



## Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza al 31.12.2023

Da ottobre 2018 l'Istat ha avviato il Censimento permanente della popolazione e delle abitazioni che ha sostituito il tradizionale Censimento della popolazione decennale. Questo nuovo approccio si basa sull'integrazione delle informazioni disponibili dalle fonti amministrative con i dati raccolti attraverso le indagini campionarie svolte a rotazione su tutti i comuni italiani.

Per il conteggio della popolazione nel 2023 è stato possibile beneficiare dei progressi raggiunti nel corso degli ultimi anni in termini di qualità e tempestività dai Registri a supporto della produzione statistica ufficiale. In questo contesto il perno del Censimento permanente della popolazione è rappresentato dal Registro di Base degli Individui, delle famiglie e delle convivenze (RBI).

Il RBI, insieme ai registri tematici (ad esempio quelli sull'occupazione e sull'istruzione) e alla regolare acquisizione delle fonti amministrative, trattate ed utilizzate a fini statistici, ha permesso ormai da quasi un decennio di integrare i dati di fonte anagrafica comunale con le informazioni provenienti da altri archivi di INPS, MIUR, Catasto Immobiliare, Casellario dei pensionati, ecc., così come previsto dalla Legge 205/2017.

Il RBI è un ambiente informativo interno all'Istituto che, progettato per supportare i processi produttivi statistici, rappresenta l'infrastruttura fondamentale per la produzione delle statistiche ufficiali sulla popolazione e il punto di riferimento per l'estrazione dei campioni necessari alle indagini previste dal Censimento permanente (come specificato nella legge censuaria e nel Piano Generale di Censimento) e a tutte le indagini campionarie sulle famiglie.

In linea con l'impianto metodologico del Censimento permanente, e nel rispetto delle normative sulla protezione dei dati personali, il RBI e tutte le fonti amministrative non anagrafiche disponibili presso l'Istat contengono dati anonimi "statistici". Questi dati, derivanti da un processo di elaborazione e validazione statistica, sono limitati a un numero di variabili funzionali alla rappresentazione delle principali caratteristiche strutturali della popolazione.

Il RBI viene "consolidato" annualmente alla data del 31 dicembre di ogni anno. Questo processo di consolidamento si basa sull'integrazione dei flussi individuali della dinamica demografica (nascite, decessi, trasferimenti di residenza sia tra comuni che con l'estero). L'adozione del modello di contabilità demografica permette di sfruttare appieno le potenzialità della base dati micro (combinazione di flussi e stock) per produrre indicatori più accurati e innovativi sulla dinamica demografica, tenendo conto della sequenza degli eventi demografici vissuti dagli individui.

I principali output applicati al RBI "consolidato" comprendono il bilancio demografico della popolazione residente a livello comunale e la struttura della popolazione residente, suddivisa per genere, età, stato civile e cittadinanza. La produzione tempestiva di queste informazioni consente di anticipare i risultati definitivi annuali del Censimento permanente, garantendo la continuità della produzione statistica ufficiale. Questo processo soddisfa sia le stringenti tempistiche richieste dagli utilizzatori istituzionali dei dati di popolazione, sia le prescrizioni del regolamento europeo sulle statistiche demografiche. Questi output vengono successivamente rielaborati e diffusi come definitivi, una volta effettuata la correzione del RBI sulla base degli esiti del Censimento permanente.

Come per tutto il primo ciclo, il conteggio del Censimento permanente del 2023 ha l'obiettivo di correggere gli errori di sovra e sotto-copertura del RBI. In particolare, vengono individuate le persone registrate come residenti ma non presenti negli archivi amministrativi (sovra copertura) e quelle identificate nei dati amministrativi come abitualmente dimoranti ma non risultanti tali nel Registro (sotto copertura). Questa correzione, opera a livello micro, attraverso la riclassificazione dei record individuali nel Registro determinandone lo status di abitualmente dimoranti o meno sulla base dei segnali di vita amministrativi. Si consolida così un'innovazione metodologica significativa che garantisce la corrispondenza precisa, in termini di "teste", tra il conteggio e i record di individui abitualmente dimoranti nel RBI.

Questo sistema garantisce, e continuerà a garantire in misura crescente nei prossimi anni, una piena integrazione con gli output delle statistiche demografiche, contribuendo così a una maggiore stabilità e coerenza dell'intero sistema dei dati relativi alla popolazione. Il processo censuario farà sempre più uso delle



informazioni provenienti dagli archivi amministrativi per stimare il conteggio di popolazione e, ove possibile, alcune variabili tematiche previste dal Regolamento europeo sui Censimenti della popolazione e delle abitazioni.

