



## *Research paper*

*First published online: November 25, 2025*

**Gianluigi Salvucci\***

# **L'INTEGRAZIONE DEI DATI GEOGRAFICI A SUPPORTO DELLA STATISTICA UFFICIALE**

## **Abstract**

This study introduces the Aggregate Spatial Similarity Index Relocation - ASIR, a methodology aimed at enhancing the accuracy and quality of geographic data. ASIR facilitates the repositioning of geographic objects, ensuring better alignment across diverse data sources. This approach streamlines the integration and conflation process, reducing manual effort and ensuring excellent results in terms of data consistency and completeness. Implementing ASIR within the context of building registries contributes to effective geographic data management, crucial for territorial planning and management activities.

**Keywords:** geographic data quality, data integration, conflation, building

---

\*Dipartimento per le Statistiche Economiche, Ambientali e Conti Nazionali (DIAE). Direzione Centrale per le Statistiche Ambientali e Territoriali (DCAT). Servizio Ambiente e Territorio (ATA). Istat, Roma, Italia.

## 1 Introduzione

La discussione sulla qualità dei dati geografici è una tematica ampiamente esaminata in diversi contesti di ricerca. Nel campo geografico, l'attenzione si concentra sull'accuratezza e la precisione dei dati, con particolare riferimento al posizionamento corretto degli edifici nel contesto catastale italiano. A livello internazionale, l'attenzione si sposta sulla valutazione della completezza e dell'accuratezza dei dati presenti in OpenStreetMap, dove diverse metodologie statistiche sono state impiegate per valutarne la qualità. L'alta qualità dei dati geografici è essenziale per garantire un'efficace integrazione tra diverse fonti. Solo attraverso il confronto di dati accuratamente posizionati è possibile evitare duplicazioni e costruire un archivio affidabile. Questa problematica riveste particolare importanza nel nuovo Sistema dei Registri Statistici di Base, cuore del nuovo Sistema dei Censimenti dell'Istituto Nazionale di Statistica (Istat), soprattutto per quanto riguarda il Registro degli Edifici. Il Registro degli Edifici si basa principalmente su dati di fonte amministrativa, principalmente derivanti dal catasto. Per valutarne la completezza e facilitare l'integrazione con edifici provenienti da altre fonti, è essenziale riconoscerli in modo univoco. Il lavoro di Ďuračiová (2023) ha introdotto una metodologia innovativa per il riconoscimento delle sagome edilizie, fondata sull'indicatore Aggregate Spatial Similarity Index Relocation (ASI). Questo indicatore si basa sull'aggregazione di misure elementari finalizzate a descrivere le caratteristiche geometriche delle forme e a distinguere se due edifici sono uguali oppure differenti solo per risoluzione cartografica. Nel presente studio, l'ASI costituisce il punto di partenza per lo sviluppo del modello ASIR (Aggregate Spatial Similarity Index Relocation). È importante sottolineare che, mentre ASI è un indicatore di similarità di forma, Aggregate Spatial Similarity Index Relocation (ASIR) è un modello logico che estende l'ASI, introducendo la componente di ricollocazione (Relocation). ASI e ASIR sono quindi concetti distinti: il primo misura la similarità tra oggetti, il secondo utilizza tale misura per spostare e riallineare automaticamente gli oggetti geografici tra fonti differenti. A partire da questo indicatore si è sviluppato il modello geografico ASIR, che si avvale di indicatori costruiti mediante funzioni di aggregazione basate sulla logica fuzzy, introdotta da (Zadeh 1965). In questo contesto, la logica fuzzy non costituisce un indicatore autonomo, ma un metodo di calcolo che consente di combinare in modo flessibile più misure elementari, rappresentando gradi intermedi di similarità tra oggetti geografici. Esso consente di ricollocare gli oggetti geografici provenienti da una fonte per ottimizzarne la corrispondenza con una fonte di riferimento. Questo strumento consente di valutare la corrispondenza tra gli oggetti e di integrare quelli mancanti, evitando duplicazioni e facilitando l'integrazione degli attributi degli edifici provenienti da altre fonti. Tale processo, noto come *conflation*, richiederebbe un enorme dispendio di lavoro visivo che grazie al modello ASIR è stato evitato con risultati eccellenti. Tradizionalmente la qualità dei dati geografici viene valutata confrontando la distanza tra la rappresentazione cartografica e la realtà osservata. L'indicatore ASI, proposto da Ďuračiová (2023), ha permesso un avanzamento importante in questa direzione, in quanto consente di distinguere se due oggetti edilizi sono effettivamente uguali oppure se le differenze derivano unicamente dalla diversa risoluzione o dalla modalità di rappresentazione.

## 1.1 Il georiferimento del catasto: una questione ancora aperta

Il catasto italiano è stato progettato su un sistema di riferimento multiorigine, concepito per garantire la coerenza interna dei singoli “fogli”, considerabili come unità minime di consistenza geografica. Tale coerenza, tuttavia, non assicura l’allineamento con l’origine geografica reale. Il sistema si basa su coordinate Cassini-Soldner e Gauss-Boaga derivate da punti fiduciali, ma prassi e consuetudini amministrative hanno spesso compromesso l’affidabilità del posizionamento automatico, rendendo necessarie correzioni puntuali particolarmente onerose, in termini di quantità di ore lavoro perché richiedono interventi manuali a video e confronto con ortofoto. Questi spostamenti locali possono variare da semplici traslazioni a più complesse roto-traslazioni, richiedendo procedure di rettifica mediante l’individuazione di punti omologhi e trasformazioni analitiche. Nell’ambito della costruzione del Registro degli Edifici dell’Istat sono state sperimentate diverse metodologie per migliorare il georiferimento degli elementi catastali. In alcuni casi è stata realizzata una banca dati dedicata in cui, per ciascun foglio catastale, il sistema di riferimento contenuto nel file geografico è stato corretto con uno spostamento uniforme, così da garantire il riposizionamento automatico nelle versioni successive. Tuttavia, questa procedura si è rivelata estremamente onerosa: il catasto comprende oltre 300.000 fogli di mappa, il cui controllo visivo sistematico sarebbe risultato praticamente impossibile data la limitata disponibilità di risorse. Inoltre, la correzione del sistema geografico di riferimento non include possibilità di rotazioni ma solo di traslazioni per cui potrebbe non riuscire a riposizionare correttamente gli edifici. Il georiferimento si configura quindi come una necessità imprescindibile per colmare la mancanza di una rilevazione diretta degli edifici. Errori di posizionamento avrebbero potuto compromettere per esempio, l’attribuzione delle sezioni di censimento, che rappresentano l’unità statistica territoriale di massima risoluzione per la diffusione dei dati. Per questo motivo, il Registro degli Edifici, fondato prevalentemente su dati catastali, è stato integrato con il Registro degli Indirizzi, così da migliorare l’imputazione delle sezioni mediante il confronto tra coordinate catastali e coordinate associate agli indirizzi amministrativi. Questo perché per quanto gli indirizzi sono spesso diffusi con un qualche georiferimento, non è detto che individuino l’edificio ad essi associati.

## 2 Lavori precedenti: un’analisi critica

Il modello ASIR si basa sul riconoscimento degli oggetti all’interno di diverse basi di dati geografiche, ponendo particolare enfasi sull’identificazione delle coppie di oggetti attraverso forme e posizioni, senza la necessità di codici di collegamento. L’analisi dei lavori precedenti in questo campo ha evidenziato due principali filoni di ricerca: la valutazione della similarità tra poligoni e le tecniche per la sua misurazione, nonché il concetto di *conflation*. Quest’ultimo si riferisce alla fusione di elementi geografici e attributi per arricchire le informazioni all’interno di archivi geografici integrati. Tuttavia, nonostante l’abbondanza di letteratura in merito, l’obiettivo finale rimane la ricerca di metodologie in grado di riconoscere e ricollocare gli edifici tra diverse banche dati senza una codifica comune. Rispetto agli studi precedenti, ASIR introduce la possibilità non solo di individuare corrispondenze tra oggetti, ma anche di ricollocarli automaticamente grazie alla componente di

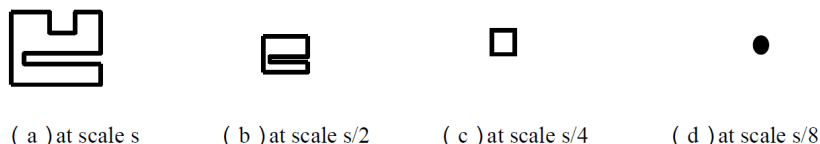
*relocation*. Ciò consente di ridurre in maniera significativa il lavoro manuale e di fornire indicatori quantitativi della qualità del riposizionamento, colmando un vuoto ancora presente in letteratura.

