3 – ASIR relocation

From ASI to ASIR: towards a Relocating Algorithm



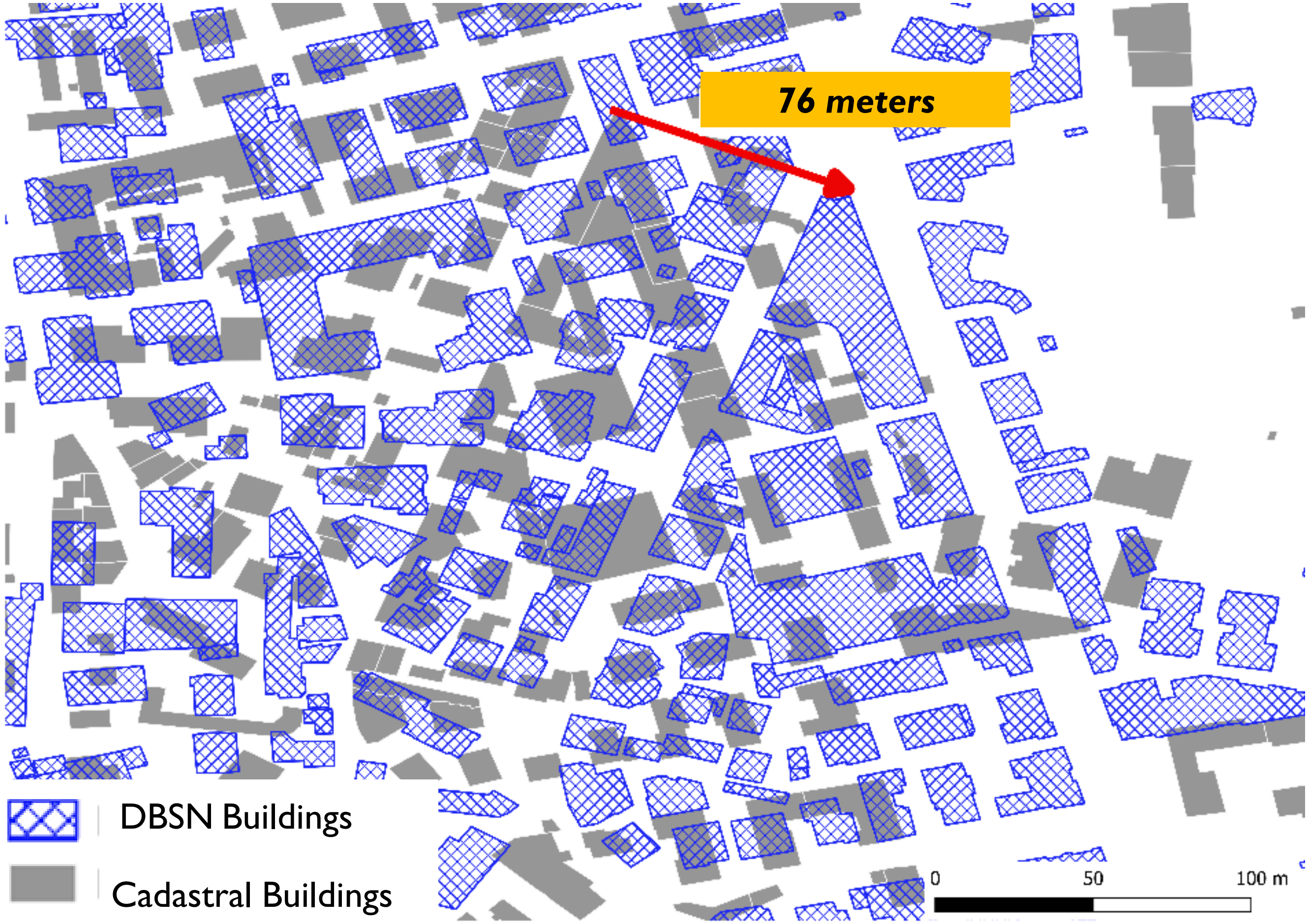
Empirical Findings: Case Study, The Municipality of Cassino (FR)

- The urban fabric of Cassino is complex and densely populated, with an intricated road network. The municipality is divided into 154 map sheets.
- Before the application of ASIR, no buildings had been identified, indicating a critical situation.
- ASIR identified identical buildings and those with different resolutions, significantly improving the situation.
- After the application of ASIR, 130 out of 154 map sheets were correctly positioned, without any human intervention.