## I segnali di vita e le fonti utilizzate per il conteggio 2023 della popolazione

Per il conteggio della popolazione del 2023, il Servizio Censimento della popolazione ha creato un repository che integra le informazioni provenienti dalle fonti amministrative con quelle anagrafiche organizzate nel Sistema Integrato di Microdati (SIM). Il SIM presenta una base dati relazionale di dati amministrativi, progettata per garantire l'anonimato degli individui, e sostenere i processi di produzione statistica, sia per le statistiche sociali che economiche. Un elemento chiave di questo sistema è l'assegnazione di un codice ID univoco e costante, che consente di identificare ogni individuo e unità economica nei diversi archivi e di creare relazioni tra le fonti, garantendo al contempo il trattamento di dati privi di identificativi diretti.

Grazie alle indicazioni degli esperti di fonti amministrative dell'Istat che utilizzano gli archivi per la produzione delle statistiche sulla popolazione, il Servizio Censimento permanente della popolazione ha selezionato le fonti rilevanti per il conteggio della popolazione dimorante abitualmente. Inoltre, un gruppo di esperti tematici e metodologi, ha definito un ordine gerarchico tra le fonti, fondamentale nel processo di integrazione che costituisce la base di riferimento per il conteggio della popolazione. Questa metodologia ha permesso poi di rilevare i cosiddetti "segnali di vita" degli individui, indicativi della loro "dimora abituale in Italia". Per il conteggio di popolazione del 2023, il database ha integrato le informazioni anagrafiche individuali del Registro di Base degli Individui (RBI) con quelle del Registro Tematico del Lavoro (RTL) dell'Istat, dagli archivi dell'istruzione, dalle dichiarazioni fiscali, dagli archivi della previdenza sociale e dal catasto immobiliare dell'Agenzia delle Entrate.

Tutte queste informazioni sono state raccolte con un riferimento temporale compreso tra il 1° gennaio 2022 e il 31 dicembre 2023, garantendo una fotografia completa e aggiornata della popolazione residente.

Nell'ambito delle fonti utilizzate, il Prospetto 1 illustra, per gli anni 2022 e 2023, la denominazione degli archivi, l'ente titolare fornitore, il numero di record trattati per ciascun anno e la classificazione dei dati in base al tipo di segnale di vita. Quest'ultima, secondo la logica gerarchica definita dagli esperti del Censimento permanente della popolazione, permette di attribuire agli individui la condizione di dimora abituale in Italia. Complessivamente, i record trattati nel repository di integrazione per ciascun anno ammontano a centinaia di milioni. Questo numero include sia le molteplici occorrenze che possono riferirsi allo stesso individuo in momenti diversi nello stesso archivio, sia la possibilità che lo stesso individuo compaia contemporaneamente in più archivi. Nel costruire la base dati integrata delle fonti amministrative e anagrafiche, si è tenuto conto sia della periodicità dei dati sia della loro rilevanza rispetto alla definizione di dimora abituale stabilita dal regolamento europeo (regolamento CE n. 1260/2013). Dal Prospetto 1 emerge che le fonti assicurative dell'INPS, relative agli individui che partecipano al mercato del lavoro, in particolare per il settore privato, sono ampiamente disponibili per entrambi gli anni (2022 e 2023). Diversamente, alcune fonti risultano parzialmente o completamente mancanti per uno dei due anni. Le fonti del MIUR, riguardanti gli Archivi del personale, e quelle previdenziali relative al Casellario dei pensionati e assistenziali, hanno un riferimento data che copre solo il 2022.

Si registrano ritardi rispetto alla data di riferimento per alcune fonti delle dichiarazioni fiscali, in particolare per la Banca dati reddituale e le informazioni derivanti dai modelli UNICO/730. Tali ritardi, di circa 20 mesi, sono attribuibili alla natura delle fonti fiscali, che si basano sulle dichiarazioni dei redditi relative all'anno precedente rispetto alla data di riferimento.

Negli ultimi anni il sistema integrato dei dati amministrativi e anagrafici ha subito una notevole evoluzione. Oltre al crescente utilizzo di registri tematici, come l'RBI e l'RTL, sono stati realizzati significativi progressi tecnici e informatici. Questi miglioramenti hanno permesso di ottimizzare l'utilizzo delle informazioni disponibili, riducendo i tempi di elaborazione e aumentando l'efficienza operativa. Parallelamente, sono state intraprese importanti attività di ingegnerizzazione dei processi e dei prodotti, tra cui l'aggiornamento dei codici di provincia e comune. Tali interventi hanno migliorato la qualità della localizzazione geografica dei segnali provenienti dalle fonti amministrative.