## 2.1 La *conflation* dei dati

L'integrazione di dati geografici, e quindi il loro confronto, sono spesso associati al termine "*conflation*" (Adams et al. 2010; Lei 2020). Questo deriva dal latino "conflat-" che significa "soffiare insieme", evocando l'immagine di voci che si fondono in un'unica melodia. Questa pratica rappresenta un'esigenza fondamentale nella cartografia a larga scala; infatti, non è praticabile disegnare ex novo tutti gli oggetti di una base dati ad ogni occasione. È quindi compito del cartografo trovare soluzioni per integrare fonti dati diverse, pur rispettando i diritti d'uso delle diverse fonti geografiche (Ponti 2007; Coccagna 2011). Questo processo è stato largamente dibattuto nell'ambito dell'integrazione spaziale dei dati (Ďuračiová e Igondová 2017). Il processo comporta una serie di procedure di integrazione spaziale dei dati, quasi sempre basate sulla sovrapposizione geografica (Balasubramani e Cruz 2018), dove non è raro che lo stesso elemento sia rappresentato in modo leggermente diverso e non sia perfettamente posizionato. Diversi studi si sono concentrati sull'analisi della completezza dei dati presenti in OpenStreetMap, tra cui il lavoro di Ďuračiová (2023), che ha fornito la base per il ricalcolo dell'indicatore ASI. L'ASI si distingue dagli altri indicatori noti in letteratura perché non si limita a valutare la completezza di una base dati rispetto all'altra, ma esprime un giudizio qualitativo sulle coppie di oggetti trovati. Quando si confrontano basi dati geografiche diverse per creare relazioni o integrazioni, sorgono inevitabilmente dei dubbi. Il processo di *conflation* prevede due tipologie di riferimento: quella orizzontale e quella verticale (Yuan e Tao 1999). La *conflation* verticale impiega un algoritmo in cui ogni fonte ha la stessa importanza delle altre; pertanto, è necessario individuare una regola per selezionare l'elemento da integrare in ciascun caso. Si prendano, ad esempio, due basi dati A e B. Nella base A è presente un edificio di forma simile d in Figura 1, mentre nella base B è presente un edificio identico alla figura c della Figura 3. Un algoritmo di *conflation* orizzontale dovrebbe essere in grado di riconoscere che i due poligoni si riferiscono allo stesso oggetto e, pertanto, preferire il poligono della base A perché ha un perimetro più esteso. Un algoritmo di *conflation* orizzontale dovrebbe essere in grado di riconoscere che i due poligoni si riferiscono allo stesso oggetto e, pertanto, preferire il poligono della base A. Questo processo diverge da quello verticale, in cui si stabilisce un ordine con cui integrare le diverse fonti. Questa procedura è più semplice da seguire se esiste una conoscenza adeguata delle fonti tale da poter effettuare un ordinamento gerarchico qualitativo (Yuan e Tao 1999). In sintesi, la letteratura sulla *conflation* ha fornito strumenti preziosi per integrare fonti differenti, ma rimane limitata all'individuazione e fusione degli oggetti. Tuttavia, il processo fin qui descritto attraverso la letteratura, non tiene conto del problema del riposizionamento degli edifici e di eventuali problemi topologici conseguenti. Il modello ASIR, proposto in questo lavoro, si pone come un'evoluzione di questi approcci: non solo riconosce le corrispondenze tra oggetti, ma introduce la componente di *relocation*, che permette di ricollocarli automaticamente, riducendo l'intervento manuale e migliorando la

qualità complessiva dei dati integrati, perché posizionandolo al meglio si riducono problemi topologici (ad esempio di *overlapping*).

Figura 1. Modifiche teoriche di un contorno al diminuire della scala geografica

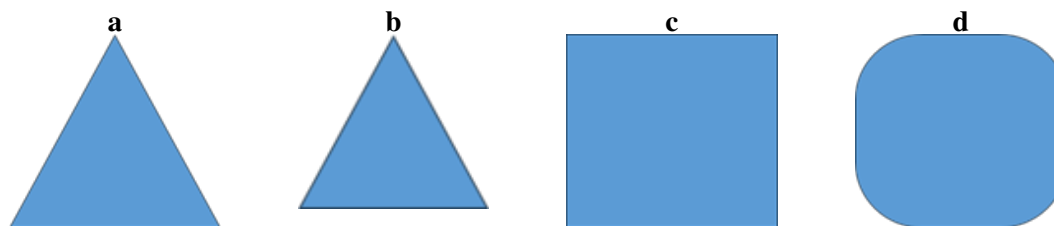


Fonte: Yan 2009

## 2.2 La similarità geometrica

Quando si confrontano due insiemi di oggetti geografici al fine di individuare una relazione che li colleghi tra loro, sorge la domanda se questi oggetti siano compatibili con lo stesso oggetto nella realtà. Ad esempio, se sovrapponessimo due fonti di edifici differenti, la relazione da verificare sarebbe la corrispondenza per sovrapposizione degli oggetti. Tuttavia, questa corrispondenza non garantisce necessariamente la compatibilità e la coerenza con la realtà. Il confronto diretto con ortofoto non è praticabile su numeri elevati di edifici. Per affrontare questa sfida, ci affidiamo al concetto di similarità geometrica. Poiché non possiamo fare affidamento su un codice che correla gli elementi delle diverse basi di dati geografiche, questa relazione deve essere estrapolata dal confronto analitico dei poligoni. La similarità geometrica si basa sul confronto classico della cartografia, che include la similitudine di superfici, perimetri e angoli. Questo approccio ci consente di valutare la coerenza tra gli oggetti geografici senza la necessità di codici di collegamento, offrendo un metodo affidabile per stabilire relazioni significative tra le diverse fonti di dati geografici.

Figura 2. Differenti concetti di similarità

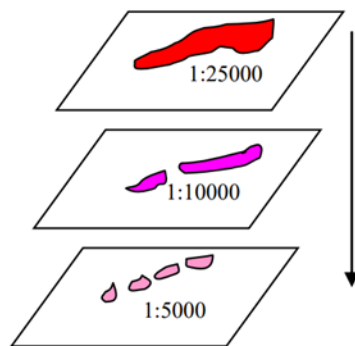


Fonte: elaborazione propria.

Considerare una possibile diversa “grafia” dei cartografi quale criterio consentirebbe di creare delle coppie di oggetti simili tra quelli confrontati in Figura 2. Se si desse un maggior enfasi agli angoli, allora in Figura 1 sarebbero simili i poligoni a e b, ma non c e d. Al contrario, se si desse più valore al perimetro/superficie, vincerebbe la similitudine dei poligoni c e d. Quando si applica questo ragionamento al contesto specifico della cartografia, l’esempio astratto precedentemente illustrato assume un significato differente, poiché

emergono diverse problematiche. In primo luogo, bisogna considerare che, dal punto di vista cartografico, la scala e la risoluzione determinano i contorni degli oggetti in termini di accuratezza posizionale e precisione dei contorni (Thompson 1988). Ad esempio, se consideriamo uno stesso insediamento disegnato cartograficamente a diverse scale geografiche, possiamo osservare come il disegno dello stesso oggetto (in questo caso un insediamento) cambi in base al livello della scala geografica (Yan, 2014). La scala e la risoluzione influenzano fortemente la rappresentazione cartografica: lo stesso oggetto può apparire diverso a scale diverse, rendendo difficile l'automatizzazione del confronto (Figura 3).

*Figura 3. Rapporto scala risoluzione*

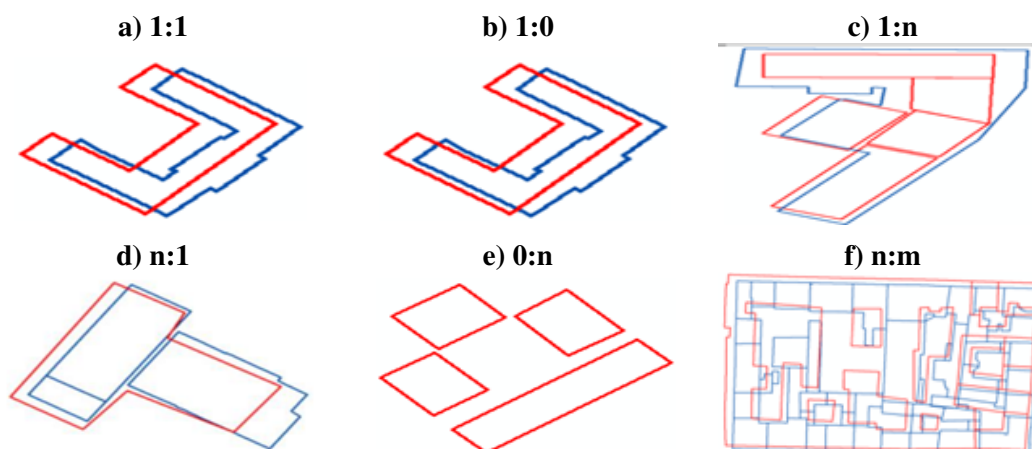


*Fonte: Yan 2010*

La scala e la risoluzione sono strettamente collegate: al crescere della scala aumenta il livello di dettaglio richiesto nella rappresentazione cartografica. Lo stesso oggetto, ad esempio un insediamento, può quindi apparire molto diverso se disegnato a 1:25.000 o a 1:5.000, rendendo complessa l'identificazione automatica come entità unica. In un altro studio, Yan (2009) evidenzia come la rappresentazione di un insediamento possa persino passare da un poligono a un punto, modificando il tipo geometrico ma continuando a descrivere la stessa entità territoriale. Gli esempi riportati mettono in evidenza il problema del confronto di oggetti che rappresentano le stesse entità pur risultando diversi nella loro rappresentazione cartografica. Oltre al caso della scala e risoluzione, esiste un altro problema essenzialmente semantico legato alla definizione di oggetto, per cui un edificio alla stessa scala può essere interpretato in maniera differente, sia per errori sia per una diversa interpretazione dello stesso. In questi casi, è necessario tener conto di questo fattore nel processo di analisi della similarità. Fan et al. (2014) affrontano in maniera esaustiva il problema nella sua organicità. Supponiamo di avere due basi di dati, una rossa e una blu, come mostrato in Figura 4, e di stabilire un criterio per la scelta degli oggetti da confrontare basato sulla sovrapposizione. Gli autori mostrano i diversi casi in cui ci si può ritrovare. Il caso più semplice è la corrispondenza biunivoca degli edifici, come mostrato in Figura 4-a, per arrivare al caso più complicato esposto in Figura 4-f con una relazione molti a molti.