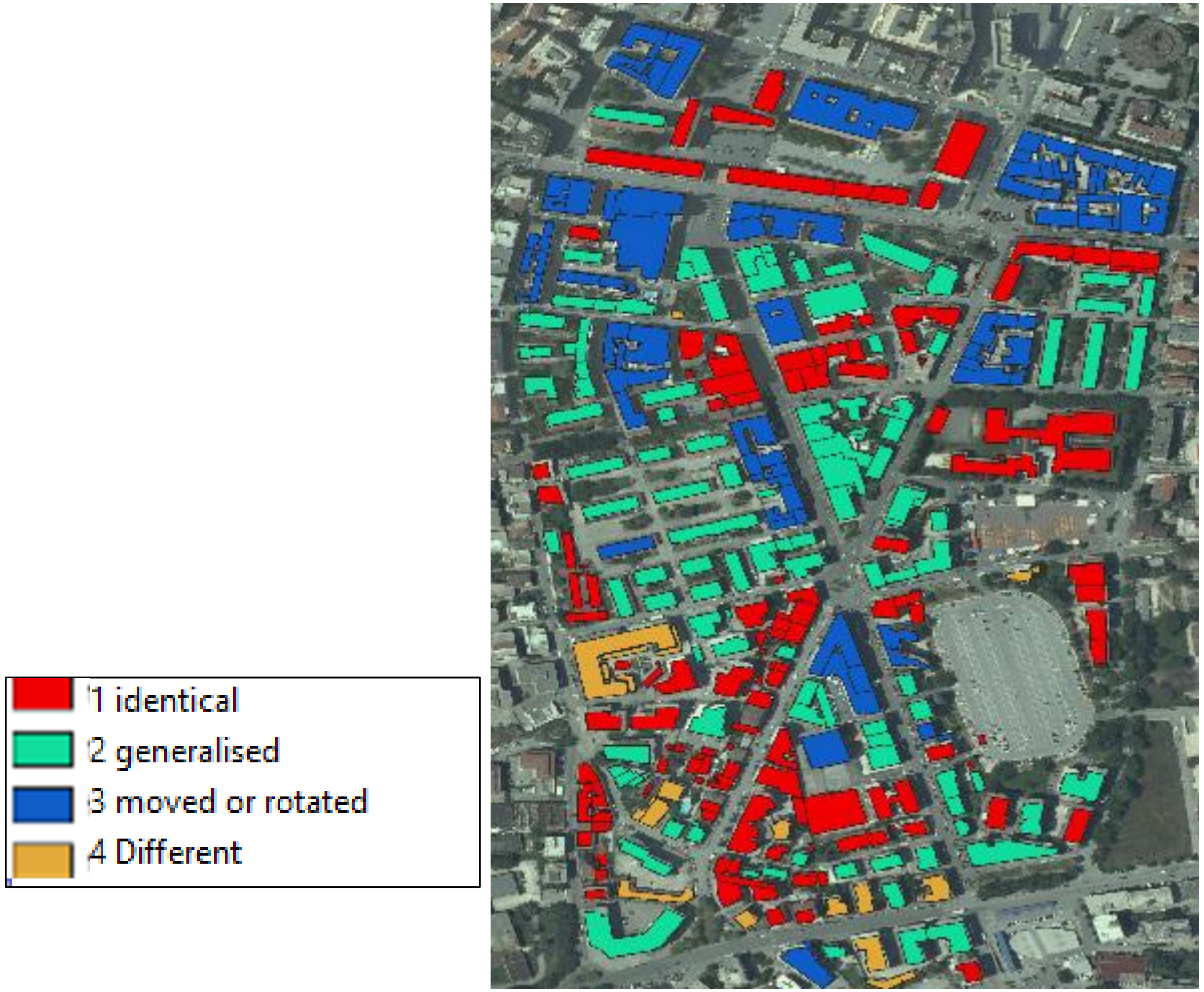
Classification of buildings in the municipality of Cassino before and after the application of ASIR.