**PROSPETTO 1 - ARCHIVI AMMINISTRATIVI E REGISTRI TEMATICI CHE FORNISCONO SEGNALI DI VITA DIRETTI E INDIRETTI RISPETTO ALLA DIMORA ABITUALE IN ITALIA AL 31.12.2023**

SEGNALI DI VITA PER ORDINE GERARCHICO	ENTE TITOLARE	DENOMINAZIONE ARCHIVIO	N. record	
			2022	2023
1) Segnali diretti di lavoro	INPS	UNIEMens	187.749.934	190.175.961
		DMAG (dichiarazione sulla manodopera agricola)	8.209.575	8.153.875
		Lavoratori autonomi in agricoltura	435.085	426.343
		Archivio dei parasubordinati collaboratori	8.233.876	8.479.401
		Rapporti di lavoro domestico	1.186.231	1.124.500
		Gestione Dipendenti Pubblici - Posizioni degli assicurati iscritti	43.047.448	
		Lavoro occasionale (Libretto Famiglia e Prestazioni Occasionali)	1.423.775	1.539.669
		Archivio dei lavoratori autonomi: artigiani e commercianti	3.617.948	3.500.322
	Anagrafe Tributaria delle persone giuridiche e delle persone fisiche con partita IVA	6.758.764	5.652.721	
	MIUR	Archivio del personale universitario	516.793	
		Archivio del personale delle scuole statali	1.186.516	
	INAIL	Lavoratori interinali	1.770.348	1.633.766
MEF	NOIPA - Anagrafica (Cedolini Stipendiali dipendenti PA)	23.620.167	23.656.191	
ISTAT	Registro Tematico del Lavoro (RTL)	27.274.328	27.702.352	
2) Segnali diretti di studio	MIUR	Anagrafe degli studenti delle scuole	17.064.012	16.775.850
		Archivio degli iscritti e delle iscrizioni universitarie	2.026.203	
		Archivio delle lauree e dei laureati	528.403	532.122
		Anagrafe Nazionale degli Studenti dei corsi Post laurea - Iscritti e iscrizioni	185.679	
		Anagrafe Nazionale degli Studenti dei corsi Post laurea - Chiusura carriera	83.407	83.225
3) Segnali diretti di assistenza sociale e contratti di affitto ad uso abitativo	INPS	Beneficiari CIG a pagamento diretto	446.432	214.951
		Trattamenti non pensionistici	5.252.590	
		Casellario dei Pensionati e dei trattamenti pensionistici	22.771.903	
		Possessori della Social Card	369387	
		Reddito di cittadinanza (RDC)	1.690.536	1.367.845
		Beneficiari dell'indennità una tantum di 150/200 euro	2.801.067	408.545
	Assegno unico e universale(AUU)	85.462.887	107.620.676	
	MEF	Banca dati statistica reddituale	42.026.375	
Agenzia delle Entrate - Contratti di Locazione		4.393.391	4.406.851	
4) Segnali indiretti	MEF	Agenzia delle Entrate - Dichiarazioni Modello Redditi Persone Fisiche	13.646.487	
		Agenzia delle Entrate Dichiarazioni 730	36889637	
		Agenzia delle Entrate - Certificazione Unica	51962473	
		Agenzia delle Entrate - Catasto delle Unità Immobiliari	116.424.595	
	AQUIRENTE UNICO	Registro Centrale Ufficiale (RCU) - Settore Elettrico Consumi Consolidati	25.800.608	28.063.837
		Registro Centrale Ufficiale (RCU) - Settore Gas Consumi Consolidati	21892430	21.883.169
	ACI	ACI - Parco veicoli circolanti	48.691.246	49.343.042
Fonti anagrafiche e altre fonti utilizzate	ISTAT	Registro Base degli INDIVIDUI (RBI)	103208269	104.784.117
	MIN. INTERNO	Permessi di soggiorno	4.175.975	3.746.408
		Acquisizione e reiezione della cittadinanza italiana	88.185	88.935
	MAE	MAE - Ministero Affari Esteri - Archivio Italiani all'Estero	9.660.960	10.099.646
MEF	Agenzia delle Entrate - Anagrafe delle Persone Fisiche (AT)	104.621.798	106.379.848	

Fonte: Istat 2024



I ricercatori tematici del Servizio del Censimento permanente della popolazione hanno inoltre consolidato una definizione dei segnali di vita - noti nella letteratura internazionale come “*Signs of administrative life*”.