I due casi opposti alla corrispondenza biunivoca sono quelli descritti dalle situazioni "c" e "d" (Figura 4), in cui i poligoni sono scomposti. Ne deriva che qualsiasi confronto direbbe che sono difforni pur identificando lo stesso edificio o gruppo di edifici contigui.

Figura 4. La relazione tra edifici di due banche dati



Fonte: rielaborazione dell'autore su Fan et al. 2014

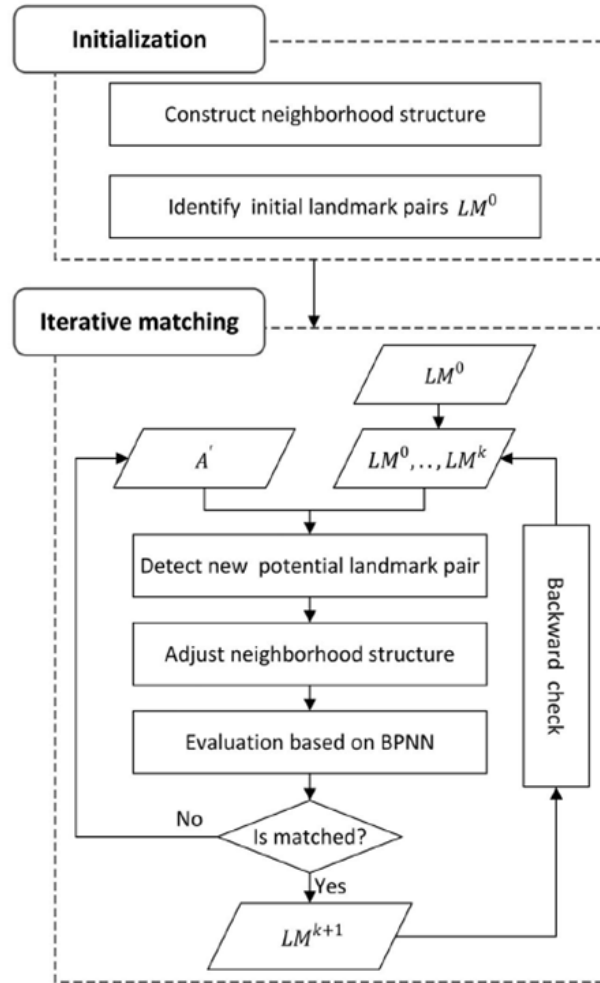
Ben diversi sono i casi "b", e, "d" e (Figura 4). Non si può dire se questi edifici non hanno un corrispondente o se ne esistesse qualcuno in un certo intorno. Il problema dell'identificazione in base al confronto dei parametri di superficie e perimetro risente della risoluzione con cui sono stati disegnati gli elementi. Per questo motivo, sarebbe opportuno che i dati venissero pre-lavorati cercando di rendere omogenee le diverse caratteristiche. Questo aspetto influisce sulle soglie di accettazione del confronto come stesso edificio. Come vedremo nella descrizione del processo adottato, è opportuno fondere gli edifici contigui e quelli entro una certa distanza per attenuare questi effetti.

### 2.3 Analisi delle relazioni geografiche per l'integrazione dei dati

La similarità e la *conflation* risentono dell'insieme degli oggetti considerati simultaneamente nel processo di confronto. Un conto è dire che due poligoni sono più o meno simili, altra storia è dire che una cella di griglia contiene più o meno gli stessi oggetti di una base dati piuttosto che di un'altra. Anche in questo caso va considerato il fine ultimo del confronto, che può riguardare la realizzazione di un indicatore di analisi fine a sé stesso o la costruzione di un indicatore per poter intervenire nel processo di giudizio qualitativo o integrativo. Il tema parte con il riferimento al lavoro Nedas e Egenhofer (2003) precursori di un'analisi dettagliata dei diversi tipi di indicatori in un contesto organico. Solitamente, la relazione alla base del confronto deriva dalla sovrapposizione geografica. Salvucci e Salvati (2021) hanno utilizzato metriche diverse capaci di esprimere giudizi generici senza entrare nel dettaglio della relazione tra gli edifici confrontati. Le metriche più utilizzate si basano su confronti di superficie, perimetro, numero dei vertici e numerosità degli elementi confrontati. Emergono due filoni in letteratura: il più diffuso esprime giudizi di completezza per area, il secondo confronta coppie di oggetti. Nel primo caso rientra il lavoro di Hecht (2013), che parte dalla suddivisione dello spazio in una griglia esagonale esprimendo giudizi relativi alla composizione degli edifici nella cella, mentre in Salvucci e Salvati (2021) si è fatto uso delle sezioni di censimento. I lavori che fanno riferimento a giudizi puntuali

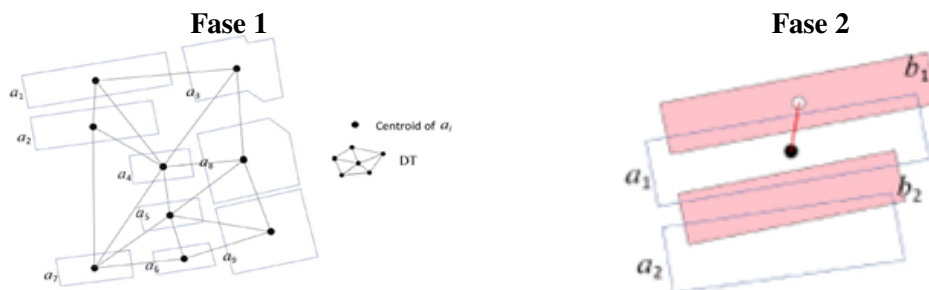
spesso non sono automatizzati e richiedono un intervento visivo, come ad esempio Franzini, Annovazzi-Lodi, e Casella (2020), che individuano in maniera puntuale gli edifici attraverso GoogleMap, un processo improbabile per esprimere un giudizio su numeri di confronti particolarmente elevati. Recentemente si sta abbandonando l'idea che la relazione da verificare sia solo quella per sovrapposizione, mettendo in discussione quale siano le coppie di elementi da dover confrontare. In tal direzione si apre il percorso per un'analisi più complessa esposta in Liu e coautori (2020), che abbandona il presupposto del confronto basato sulla sovrapposizione. L'algoritmo riportato in Figura 5 riassume tutto il ragionamento fin qui esposto e che costituisce uno dei capisaldi del presente lavoro. Si parte dalla costruzione di una matrice di vicinato, dove per ogni oggetto, in questo caso edificio, si individuano i corrispondenti da confrontare entro una certa distanza. Stabilita una metrica di confronto geometrico, l'edificio viene confrontato con tutti gli altri dell'altra fonte dati trovando le relazioni vere, ovvero quelle che hanno la più alta similarità. Questo risolve i casi b ed e di Fan et al. (2014) riportati in Figura 4 perché cerca un possibile corrispondente entro un certo raggio di distanza. Quindi in estrema sintesi Liu et al. 2020, data una matrice di distanza dei centroidi degli edifici, filtra le relazioni da esaminare entro un raggio prefissato ed identifica la relazione vera in base ad una metrica, come riepilogato in Figura 6.

Figura 5. Algoritmo di confronto proposto da Liu et al.(2020)



Fonte: Liu et al. 2020

Figura 6. I due momenti a confronto per il riconoscimento degli edifici corrispondenti

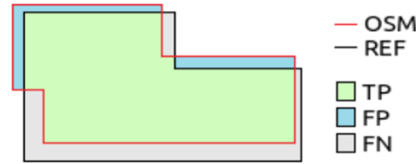


Fonte: Rielaborazione su Liu et al. 2020

## 2.4 Metriche per il confronto dei dati geografici

Alcune delle metriche utilizzate in questi processi di *conflation* sono state ampiamente studiate (Yan 2010), e ne esamineremo alcune per illustrare la complessità del tema. Il metodo geometrico di riconoscimento più diffuso si basa sul confronto tra la superficie intersecata e la superficie iniziale degli edifici delle due basi dati. Un esempio eloquente di questo approccio è fornito da Brovelli et al. (2016), come illustrato nella Figura 7.