	Before ASIR application		After ASIR application	
	abs	%	abs	%
Different	23.686	92,91	12.794	50,18
Moved	1.808	7,09	6.270	24,59
Generalised	0,00	0,00	812	3,19
Identical	0,00	0,00	5.618	22,04
total	25.494	100,00	25.494	100,00

Empirical Findings: Case Study, The Municipality of Cassino (FR)

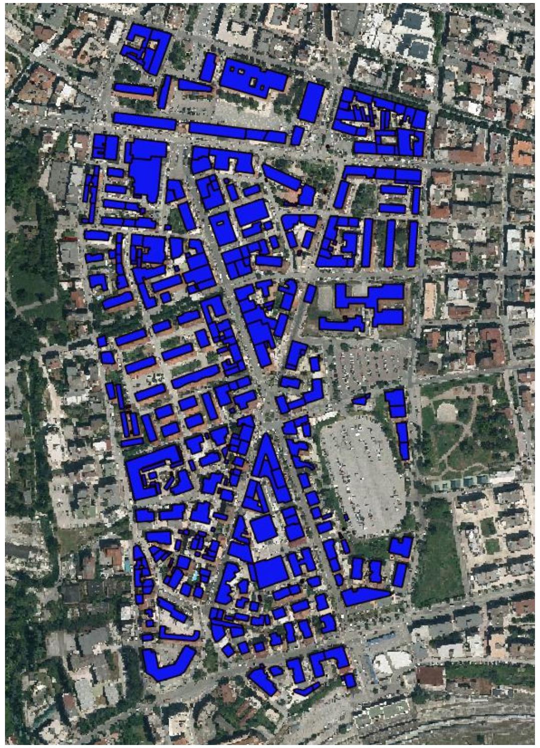


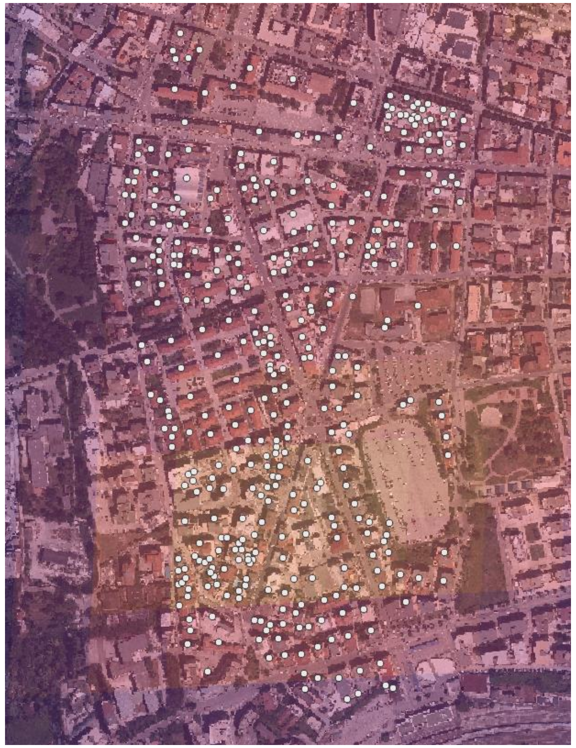
Empirical Findings: Case Study, The Municipality of Cassino (FR)

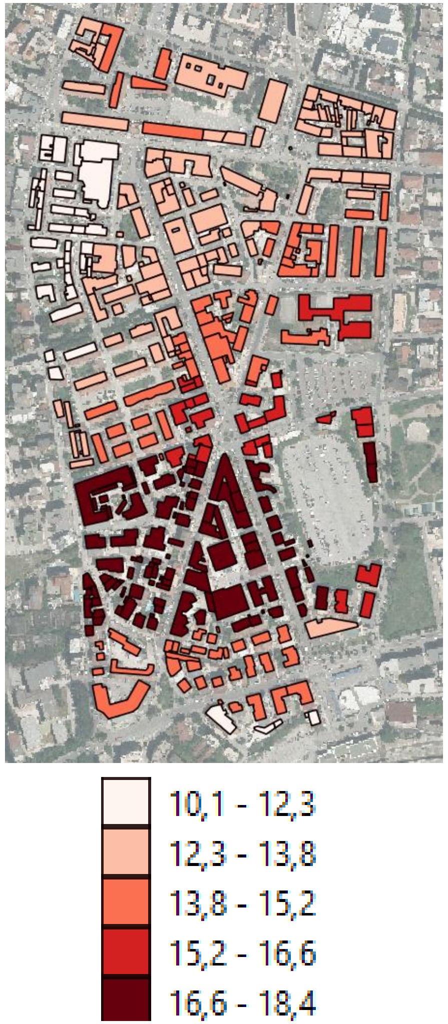


Conflation: Comparing Cadastre Positioning and Integrating Data in Registry

1. The Cadastre lacks certain crucial information, such as building height.
2. This information can only be retrieved if we are certain that the buildings are correctly positioned, utilizing Lidar survey data.
3. ASIR ensures the quality level of this operation by accurately repositioning the buildings, facilitating the retrieval of missing information.