*I “segnali di vita amministrativi” si riferiscono ad attività svolte dagli individui desumibili dagli archivi amministrativi. Queste attività permettono di identificare chiaramente un periodo di tempo durevole (ad esempio, un anno) e un luogo (un Comune) in cui si realizzano. Svolgere un lavoro autonomo o lavorare per un’impresa, essere un dipendente pubblico, avere un regolare contratto d’affitto annuale per una abitazione, frequentare una scuola o l’università sono esempi di segnali di vita amministrativi diretti. Invece, si definiscono segnali di vita indiretti quelle situazioni, sempre desumibili dagli archivi amministrativi, che identificano uno status o una condizione, ad esempio essere percettori di reddito di cittadinanza o di una pensione di vecchiaia, oppure essere familiari a carico per i quali il dichiarante del reddito indica di avere a suo carico il coniuge, i figli o altro parente.*

La definizione adottata consente di delineare anche una classificazione gerarchica dei segnali, come indicato nella prima colonna del Prospetto 1. I segnali relativi al lavoro, allo studio, ai contratti di affitto, alle dichiarazioni contenute nella Banca dati reddituale<sup>1</sup> e ai sussidi socio-assistenziali dell’INPS sono considerati segnali di vita diretti, in quanto strettamente collegati alla dimora abituale in Italia.

Particolarmente rilevanti per l’individuazione della popolazione abitualmente dimorante in Italia sono gli archivi dell’INPS relativi al lavoro e quelli del MIUR riguardanti la frequenza di corsi scolastici e universitari.

Il Prospetto 2 evidenzia come queste fonti offrano un livello informativo dettagliato, che comprende non solo la durata dell’attività svolta, ma anche la sua localizzazione (Comune e indirizzo) e alcune caratteristiche specifiche (ad esempio la tipologia di contratto di lavoro o il corso di studi). Tali informazioni rappresentano elementi fondamentali per valutare la solidità del segnale di vita sul territorio.

#### PROSPETTO 2 - ATTRIBUTI DEGLI ARCHIVI RELATIVI AD ATTIVITÀ DI LAVORO E/O STUDIO

Nome dell’attributo del Database integrato	Descrizione
Codice individuo	Identificativo univoco di ciascun individuo che permette l’integrazione tra le varie fonti e anni differenti
Codici Provincia e Comune	Luogo di lavoro o di studio a seconda del tipo di fonte da cui proviene l’informazione
Durata dell’attività (Presenze mensili)	Dettaglio dei mesi, da gennaio a dicembre degli anni considerati, a cui l’informazione si riferisce
Lavoro - Qualifica	Operaio, impiegato, dirigente
Lavoro - Orario di lavoro	Tempo pieno, parziale orizzontale
Lavoro - Durata del contratto	Indeterminato, determinato, stagionale
Studio - Tipologia di corso	Scuola dell’obbligo, Università
Unità economica di riferimento	Unità locale o scuola

Fonte: Istat, 2024

Invece, tutte le fonti fiscali (come i modelli UNICO, modelli 730, ecc.), il possesso di automobili registrate nell’archivio ACI e il possesso di immobili presenti nell’archivio catastale sono considerate fonti di segnali indiretti rispetto alla dimora abituale. Ad esempio, dal “Quadro Familiari a carico” dei modelli UNICO Persone Fisiche e 730, è possibile ricavare le relazioni principali tra “coniugi” e tra il genitore dichiarante e i “figli”. Inoltre, questi modelli fiscali contengono informazioni sugli “altri familiari” a carico come il coniuge separato legalmente ed effettivamente; i discendenti dei figli; i genitori, i generi e le nuore; i suoceri; i fratelli e le sorelle; e i nonni purché vivano con il dichiarante o ricevano da lui assegni alimentari non risultanti da provvedimenti dell’Autorità giudiziaria.

Poiché tali segnali di presenza derivano “indirettamente” dalla dichiarazione di un percettore di reddito, essi sono stati classificati come segnali indiretti. Allo stesso modo, il possesso di un’automobile o di un’unità immobiliare non è considerato un segnale diretto ed è anch’esso incluso nella categoria dei segnali indiretti.