Figura 7. Superficie di intersezione quale indice di similarità



Fonte: Brovelli et al. 2016

Questo indicatore è ampiamente utilizzato in letteratura ed è abbastanza intuitivo, ma il problema principale risiede nella determinazione della soglia oltre la quale si decide se le due sagome appartengono allo stesso edificio o meno. Le differenze possono derivare dalla diversità nella risoluzione con cui sono state disegnate le due sagome, da problemi di traslazione e possono portare a risultati completamente errati, anche se riguardano lo stesso oggetto.

Gli indicatori proposti per il confronto delle superfici dei poligoni sono piuttosto articolati e richiamano una vasta letteratura sul confronto delle immagini. Il principio guida per identificare un caso di uguaglianza tra due poligoni è che:

$$A = A \cap B = B$$

Questo principio stabilisce che due poligoni sono considerati uguali se la loro intersezione è uguale all'area di entrambi i poligoni stessi. L'equazione esprime chiaramente il concetto di uguaglianza, poiché  $A = B$  se e solo se la loro intersezione è uguale ai poligoni stessi. Tuttavia, nella pratica, persino la stessa persona potrebbe non essere in grado di disegnare lo stesso edificio senza spostare nemmeno un vertice. Pertanto, il confronto deve basarsi su indicatori del tipo di quello di Jaccard. Le metriche proposte per comprendere questa dimensione includono l'indice di Jaccard (1901) che è stato utilizzato nella versione Tanimoto (1958). L'indicatore originario di Jaccard si basa sull'equazione precedente considerando il seguente rapporto:

$$\text{sim}_J = \frac{A \cap B}{A \cup B}$$

Il rapporto può valere 1 se e solo se  $A = B$ . Per facilitare il calcolo, evitando di calcolare sia l'intersezione che l'unione, la versione Tanimoto considera:

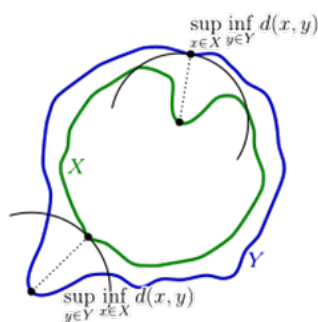
$$\text{sim}_T(A, B) = \frac{A \cap B}{A + B - A \cap B}$$

Perfettamente identica alla precedente ma molto più veloce in termini computazionali. Un'altra metrica utilizzata è quella di Sorensen (1948) leggermente diversa dalle precedenti, che parte sempre dall'equazione iniziale facendo una diversa considerazione: se due insiemi sono uguali, allora:

$$\text{sim}_T = \frac{2(A \cap B)}{A + B}$$

Questi indicatori sono abbastanza semplici da calcolare ma nell'ASI proposto da Ďuračiová (2023) il confronto delle sagome è molto più complesso. In particolare, ci si sofferma su alcuni confronti nel perimetro tra le due sagome. Si è potuto constatare che questo fatto, più degli altri, è fondamentale per migliorare il riconoscimento degli edifici. Bisogna considerare che la forma degli edifici è in parte imposta dai piani regolatori, che applicano quote di superfici edificabili uguali per quartieri dove i lotti sono molto simili. Questo determina che la superficie degli edifici non è un indicatore robusto per distinguere e identificare le coppie di edifici omologhi tra le diverse fonti. Migliore è la proposta di Ďuračiová, che si basa sulla distanza di Frechet e quella di Hausssdorf e che riesce a confrontare i contorni degli edifici. Applicazioni di queste metodologie si ritrovano nei campi più disparati. A solo titolo di esempio e soprattutto per capire le potenzialità degli algoritmi proposti si rimanda a Yu et al. (2022) che propone un'applicazione basata sulla distanza di Hausssdorf per il riconoscimento della forma regolare delle patate. Come noto, questa distanza rappresenta la massima distanza tra le massime differenze tra i due poligoni. In campo geografico questa distanza è stata largamente utilizzata per capire quali dati poter integrare (Yuan e Tao 1999) e per poter effettuare matrici di confronto tra oggetti appartenenti a diverse fonti (Min, Zhilin, e Xiaoyong 2007). Più utilizzata è la distanza di Frechet per l'integrazione dei dati geografici, nasce in applicazione di integrazione di grafi stradali basandosi sul confronto di linee.

Figura 8. Esempificazione del calcolo della distanza di Hausssdorf

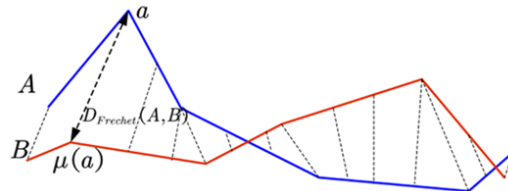


Fonte: Rocchini – Wikipedia

Ad esempio, Savary e Zeitouni (2005) evidenziano come i tracciati delle linee siano più o meno difforni tra loro pur evidenziando similitudini. Si pensi ad esempio alle tracce GPS ricavabili dai percorsi automobilistici. L'intuizione di Ďuračiová, di applicare a dei poligoni una metrica nata per le linee offre un'opportunità di una accuratezza maggiore nel riconoscimento geometrico riducendo gli accoppiamenti fallaci. Si sottolinea che nel-

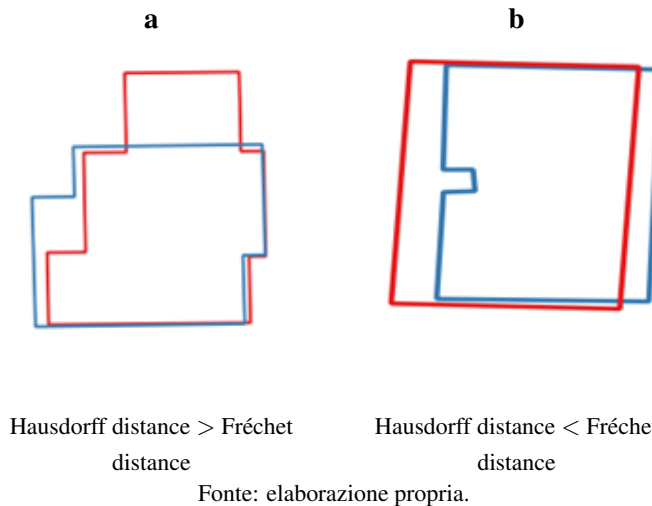
l'ambito dell'ASI i poligoni sono considerati delle linee chiuse e pertanto si applicano le metodologie di confronto tipiche di queste geometrie che risultano molto più discriminanti rispetto a quelle delle superfici dei poligoni. Per quanto riguarda la distanza di Frechet consente di stabilire la distanza minima tra le due linee durante il percorso (Alt e Godau 1995). L'esempio classico è quello del percorso effettuato da un cane e il suo padrone, la distanza individuata è molto simile al concetto della lunghezza minima che deve avere il guinzaglio durante il tragitto, se entrambi percorrono il percorso nella stessa direzione. Si tratta dunque di un indice di similarità molto forte il cui algoritmo è stato riprodotto non essendo disponibile tra i moduli python attualmente in circolazione.

Figura 9. Esempificazione dell'individuazione della distanza di Frechet



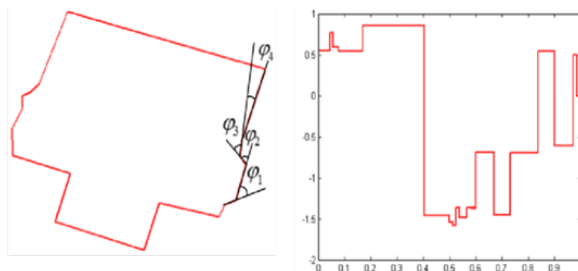
Fonte: <https://enricogiannini.com/> (ultimo accesso 25/06/2025)

Figura 10. Esempificazione del diverso comportamento delle distanze tra i perimetri considerati



Osservando la Figura 10, la distanza di Frechet sarebbe più sensibile a variazioni minime mentre quella di Hausdorff evidenzia casi più eclatanti. Esistono altre funzioni utilizzabili come, ad esempio, quella di Arkin utilizzata da (Fan et al. 2014) che è molto simile ad una distanza di Frechet. Arkin propone di calcolare l'integrale della funzione che esprime una relazione tra la misura del lato in ascisse e l'arcotangente in ordinata. Due poligoni sono simili se hanno un risultato pressoché identico vale a dire la differenza tra gli integrali

Figura 11. Esempificazione della funzione di Arkin



Fonte: Fan et al. 2014

pari a zero. Per quanto la funzione sia intuitiva il calcolo è estremamente oneroso, non essendo tra l'altro disponibile alcun modulo python capace di eseguirlo.

## 2.5 La difficoltà del riconoscimento degli edifici e i tentativi precedenti

Nel tentativo precedente di creare una matrice di accoppiamento automatica, la metrica utilizzata si basava principalmente sul confronto delle superfici e dei perimetri degli edifici. Tuttavia, questi tentativi sono falliti a causa della notevole somiglianza tra gli edifici circostanti. Questa somiglianza è attribuibile a ragioni storiche e amministrative, come i vincoli imposti dal piano regolatore che stabiliscono regole uniformi per varie parti del comune. Nei tentativi precedenti una volta posizionato il vettoriale del catasto su una base cartografica corretta, veniva stabilito un raggio di ricerca e si spostavano gli edifici cercando il più simile. I tentativi fallivano perché il numero dei falsi positivi era enorme e non consentiva di discriminare correttamente. Si pensi ad un poligono banale quale può essere un rettangolo, data superficie e perimetro quanti rettangoli si possono ottenere?

