

NOTA METODOLOGICA

La produzione dei dati per sezioni di censimento del Censimento Permanente al 31.12.2021.

Il nuovo censimento permanente della popolazione e delle abitazioni si basa sull'integrazione di dati di fonte amministrativa e di dati acquisiti attraverso le rilevazioni che ogni anno coinvolgono un campione rappresentativo di comuni e di famiglie.

I dati da fonte amministrativa insieme a quelli provenienti da rilevazioni campionarie condotte dall'Istat costituiscono un flusso continuo di informazioni che vanno a definire e aggiornare i Registri statistici che supportano la produzione censuaria: il Registro Base degli Individui (RBI), il Registro Tematico del Lavoro (RTL), il Registro Statistico di Base dei Luoghi (RSBL). Questi registri contengono dati statistici e informazioni territoriali per tutte le unità censuarie: individui, famiglie e alloggi. I primi due registri (RBI e RTL) includono dati sulle principali caratteristiche strutturali e socioeconomiche della popolazione; il Registro RSBL contiene informazioni sui luoghi ed è articolato in quattro componenti: suddivisioni amministrative e statistiche, sezioni di censimento, indirizzi e, infine, edifici e abitazioni.

Al fine di integrare le informazioni di fonte amministrativa, l'Istat conduce ogni anno due indagini campionarie: una rilevazione di tipo areale per valutare la qualità dei dati impiegati per il conteggio della popolazione a livello comunale e una rilevazione da lista anagrafica per raccogliere le variabili tradizionalmente osservate in occasione del censimento e, di conseguenza, costruire l'insieme di dati necessari per produrre gli output ai vari livelli classificatori e territoriali.

Per la determinazione dei dati sub-comunali è necessario geo-codificare le unità statistiche alle sezioni di censimento. Questo tipo di operazione in passato avveniva durante la rilevazione, in modo contestuale alla raccolta dei dati. Con la nuova strategia del censimento permanente la geo-codifica delle unità statistiche proviene dall'integrazione tra RBI e RSBL. La lavorazione riferita ai dati attualmente diffusi si è basata principalmente sulla disponibilità di informazioni sugli indirizzi degli individui della popolazione censita, sull'associazione tra gli indirizzi e le sezioni di censimento 2011 e sull'adozione di opportuni metodi statistici per recuperare le informazioni delle unità che, per vari motivi, non è stato possibile geo-codificare. Pertanto, l'accuratezza dell'intero processo di geo-codifica della popolazione 2021 alle sezioni di censimento 2011 è dipesa principalmente dalla qualità delle informazioni presenti negli archivi amministrativi che alimentano i registri.

L'operazione di geo-codifica alle sezioni di censimento, preliminare alla produzione dei dati rilasciati, è avvenuta in tre fasi successive (Figura 1, pag. 3):

1. i processi di associazione degli indirizzi alle sezioni di censimento in RSBL e di associazione degli indirizzi agli individui in RBI;

2. la fase di linkage tra RBI e RSBL tramite gli indirizzi;
3. la procedura di assegnazione delle unità non geo-codificate alle sezioni di censimento.

Nella prima fase, nell'ambito del registro RSBL, gli indirizzi sono stati associati in modo univoco alle sezioni di censimento disponibili nel registro (Basi Territoriali 2011), utilizzando sia le coordinate geografiche sia le informazioni relative agli indirizzi presenti negli archivi amministrativi (Archivio Nazionale dei Numeri Civici delle Strade Urbane, Liste Anagrafiche Comunali, Anagrafe Tributaria). Oltre gli archivi amministrativi predetti sono stati utilizzati anche gli indirizzi del Censimento della Popolazione del 2011 e le informazioni provenienti dall'indagine effettuata presso i Comuni tra il 2020 e il 2021. Il processo di geo-codifica dell'indirizzo alla sezione di censimento prevede un algoritmo che verifica il livello di concordanza e di divergenza delle informazioni, sia geografiche che non, attribuendo (secondo una gerarchia delle fonti e criteri predefiniti) un'associazione univoca tra indirizzo e sezione di censimento. Sempre nella prima fase, si è proceduto infine ad associare gli indirizzi alla popolazione del RBI individuata come dimorante abitualmente al censimento del 2021.

La seconda fase è consistita in una prima integrazione dei registri RBI e RSBL, tramite l'impiego di opportune tecniche di *record-linkage*, per collocare individui e famiglie sul territorio. Gli individui presenti in RBI sono stati geo-codificati tramite una chiave di aggancio univoca riferita all'indirizzo (prevalentemente di residenza) e alla sezione di censimento 2011 associata in RSBL. A causa di alcuni problemi strutturali connessi alla qualità e all'aggiornamento delle informazioni contenute nelle fonti impiegate per la costruzione dei registri RBI e RSBL, non è stato possibile collocare sul territorio in modo completo tutta la popolazione con la sola operazione di record-linkage. Alla fine di questa attività una parte residuale è rimasta non geo-codificata (circa il 2,5% della popolazione censuaria relativa al 2021).