<sup>1</sup>Si tratta delle dichiarazioni contenute nella Banca dati reddituale. Attraverso apposite campagne ordinarie, l’INPS richiede ai soggetti beneficiari di prestazioni previdenziali e/o assistenziali collegate al reddito (pensionati, disoccupati beneficiari di cassa integrazione, ecc.) l’annuale dichiarazione dei redditi (modello RED) che incidono sulle prestazioni in godimento.



Ulteriori segnali indiretti derivano dal RBI e si riferiscono alla composizione del nucleo familiare registrato in anagrafe. In questo caso, le relazioni principali individuate sono quelle tra l'intestatario della famiglia e il "coniuge" e tra intestatario e i "figli".

La base dati della popolazione residente utilizzata per verificare la coerenza con i segnali di vita in Italia è costituita dagli individui identificati come residenti nel RBI attraverso un processo prodromico al registro stesso che considera sia i flussi in ingresso e in uscita della popolazione (MIDEA -Micro-DEmographic Accounting) che la definizione della popolazione del RBI eleggibile ad essere residente in un Comune (ANVIS - Anagrafe Virtuale Statistica). La base dati ANVIS viene quindi consolidata ogni anno mediante un aggiornamento a livello micro, integrando i flussi della dinamica demografica come iscrizioni per nascita, cancellazioni per decesso, iscrizioni/cancellazioni da e per l'estero, e iscrizioni/cancellazioni da e per altri Comuni italiani. Questi flussi si riferiscono al periodo compreso tra il 9 ottobre 2011 e il 31 dicembre 2023.

Le informazioni anagrafiche sono state ottenute attraverso l'integrazione dei dati contenuti nei diversi archivi inclusi nel SIM. In particolare, i dati utilizzati per ricostruire il profilo demografico degli individui hanno riguardato:

- Data di nascita;
- Genere;
- Cittadinanza;
- Paese di nascita.

Qualora la cittadinanza non fosse disponibile in uno degli archivi considerati (RBI e Permessi di Soggiorno) si è fatto riferimento all'informazione relativa al paese di nascita. In caso di incongruenze tra le informazioni anagrafiche presenti nelle diverse fonti per uno stesso individuo, è stata applicata una scala gerarchica delle fonti per selezionare i dati più affidabili. L'Anagrafe Tributaria, insieme all' RBI, rappresenta la fonte più completa per le variabili anagrafiche relative a data di nascita, genere e paese di nascita. Inoltre, si sottolinea che ai fini dell'analisi della localizzazione effettiva degli individui in termini di dimora abituale, risulta particolarmente utile la presenza, in questa fonte, della variabile relativa al Comune del domicilio fiscale.

### **Il processo di integrazione del RBI e delle altre fonti amministrative e i profili di continuità per la dimora abituale in Italia**

Il Prospetto 3 illustra il flusso di lavoro, e le fasi principali del processo di integrazione delle fonti amministrative con RBI. Questo percorso, che parte dallo scarico delle fonti dal SIM fino alla determinazione della sotto e sovra copertura del RBI segue una struttura metodologica articolata. Dopo lo scarico delle fonti dal SIM (*Step 1*), i dati derivanti dagli archivi amministrativi selezionati sono stati riorganizzati per consentire l'analisi della popolazione dimorante abitualmente in Italia.

Il processo di integrazione prevede il trattamento di informazioni provenienti da oltre 40 archivi amministrativi, ciascuno relativo a più annualità. Questi archivi contengono i dati fondamentali sui segnali di vita degli individui e, per ogni evento amministrativo registrato, riportano informazioni sulla localizzazione dell'evento stesso attraverso i codici di provincia e comune. Oltre alla localizzazione degli eventi, le stesse variabili sono utilizzate per ricavare ulteriori informazioni, quali la residenza, il luogo di nascita o il comune di provenienza.

Tuttavia, il processo di integrazione deve affrontare alcune criticità:

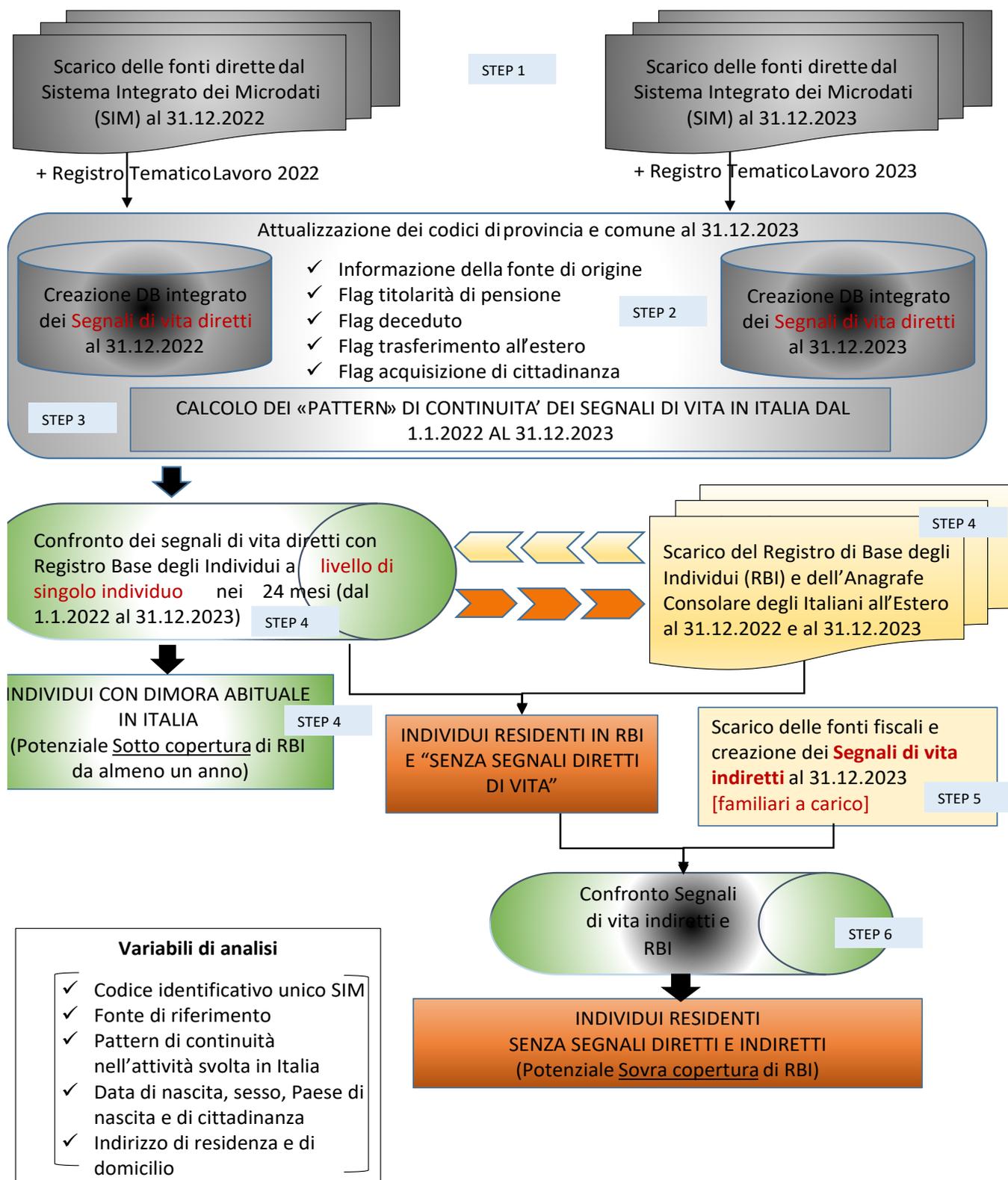
- valori validi ma disallineati temporalmente, o non aggiornati alla data di riferimento;
- valori mancanti, totali o parziali.

Il disallineamento temporale si verifica in particolare quando gli archivi vengono acquisiti in momenti diversi.

Ciò comporta, ad esempio, che uno stesso comune possa essere associato a codici territoriali differenti in base alle variazioni amministrative intercorse. Inoltre, il calcolo degli output derivanti dall'integrazione delle fonti amministrative richiede l'utilizzo di dati relativi a due anni consecutivi, rendendo il processo di produzione ancora più sensibile a tali problematiche. Il problema dei valori mancanti risulta particolarmente rilevante per alcune fonti che forniscono segnali diretti di presenza sul territorio come ad esempio le posizioni degli assicurati iscritti all'ex INPDAP.