Dato che

$$\text{sup} = l \times b$$

E allora

$$b = \frac{\text{sup}}{l}$$

il perimetro sarebbe:

$$P = 2l + \frac{\text{sup}}{l}$$

Quindi dato P e sup il numero di rettangoli ottenibili dipende esclusivamente da l ovvero un numero infinito. Considerando che gli edifici possono assumere un numero di lati indefinito e quasi sempre irregolari si può comprendere il volume dei falsi positivi possibili. Questo problema si è dimostrato insormontabile e si è dovuto abbandonare in favore di riposizionamenti manuali.

### 3 L'ASI e gli indicatori elementari

L'utilizzo dell'indicatore ASI restituisce una classificazione degli edifici confrontati.

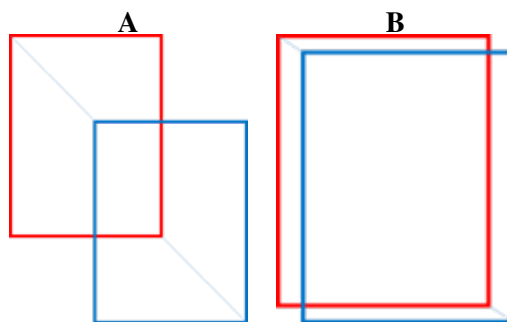
L'indicatore è una sintesi di varie misure ottenuta con tecniche fuzzy set che garantiscono un elevato grado di identificazione degli oggetti. Esaminiamo nel dettaglio le misure considerate in Tabella 1. Nell'algoritmo proposto da Ďuračiová si distinguono le misure che tengono conto del posizionamento geografico relative dei due elementi a confronto con quelle che vengono effettuate a prescindere dalla loro posizione, degli oggetti geografici rappresentanti gli edifici provenienti da fonti diverse. Le misure che considerano la Distanza Hausdorff e quella di Fréchet assumono un significato solo se si considera dove sono posizionati i due elementi. L'ASI effettua la classificazione di due basi dati che si presuppone siano state collocate bene nello spazio geografico. Se questa condizione viene meno bisognerà riconoscere la somiglianza dei sedimi degli edifici. Lo stesso edificio in due basi dati differenti avrà delle distanze di Hausdorff e Fréchet diverse se uno dei due è spostato rispetto all'altro. Si veda in Figura 12 relativamente alla distanza di Hausdorff, ossia la distanza massima tra i due poligoni. Se si confronta lo stesso identico poligono, ma questi non sono allineati si ottiene una distanza maggiore di quella effettiva Figura 12-A che in questo caso sarebbe pressoché nulla Figura 12-B

*Tabella 1. Sintesi degli indicatori utilizzati per il calcolo dell'ASI rettificato*

| <b>Posizionamento relativo</b> | <b>Dimensione di confronto</b>                   | <b>Indicatori utilizzati</b>                              |
|--------------------------------|--------------------------------------------------|-----------------------------------------------------------|
| Sovrapposizione edifici        | Accuratezza della forma, diversità tra le sagome | Distanza di Hausdorff; Distanza di Fréchet                |
| Sovrapposizione edifici        | Accuratezza posizionale                          | Indice di somiglianza di Jaccard; Indice di Sørensen–Dice |
| Senza riferimento              | Somiglianza delle sagome                         | Somiglianza di superficie, perimetro e numero di vertici  |

Fonte: elaborazione propria.

*Figura 12. Esempificazione della distanza di Hausdorff tra due poligoni identici ma posizionati in maniera difforme*



Fonte: elaborazione propria.

L'algoritmo originario di Ďuračiová da cui si ottiene l'indicatore ASI, tiene conto di

queste particolarità degli indicatori in un processo logico di scelta. Come in tutte le logiche fuzzy set gli indicatori elementari vengono riproporzionati, in questo caso con tecniche min-max con range 0,1. L'algoritmo prevede tre momenti successive su tre diversi aspetti del confronto delle forme:

- a Similitudine tra le distanze fra i perimetri delle forme nella loro posizione
- b Similitudine tra le superfici intersecate nella loro posizione
- c Similitudine tra forme, superfici e perimetro in assoluto

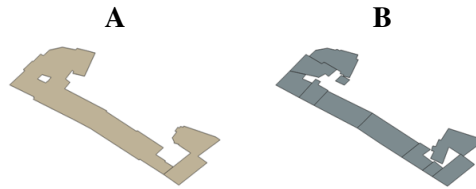
Il punto a,b dell'algoritmo ASI viene calcolato in base alla posizione effettiva mentre il punto c considera la similitudine tra le forme in astratto in base alle loro caratteristiche. In estrema sintesi il valore dell'ASI viene scelto tra gli indicatori di partenza secondo questa sequenza:

1.  $sim_d = Max$  (distanza Hausdorff, Frechet normalizzate)
2.  $sim_s = Minimo$  (similitudine Tanamoto, Sørensen–Dice)
3.  $sim_h = Minimo$  (similitudine Area, Perimetro, Vertici)
4.  $Max(sim_d, sim_s)$
5.  $Minimo(punto4, 3)$

Si rimanda al testo originale Ďuračiová (2023) per l'approfondimento delle singole componenti. Il valore dell'ASI potrebbe restituire un valore che non consente di identificare lo stesso oggetto considerando il peso della posizione geografica degli elementi; tuttavia, se le forme sono molto simili per valori di perimetro e superficie questo diventa un indicatore di un effetto traslazione per cui lo stesso oggetto è stato identificato ma uno dei due è spostato rispetto all'altro. Attenzione questa deduzione vale solo perché Ďuračiová mette in confronto oggetti tra loro sovrapposti. Un altro problema specifico del confronto tra base dati relative agli edifici è la diversa interpretazione dell'unità minima di riferimento. Lo stesso edificio potrebbe essere stato disegnato diviso nell'altra banca dati, o viceversa. Ovviamente gli edifici risulterebbero difforni pur rappresentando la stessa entità. A tal proposito, si osservino le Figure 13 A e B dove, nonostante il perimetro sia identico, la Figura 13-A riporta un insieme di edifici che in Figura 13-B sono suddiviso all'interno dello stesso perimetro in maniera difforme. Qualsiasi tipo di confronto restituirebbe un esito negativo. Ďuračiová, propone di considerare dei cluster di edifici costituiti dalla fusione degli stessi entro una certa distanza. Da test preliminari effettuati nell'elaborazione che si sta descrivendo questa distanza è stata posta pari a 1 metro. In pratica prima di effettuare il confronto sono stati costruiti i cluster di edifici di entrambe le basi dati con una operazione di buffer di 1 metro e successiva fusione.

Questa operazione attenua il problema della diversa risoluzione delle due fonti perché consiste in una sorta di *smoothing* dei poligoni ingrossati di 1 mt. L'algoritmo elaborato

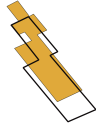
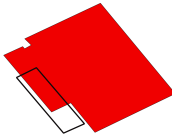
Figura 13. Esempificazione del problema della difformità degli edifici e possibili soluzioni proposte



Fonte: elaborazione propria.

da Ďuračiová è stato interamente riprodotto e applicando gli stessi parametri per la classificazione dell'ASI si ottengono le 4 tipologie di edifici. Si faccia attenzione che il calcolo viene effettuato sui cluster ma il tipo viene riportato sul dato dell'edificio iniziale. Questo ci consente una più facile interpretazione dei dati e ci permette di capire se il posizionamento effettuato degli edifici catastali risulta o meno corretto. L'obiettivo qualitativo è dunque quello di ridurre al minimo la quota di tipo 0, ovvero i non riconoscibili. Gli identici e quelli a differente risoluzione non devono essere ulteriormente spostati. Il tipo 3 traslato può essere oggetto di un semplice intervento successivo con uno spostamento da un centroide all'altro facendo verosimilmente diventare la nuova tipologia in 1 o 2 una volta spostato.

Tabella 2. Classificazione dei tipi di corrispondenza utilizzati nel calcolo dell'ASI

| Tipo   | Declaratoria           | Algoritmo                                              | Esempio                                                                               |
|--------|------------------------|--------------------------------------------------------|---------------------------------------------------------------------------------------|
| Tipo 1 | Identico               | $ASI > 0.85$                                           |  |
| Tipo 2 | Differente risoluzione | $sim_d > 0.75$ ; $sim_s > 0.75$ ;<br>$sim_{sh} > 0.75$ |  |
| Tipo 3 | Traslato               | $sim_v > 0.75$ ; $sim_s < 0.75$ ;<br>$sim_{sh} > 0.75$ |  |
| Tipo 4 | Non riconoscibile      | —                                                      |  |

Fonte: elaborazione propria.