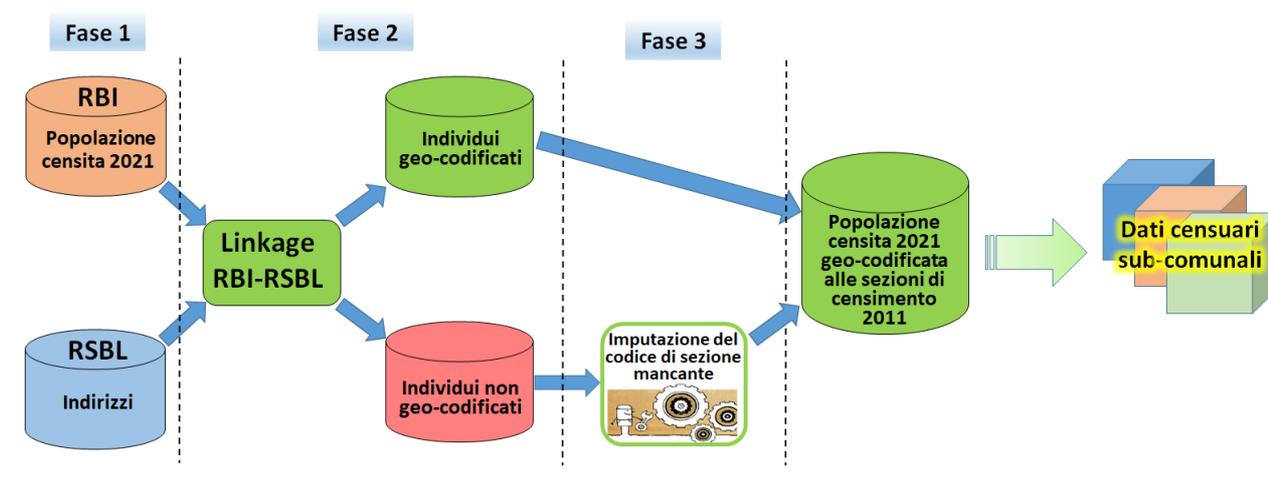
Per completare la geo-codifica di tutte le unità si è resa necessaria, quindi, una terza fase di processo in cui, tramite l'impiego di una procedura sviluppata dall'Istat, è stato possibile assegnare il codice di sezione alle unità non geo-codificate. Tale operazione, implementata solo a fini statistici, ha utilizzato procedure deterministiche e probabilistiche ordinate in modo gerarchico (Carbonetti *et al.*, 2023a; Carbonetti *et al.*, 2023b). L'esito dell'intera procedura di recupero del codice di sezione, ove mancante, ha permesso di completare la collocazione di tutta la popolazione censuaria.

Unitamente ai dati vengono rilasciati anche alcuni indicatori relativi all'esito delle diverse fasi del processo di geo-codifica della popolazione al territorio. Tali indicatori, calcolati a livello comunale a partire dal dato di popolazione residente al 31.12.2021, sono i seguenti:

- quota di popolazione con sezione proveniente dai registri (fase 2);
- quota di popolazione con sezione attribuita mediante procedure deterministiche (fase 3);
- quota di popolazione con sezione attribuita mediante procedura probabilistica (fase 3).

Una volta completata la geo-codifica di tutte le unità censite alle sezioni di censimento 2011 si è proceduto alla produzione, validazione e rilascio dei dati sub-comunali¹.

Figura 1 – Fasi dell’operazione di geo-codifica della popolazione censita 2021 alle sezioni di censimento 2011.



Riferimenti bibliografici

Fardelli D., Orsini E. and Pagano A. (2021). “The address component of the Statistical Base Register of Territorial Entities”. Book of Short Papers SIS 2021, pp. 1206-1211. Pearson.

<https://it.pearson.com/content/dam/region-core/italy/pearson-italy/pdf/Docenti/Universit%C3%A0/pearson-sis-book-2021-parte-2.pdf>

Fortini M., Tuoto T. (2020). “Probabilistic record linkage with less than three matching variables”. Book of Short Papers – SIS 2020, pp. 3-8. Pearson.

<https://it.pearson.com/content/dam/region-core/italy/pearson-italy/pdf/Docenti/Universit%C3%A0/Pearson-SIS-2020-atti-convegno.pdf>

Carbonetti G., Daddi S., De Matteis G., Di Zio M., Fardelli D., Ferrara R., Lipizzi F., Orsini E. (2023a), “New perspectives for the quality of sub-municipal data with the Italian permanent population and housing census”. Book of Short Papers – ASA Conference 2022: Data-Driven Decision Making, Genova 12-14 Settembre 2022. Firenze University Press.

¹ Per questo rilascio dei dati l’assegnazione della popolazione alla sezione non ha previsto un controllo sulla componente abitazioni/edifici per valutare la disponibilità di abitazioni per la popolazione assegnata alle sezioni. Per rilasci futuri basati sulle Basi territoriali 2021, oltre alla geo-codifica degli individui a queste nuove basi, si terrà conto anche della componente edifici e abitazioni del RSBL.

Carbonetti G., De Matteis G., Di Zio M., Fardelli D., Ferrara R., Lipizzi F. (2023b), “Enumeration area imputation methods for producing sub-municipal data in the Italian permanent population and housing census”. *Statistical Journal of the IAOS*, vol. 39, no. 1, pp. 123-136, 2023. IOS Press.

DOI: 10.3233/SJI-220113

<https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji220113>