**PROSPETTO 3 - FLUSSO DI LAVORO DEI SEGNALI DI VITA PER LA DETERMINAZIONE DEL CONTEGGIO DI POPOLAZIONE DEL CENSIMENTO PERMANENTE DELLA POPOLAZIONE AL 31.12.2023**



Fonte: Istat, 2024



In altri casi, le informazioni sono disponibili solo parzialmente, limitando l'utilizzabilità delle fonti stesse. Per affrontare queste criticità il Servizio Censimento svolge una serie di analisi e operazioni specifiche, tra cui:

- analisi territoriali, applicate al RBI e al RTL;
- definizione della localizzazione prevalente per i segnali di lavoro e studio;
- aggregazioni longitudinali basate su variabili geografiche;
- confronti temporali su diversi archivi;
- comparazioni con informazioni provenienti da altre fonti.

Per gestire i dati mancanti e garantire una coerenza temporale, è stata definita una procedura di imputazione e allineamento temporale delle variabili di localizzazione. Questa procedura, implementata nello Step 2 del processo di integrazione, utilizza:

- informazioni interne agli archivi amministrativi.
- fonti ausiliarie esterne, per colmare le lacune o verificare la coerenza dei dati.

Nella costruzione del prodotto integrato dei dati amministrativi, descritto nel Prospetto 3, i codici di provincia e comune di tutte le fonti amministrative utilizzate come input sono stati:

- Imputati (nei casi di valori mancanti).
- Attualizzati rispetto alle variazioni amministrative e territoriali.

L'obiettivo è stato quello di rendere le informazioni consistenti rispetto a un unico istante di riferimento, fissato al 31 dicembre 2023. Per l'imputazione dei valori mancanti, quando possibile, sono stati sfruttati vincoli logici e relazionali per determinare in maniera univoca i valori da attribuire. Questo approccio ha permesso di garantire un'elevata qualità e coerenza delle informazioni integrate.

Dopo l'aggiornamento dei codici di provincia e comune, nello Step 2 del processo di integrazione si procede alla creazione della base dati integrata delle fonti di lavoro e studio dell'INPS e del MIUR oltre alle altre fonti che forniscono segnali di vita diretti. Queste includono: Casellario dei pensionati, Trattamenti non pensionistici, Banca dati reddituale, Contratti di locazione, Cassa Integrazione Guadagni, reddito di cittadinanza, di emergenza e sussidi e indennità Covid-19. Inoltre, vengono generati i *flag* informativi che identificano: eventuali decessi, titolarità di pensioni, trasferimenti all'estero (derivabili dalla Banca dati Reddituale e dal Casellario dei Pensionati), acquisizione di cittadinanza italiana.

Un aspetto cruciale per il conteggio della popolazione è stato il rispetto della definizione di dimora abituale sancita dal Regolamento UE 1260/2013. Secondo questa normativa, la "*usually resident population*" si riferisce a:

- Individui che hanno risieduto in un luogo per almeno 12 mesi prima della data di riferimento.
- Individui che, pur non avendo risieduto per 12 mesi, manifestano un'intenzione di stabilità per un periodo equivalente successivo alla data di riferimento.

Conformemente a queste indicazioni, per il conteggio della popolazione si è adottato come riferimento temporale il periodo compreso tra il 1° gennaio 2022 e il 31 dicembre 2023. Tale intervallo è stato utilizzato per identificare i segnali di vita amministrativi e rappresenta la data di riferimento del Censimento permanente della popolazione 2023. Questo approccio garantisce una coerenza metodologica con le direttive europee e consente di fornire una rappresentazione accurata della popolazione dimorante abitualmente sul territorio nazionale.

L'osservazione longitudinale dei segnali diretti su un periodo di due anni permette di identificare profili specifici di presenza degli individui sul territorio (o pattern di continuità nella presenza) un'attività descritta nello Step 3 del Prospetto 3. Questi profili consentono, in alcuni casi, di determinare chiaramente la dimora abituale in Italia, mentre in altri i "segnali di vita" risultano di bassa intensità o associati a categorie come i lavoratori stagionali, che non rientrano nella definizione di popolazione dimorante abitualmente.

Per comprendere meglio la classificazione dei segnali di lavoro e studio rispetto alla dimora abituale, il Prospetto 4 offre una rappresentazione utile dei segnali diretti di presenza in Italia.

Ad esempio, i profili in blu scuro da 1 a 4 accumulano almeno 12 mesi di segnali di presenza dal 2022 al 2023, anche se non continui, come nei casi 3 e 4 dello schema questi profili sono comunque considerati validi per determinare la dimora abituale. Invece, i segnali classificati dal profilo 5 in poi presentano invece discontinuità o non raggiungono i 12 mesi complessivi di segnali diretti di lavoro o studio nei 24 mesi di



osservazione. I profili 5 e 6 mostrano una debole intensità di presenza, rendendoli meno rilevanti per l'identificazione della dimora abituale; i profili 7-11, invece, non sono utili per determinare la dimora abituale, ma potrebbero evolversi nel tempo. Ad esempio, i profili 10 e 11 potrebbero indicare, nei censimenti successivi, segnali forti di presenza, simili ai casi 2 e 4.