#### 4 Il sistema delle matrici: linkage similarità

Il pregevole lavoro di Ďuračiová, esposto brevemente al paragrafo precedente si basa sul confronto tra un oggetto e il corrispondente per posizione. Questa situazione idilliaca viene meno se le due basi dati sono traslate l'una rispetto all'altra per una qualsiasi ragione. Lo schema proposto da Liu e coautori (2020) parte da una sistemazione iniziale dei dati, in particolare si prende in esame il caso che esista una mancata sovrapposizione. Il problema è affrontato considerando due matrici contemporaneamente quelle delle distanze e quella delle somiglianze. Per ogni edificio sulla base del centroide vengono individuati gli edifici corrispondenti nella base dati di riferimento entro un certo raggio. Per questi edifici vengono individuate le coppie possibili che identificano gli stessi oggetti solo se valgono certi parametri per gli indici di somiglianza scelti, che vanno a costituire la matrice di somiglianza. La situazione descritta da Liu e coautori ben si addice al nostro registro dove in realtà non sappiamo misurare lo scostamento tra il catasto georiferito da Istat e il DBSN nazionale IGM. Nell'esempio riportato in Figura 14 in reticolato blu gli edifici di riferimento del DBSN Nazionale prodotte da IGM mentre in grigio il risultato del posizionamento manuale la doppia freccia individua la coppia di punti omologhi, attraverso i quali si dovrebbe effettuare lo spostamento. La figura mostra la complessità della scelta ed individuazione dei punti omologhi.

*Figura 14. Sovrapposizione del Dbsn (IGM) con il prodotto del georiferimento iniziale del Catasto effettuato da Istat*



Fonte: Abbatini 2024

## 5 ASIR: dall'indicatore ASI al modello di riconoscimento e spostamento

Integrando i contributi di Ďuračiová e di Liu e coautori (2020) è stato sviluppato l'algoritmo riportato in Figura 15, articolato in due fasi principali. In riconoscimento al lavoro originario di Ďuračiová, l'algoritmo è stato denominato ASIR. È fondamentale precisare che è un modello logico volto al riconoscimento e al ricollocamento degli oggetti geografici, e non deve essere confuso con l'ASI, che è invece un indicatore di similarità di forma proposto da Ďuračiová. La prima fase di preparazione riguarda la confrontabilità geografica dei dati: tutti gli edifici vengono riportati nel sistema di riferimento EPSG:32632, che, in quanto proiettato, consente di effettuare misurazioni in metri. Successivamente si procede all'identificazione degli aggregati, fondendo in un unico poligono gli edifici situati a meno di un metro di distanza. A questo punto è possibile calcolare le diverse metriche di similarità: in base al posizionamento iniziale delle due basi dati vengono costruite le matrici delle distanze, selezionando per ciascun edificio catastale gli edifici corrispondenti entro una soglia prefissata. La seconda fase è resa possibile dalla costruzione di un geodatabase con una struttura innovativa, che consente di confrontare le geometrie delle coppie indipendentemente dal loro contesto geografico. Diversamente dai sistemi tradizionali di archiviazione (ad esempio GDB o shapefile), l'impiego del geodataframe di GeoPandas permette di trattare la geometria come una variabile qualsiasi, rendendo le operazioni di confronto più semplici e veloci.

### 5.1 Il riposizionamento: ASI Relocation (ASIR)

L'integrazione delle due procedure consente di introdurre una fase ulteriore, volta al riposizionamento automatico degli edifici. Una volta identificata la matrice degli eleggibili, per ogni coppia di edifici simili si conosce il centroide: lo spostamento dell'intero foglio catastale secondo questo vettore permette di migliorare il posizionamento del database. Per limitare i falsi positivi, particolarmente frequenti in aree caratterizzate da edifici tra loro molto simili, il riposizionamento viene validato attraverso una nuova valutazione ASI a livello di foglio catastale, che misura la variazione nella frequenza degli edifici non riconosciuti. Nella sperimentazione il raggio di confronto è stato fissato a 500 m. Le geometrie delle coppie sono state traslate virtualmente in modo da far coincidere i centroidi, consentendo di calcolare indicatori robusti, come la distanza di Hausdorff e l'indice SISO:

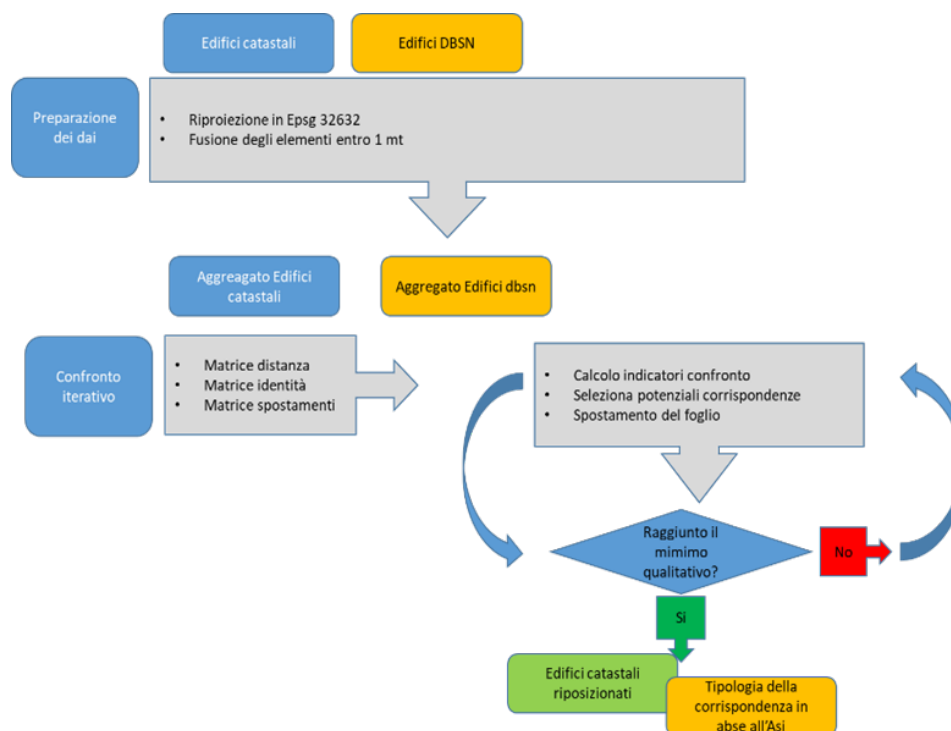
$$\text{SISO}(A, B) = \min(\sup_A, \sup_B)$$

Tra tutte le coppie possibili sono considerate valide quelle con distanza di Hausdorff  $< 5$  m e  $\text{SISO} > 0,85$ . Questo criterio fornisce una buona approssimazione dell'uguaglianza tra forme, evitando di calcolare l'ASI per singole coppie, operazione inefficiente poiché basata su normalizzazioni di gruppo.

Infine, per selezionare lo spostamento ottimale, è stata definita una misura sintetica di qualità:

$$\text{score} = \frac{n_{t1} + 2n_{t2} + n_{t3}}{2n}$$

Figura 15. Algoritmo utilizzato nel modello ASIR



Fonte: elaborazione propria

dove i tipi 1 (identici) e 2 (diversa risoluzione) hanno peso doppio rispetto al tipo 3 (edifici da spostare). Il valore massimo  $2n$  corrisponde alla situazione in cui tutti gli edifici risultino correttamente collocati. Lo spostamento migliore è quello con lo score più elevato, tenendo conto anche della riduzione della quota di edifici non riconosciuti (tipo 0).

## 6 Evidenze empiriche da un caso studio: il comune di Cassino (FR)

Il processo ASIR consente di migliorare notevolmente il sistema di georiferimento con delle ricadute per tutto il sistema del Registro dei Luoghi, tenuto dall'Istat. Il registro degli edifici a supporto del censimento edifici si compone di una parte cartografica che serve a geocodificare la sezione di censimento in cui ricade l'edificio. Questa assegnazione avviene attraverso un processo geografico che confronta la sezione in cui ricade l'edificio e la confronta con quella in cui ricadono gli edifici ad esso associate. Questa soluzione è stata scelta perché non si era ancora raggiunto un livello di accuratezza posizionale così elevato da utilizzare soltanto il posizionamento degli edifici. Si consideri che il modello ASIR, da noi sviluppato, ha lo scopo di ricollocare gli elementi catastali sulla base della loro identificazione della fonte obiettivo, in questo caso il DBSN. Considerando il processo di produzione della mappa catastale, si ritiene che l'unità minima da considerare sia il foglio di mappa.

## 6.1 Risultati ottenuti: il caso del comune di Cassino

La sperimentazione del modello ASIR è stata condotta sul comune di Cassino, scelto per la complessità del tessuto urbano e per la necessità di un significativo riposizionamento. Il contesto si caratterizza per una densità edilizia elevata e una rete stradale articolata, condizioni che rendono particolarmente difficile l'operazione manuale di riallineamento: un lavoro che richiederebbe ore di intervento è stato invece completato dall'algoritmo in circa 3 ore, in modalità completamente automatizzata. La Tabella 3 mostra la classificazione degli edifici prima e dopo l'applicazione di ASIR.