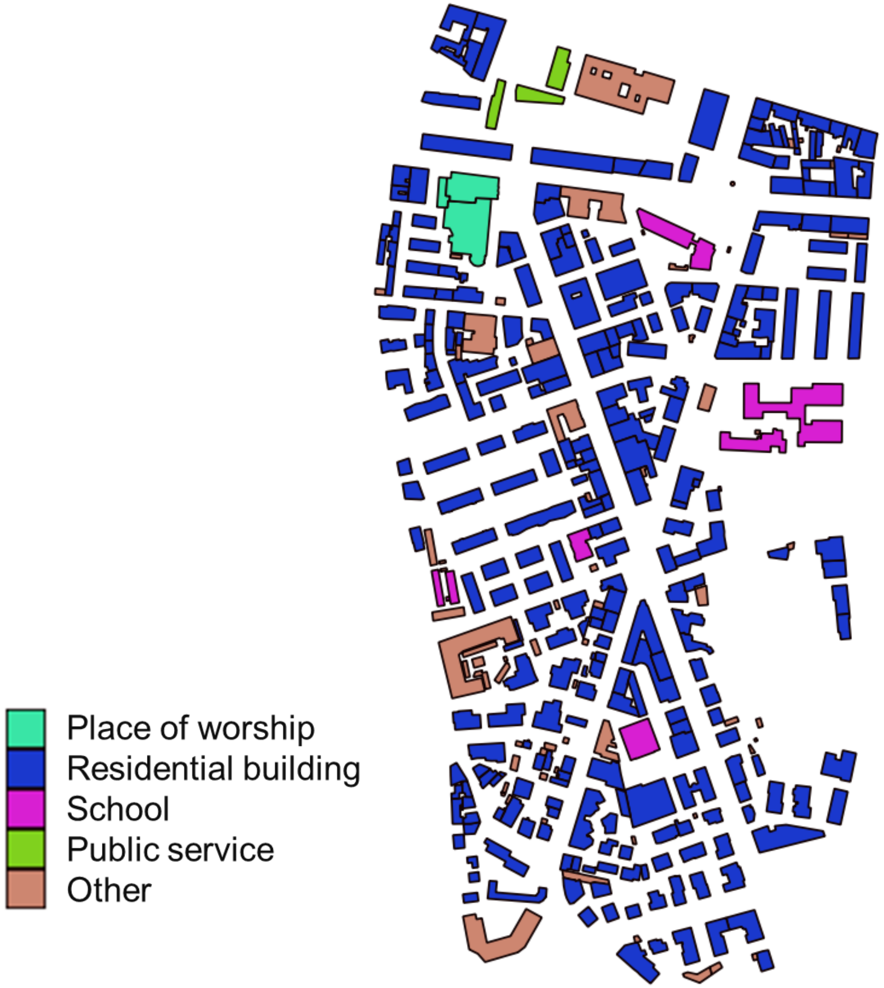
Pesaresi, Martino, e Panagiotis Politis. 2022. «GHS-BUILT-H R2022A - GHS Building Height, Derived from AW3D30, SRTM30, and Sentinel2 Composite (2018)

Conflation: Comparing Cadastre Positioning and Integrating Data in Registry

- Thanks to ASIR, recognizing buildings in the IGM database allows us to reassign to the Cadastre the characteristics detected by the IGM.

- This process serves to validate the information registered in the Cadastre and integrate any missing data.

- ASIR facilitates the integration and validation process, improving the overall completeness and accuracy of information in the Cadastre.



Conclusions: A New Starting Point

- Conflation represents a significant enhancement for our registry, enriching it with information and improving the positioning of elements.
- However, these benefits come with costs in terms of resources.
- ASIR greatly reduces processing time.
- New scenarios and opportunities in our geospatial data management process may be envisaged.

Thank you!

GIANLUIGI SALVUCCI | salvucci@istat.it

Istat | Istituto Nazionale di Statistica