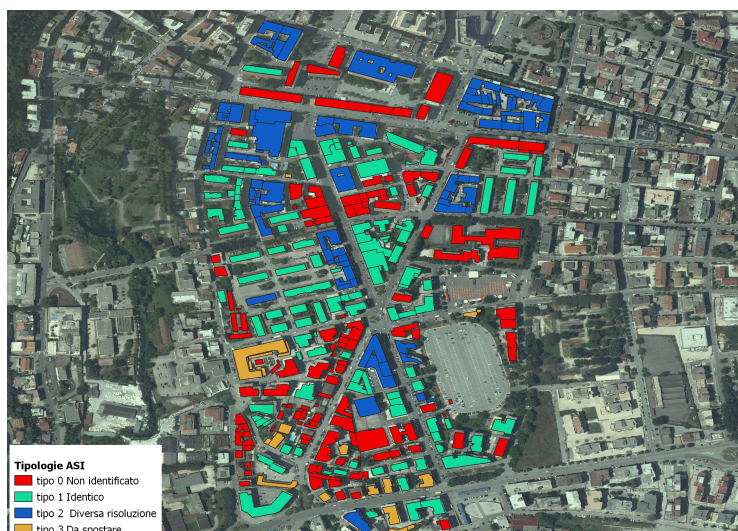
*Tabella 3. Classificazione degli edifici del comune di Cassino secondo le tipologie dell'indicatore ASI prima e dopo l'applicazione dell'ASIR e il conseguente spostamento*

| <b>Tipo ASI</b>              | <b>Senza ASIR</b> | <b>% Senza ASIR</b> | <b>Con ASIR</b> | <b>% Con ASIR</b> |
|------------------------------|-------------------|---------------------|-----------------|-------------------|
| Tipo 0 – non riconosciuto    | 23 686            | 92,91               | 12 794          | 50,18             |
| Tipo 3 – spostato            | 1 808             | 7,09                | 6 270           | 24,59             |
| Tipo 2 – diversa risoluzione | 0                 | 0,00                | 812             | 3,19              |
| Tipo 1 – identici            | 0                 | 0,00                | 5 618           | 22,04             |
| Totale edifici               | 25 494            | 100,00              | 25 494          | 100,00            |

La situazione iniziale era critica: nessun edificio risultava correttamente identificato e la quasi totalità (93%) era classificata come “non riconosciuta”. Dopo l'applicazione di ASIR, oltre 6.400 edifici sono stati classificati come identici o con diversa risoluzione, mentre la quota di “non riconosciuti” si è ridotta drasticamente al 50%. Il modello ha inoltre individuato oltre 6.200 edifici da spostare, in gran parte sovrapposti a sagome quasi identiche, segnalando quindi corrispondenze potenzialmente valide. Dal punto di vista scientifico, il risultato dimostra la capacità del modello di andare oltre una semplice misura di distanza media, fornendo una classificazione qualitativa degli edifici tramite l'indicatore ASI. Questa distinzione tra tipologie (identici, diversa risoluzione, traslati, non riconosciuti) consente non solo di valutare la qualità del riallineamento, ma anche di interpretare le cause delle differenze. Per la statistica ufficiale, il caso di Cassino evidenzia le potenzialità di ASIR come strumento per migliorare la qualità del Registro degli edifici e, più in generale, del Registro dei luoghi. La riduzione della quota di edifici non riconosciuti e l'automatizzazione del processo riducono il fabbisogno di lavoro manuale e rafforzano il ruolo dell'Istat come validatore e integratore di basi territoriali eterogenee. In Tabella 3 sono riportate le frequenze delle tipologie dopo il georiferimento iniziale. La situazione era a dir poco catastrofica dal momento che nessun edificio era stato identificato. Ne consegue che i 1808 edifici da spostare molto probabilmente sono dei falsi positivi. L'applicazione del modello ASIR ha individuato 5618 edifici identici e 812 con una risoluzione diversa, ma nella stessa posizione. Ancora 6270 edifici sono da spostare ma si sono sovrapposti ad edifici praticamente identici. Se gli edifici fossero tutti identici tra le due banche dati, si tratterebbe della stessa fonte. Ragionevolmente se dopo la ricollocazione del foglio la quota di edifici non riconosciuti è superiore al 40% si ritiene che si tratti di un errore. La situazione iniziale del comune di Cassino era particolarmente critica: nessuno dei 154 fogli di mappa risultava correttamente collocato e, di conseguenza, il sistema dei registri avrebbe dovuto basarsi esclusivamente sulle coordinate degli indirizzi. L'applicazione del modello ASIR

ha invece permesso di ricollocare correttamente 130 fogli in circa tre ore di elaborazione, in un processo interamente automatizzato e privo di intervento umano. Rimangono 24 fogli da rivedere, ma il miglioramento complessivo è evidente. A titolo esemplificativo, la Figura 16 mostra il risultato ottenuto per il foglio 86: la tipologia 1 (edifici identici) risulta prevalente e, pur in presenza di una quota residua di edifici non riconosciuti, questi sono comunque correttamente posizionati.

*Figura 16. Comune di Cassino foglio 86 del Catasto, risultato dell'Algoritmo ASIR per tipologia su ortofoto Geoportale Nazionale*



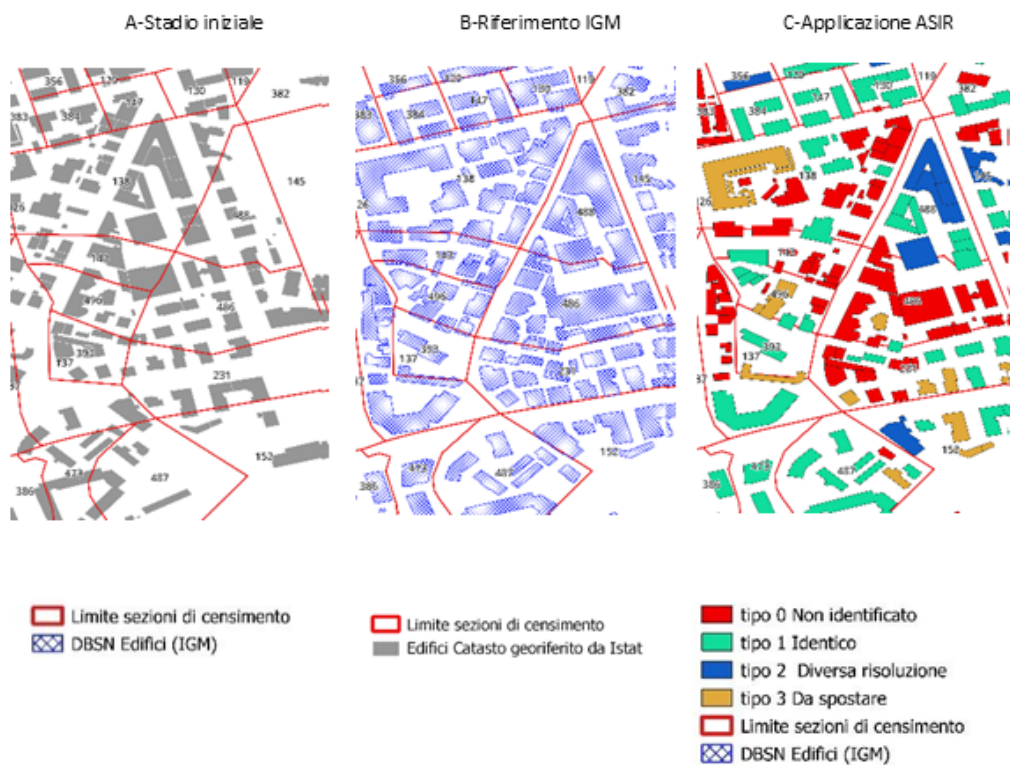
*Fonte: Abbatini (2024)*

## 7 Integrare informazioni con criteri geografici

Per l'Istat, l'obiettivo principale del georiferimento catastale è l'assegnazione corretta della sezione di censimento. Dopo il 2011, in assenza di una rilevazione sul campo, questa operazione è stata effettuata mediante sovrapposizione di archivi e controlli incrociati. La Figura 17 illustra le difficoltà tipiche del processo: se i poligoni catastali risultano mal posizionati (17-A), si producono errori di attribuzione delle sezioni; al contrario, i poligoni IGM correttamente georiferiti (17-B) garantiscono coerenza con la realtà territoriale. La ri-localizzazione ottenuta con ASIR (17-C) dimostra che è possibile migliorare sensibilmente l'accuratezza posizionale, aprendo la strada a un uso scientificamente più solido dei dati catastali. Questo miglioramento non ha solo valore cartografico: consente infatti di integrare ulteriori informazioni con criteri geografici, ad esempio attraverso il *reverse geocoding* degli indirizzi o l'attribuzione dell'altezza degli edifici a partire da dati europei (es. JRC).

La corretta posizione degli edifici catastali apre la possibilità di integrare informazioni aggiuntive. Un esempio rilevante riguarda gli indirizzi: l'archivio catastale li associa alle unità immobiliari, ma in modo incompleto e con frequenti errori ortografici. L'integrazione

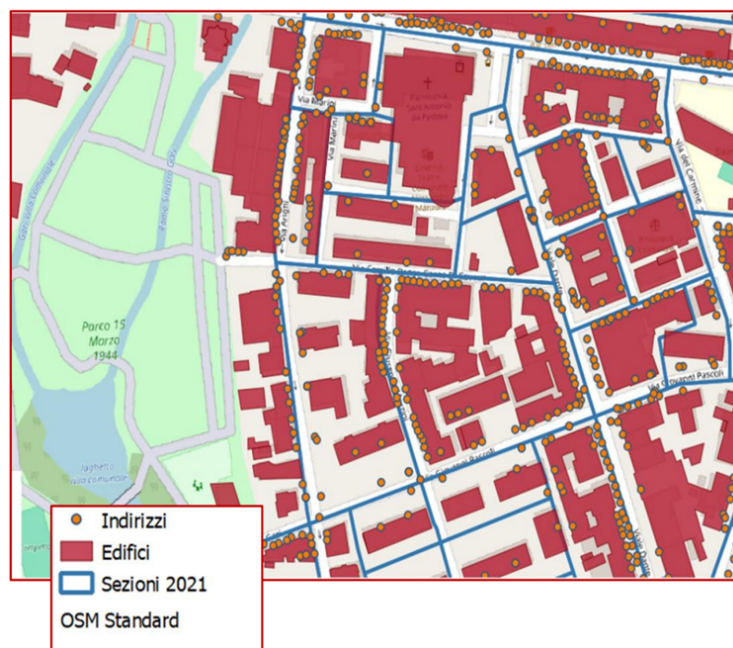
Figura 17. Individuazione della sezione di censimento e livello di qualità dell'informazione rilevata per il comune di Cassino



Fonte: elaborazione propria.

con basi di dati di indirizzi puntuali può essere realizzata tramite reverse geocoding, attribuendo a ciascun indirizzo l'edificio più vicino. Questa tecnica diventa affidabile solo a fronte di un elevato livello di accuratezza posizionale. Un limite riguarda i numeri civici che servono più edifici: in assenza di ripetizioni, il processo rischia di produrre sotto copertura nelle associazioni. Tuttavia, con il corretto posizionamento degli edifici e delle coordinate di accesso, il *reverse geocoding* permette di associare in modo sistematico indirizzi ed edifici, come mostrato in Figura 18.

Figura 18. Confronto edifici riposizionati con ASIR e indirizzi per il reverse geocoding



Fonte: Abbatini 2024

Un'ulteriore possibilità offerta dal corretto posizionamento riguarda l'integrazione di informazioni derivate da altre fonti geografiche. Un esempio significativo è l'attribuzione dell'altezza degli edifici, ricavata dallo strato prodotto dal JRC della Commissione Europea (Pesaresi e Politis 2022). Si tratta di un dato particolarmente utile, solo parzialmente disponibile nel catasto. Per integrare questa informazione, è stato utilizzato il centroide degli edifici riposizionati con ASIR (Figura 19-A); il valore di altezza estratto è stato quindi assegnato al poligono catastale corrispondente (Figura 19-B). In questo modo, la ricollocazione degli edifici diventa la base per arricchire gli archivi con variabili strutturali fondamentali, con ricadute dirette sulle analisi statistiche e territoriali.

Figura 19. Estrazione dell'altezza dell'edificio in base al centroide riposizionato con ASIR e assegnazione del poligono



Fonte: Abbatini 2024

## 8 Conclusioni

Il modello geografico ASIR di riconoscimento e ricollocazione degli edifici del registro consente di affrontare fin da subito il problema della qualità dei dati geografici. Tale problema viene gestito grazie all'indicatore ASI, che permette di valutare in modo qualitativo il grado di accuratezza posizionale raggiunto. In questo modo l'operatore può comprendere se sia necessario procedere con ulteriori ricollocazioni o se il risultato ottenuto possa essere considerato accettabile. L'integrazione della componente di Relocation rende ASIR uno strumento innovativo: non si limita a misurare la similarità tra oggetti, ma fornisce una ricollocazione automatizzata che riduce drasticamente il fabbisogno di interventi manuali. Inoltre, la classificazione fornita da ASI (identici, diversa risoluzione, traslati, non riconosciuti) arricchisce la valutazione, andando oltre una semplice misura di distanza media. Dal punto di vista della statistica ufficiale, l'adozione di ASIR contribuisce a migliorare la qualità e la coerenza del Registro degli edifici e, più in generale, del Registro dei luoghi, rafforzando il ruolo dell'Istat come integratore e validatore di basi territoriali eterogenee. Il corretto posizionamento degli edifici abilita inoltre l'integrazione di ulteriori informazioni, come indirizzi e altezze, con ricadute dirette sulla produzione e diffusione dei dati censuari. Per la geografia computazionale, ASIR rappresenta un passo avanti nella traduzione dei concetti di similarità geometrica e *conflation* in un modello operativo, basato su funzioni di aggregazione fuzzy. Ciò consente di applicare in modo sistematico tecniche finora usate in contesti sperimentali a problemi concreti di gestione dei dati territoriali. Infine, i risultati ottenuti sul caso di Cassino dimostrano la robustezza dell'approccio e ne aprono la strada a sviluppi futuri, sia estendendo la sperimentazione ad altri contesti urbani, sia applicandolo a fonti diverse (ambientali, sociali, infrastrutturali).

## Riferimenti bibliografici

- Adams, B., Li, L., Raubal, M., and Goodchild, M. F. (2010). A general framework for conflation. Technical report.
- Alt, H. and Godau, M. (1995). Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(1–2):75–91.
- Balasubramani, B. S. and Cruz, I. F. (2018). Spatial data integration. In Sakr, S. and Zomaya, A., editors, *Encyclopedia of Big Data Technologies*, pages 1–8. Springer International Publishing, Cham.
- Brovelli, M. A., Minghini, M., Molinari, M. E., and Zamboni, G. (2016). Positional accuracy assessment of the openstreetmap buildings layer through automatic homologous pairs detection. In *ISPRS Archives*.
- Coccagna, B. (2011). Libero accesso nelle politiche di open data: trasparenza, apertura e auto-organizzazione nel riutilizzo delle informazioni del settore pubblico. *Cyberspazio e diritto*, 12(2):129–160.
- Ďuračiová, R. (2023). An aggregated shape similarity index: A case study of comparing the footprints of openstreetmap and inspire buildings. *ISPRS International Journal of Geo-Information*, 12(12):495.
- Ďuračiová, R. and Igondová, M. (2017). Integration of spatial data representing buildings by determining the degree of similarity. *Czech Journal of Civil Engineering*, 3(1):29–35.
- Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014). Quality assessment for building footprints data on openstreetmap. *International Journal of Geographical Information Science*, 28:700–719.
- Franzini, M., Annovazzi-Lodi, L., and Casella, V. (2020). Assessment of the completeness of openstreetmap and google maps for the province of pavia (italy). In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management*, pages 270–277. SCITEPRESS.
- Hecht, R. (2013). Measuring completeness of building footprints in openstreetmap over space and time. *ISPRS International Journal of Geo-Information*, 2(4):1066–1091.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579.
- Lei, T. L. (2020). Geospatial data conflation: A formal approach based on optimization and relational databases. *International Journal of Geographical Information Science*, 34(11):2296–2334.
- Liu, L., Ding, X., Zhu, X., Fan, L., and Gong, J. (2020). An iterative approach based on contextual information for matching multi-scale polygonal object datasets. *Transactions in GIS*, 24(4):1047–1072.
- Min, D., Li, Z., and Cheng, X. (2007). Extended hausdorff distance for spatial objects in gis. *International Journal of Geographical Information Science*, 21(4):459–475.
- Nedas, K. A. and Egenhofer, M. J. (2003). Spatial similarity queries with logical operators. In Hadzilacos, T., Manolopoulos, Y., Roddick, J., and Theodoridis, Y., editors, *Advances in Spatial and Temporal Databases*, volume 2750 of *Lecture Notes in Computer Science*,

- pages 430–448. Springer, Berlin, Heidelberg.
- Pesaresi, M. and Politis, P. (2022). Ghs-built-h r2022a - ghs building height dataset.
- Ponti, B. (2007). Il patrimonio informativo pubblico come risorsa. i limiti del regime italiano di riutilizzo dei dati delle pubbliche amministrazioni. *Diritto pubblico*, (3):991–1014.
- Salvucci, G. and Salvati, L. (2021). Official statistics, building censuses, and openstreetmap completeness in italy. *ISPRS International Journal of Geo-Information*, 11(1):29.
- Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology. *Biologiske Skrifter*, 5:1–34.
- Tanimoto, T. T. (1958). *Elementary Mathematical Theory of Classification and Prediction*.
- Thompson, M. M. (1988). *Maps for America: Cartographic Products of the U.S. Geological Survey and Others*. U.S. Government Printing Office, Washington, DC.
- Yan, H. (2009). Concepts and classification of spatial similarity relations in multi-scale map spaces. In *Proceedings of SPIE*, page 74924T.
- Yan, H. (2010). Fundamental theories of spatial similarity relations in multi-scale map spaces. *Chinese Geographical Science*, 20:18–22.
- Yan, H. (2014). *Theory of Spatial Similarity Relations and Its Applications in Automated Map Generalization*. PhD thesis, University of Waterloo.
- Yu, Y., Jiang, H., Zhang, X., and Chen, Y. (2022). Identifying irregular potatoes using hausdorff distance and intersection over union. *Sensors*, 22(15):5740.
- Yuan, S. and Tao, C. (1999). Development of conflation components. In *Proceedings of Geoinformatics 99*, pages 1–13.