Questa classificazione rappresenta uno strumento essenziale per distinguere tra presenze stabili e temporanee sul territorio, garantendo un'analisi coerente con la definizione di dimora abituale prevista dal Regolamento UE.

**PROSPETTO 4 - CONFIGURAZIONE E CLASSIFICAZIONE DEI SEGNALI DIRETTI DI STUDIO E DI LAVORO SECONDO I PROFILI DI CONTINUITÀ DELLA DIMORA ABITUALE IN ITALIA DAL 1.1.2022 AL 31.12.2023**

Finestra temporale di osservazione (24 mesi)																								Profili del segnale di lavoro e studio		
G	F	M	A	M	G	L	A	S	O	N	D	G	F	M	A	M	G	L	A	S	O	N	D			
																									1	Continuo nel 2022-2023
																									2	Continuo, prevalentemente nel 2023
																									3	Continuo, oltre dicembre 2022
																									4	Con pause ma almeno un anno
																									5	Stagionale
																									6	Interrotto sia nel 2022 che nel 2023
																									7	Solo a dicembre 2022
																									8	Solo nei primi mesi del 2022
																									9	Interrotto e solo nel 2022
																									10	Solo negli ultimi mesi del 2023
																									11	Interrotto e solo nel 2023

Fonte: Istat, 2024

Ogni segnale viene associato a un individuo specifico e a una localizzazione territoriale determinata, come illustrato nel Prospetto 5.

Si riporta di seguito un esempio pratico: nel caso di un individuo con codice identificativo "0000018", se nel periodo considerato si rileva un record in un archivio di lavoro e un altro record in una fonte relativa allo studio, entrambi i segnali localizzati nel comune di Agliè, si otterrà un unico segnale aggregato per quel comune. Tuttavia, questo segnale sarà contrassegnato da:

- un attributo che consente di tracciare l'individuo attraverso entrambi gli archivi di origine.
- un attributo relativo alla durata della presenza, espressa in termini di attività lavorativa e di studio.

**PROSPETTO 5 - ESEMPIO DI UTILIZZO DEI SEGNALI DI VITA DIRETTI DI LAVORO E STUDIO LOCALIZZATI IN UN COMUNE**

Nome attributo	Codice territorio + codice individuo	Fonti di presenza del segnale nel periodo considerato	Attributi specifici	{mese1--mese24}
Descrizione	{Identificativo}	{Sequenza fonti: ogni posizione una fonte specifica; 1=Presenza nella fonte}	{Altre info a corredo}	{Presenza/Assenza mensile}
<i>Esempio</i>	<i>Individuo 18 di Agliè</i>	<i>UniEmens (pos.1)+Università (pos.9)</i>	<i>Tempo indeterminato</i>	<i>Presente tutti i mesi</i>
Dati di esempio	001-001-0000018	1000000010	0----1-----	11111111111111111111111111111111

Fonte: Istat, 2024

L'algoritmo che elabora i segnali diretti di lavoro e studio nel prodotto integrato delle fonti amministrative genera una stringa sintetica, rappresentata nella terza colonna del prospetto (in rosso). La stringa registra:

- un segnale di lavoro nella prima posizione,



- un segnale di studio nella posizione 9.

Nell'ultima colonna del prospetto, è riportata la classificazione del profilo di continuità. Per il caso specifico, il segnale diretto di presenza nella fonte di lavoro risulta presente in tutti i mesi. Questa metodologia permette di rappresentare in modo dettagliato la presenza e la continuità delle attività di ogni individuo, fornendo una visione precisa e integrata della popolazione dimorante sul territorio.

Riprendendo il flusso di lavoro illustrato nel Prospetto 3, una volta ridotti i segnali di lavoro, studio, e altri segnali diretti per i singoli individui e determinato il Comune prevalente in cui si svolgono tali attività si procede con l'integrazione dei segnali di vita diretti con il Registro di popolazione RBI. In particolare, nello Step 4 si effettua lo scarico del RBI al 31 dicembre 2022 e al 31 dicembre 2023. Successivamente, per ciascun individuo si confrontano i segnali di vita diretti determinando, per ciascun Comune:

- gli individui con dimora abituale in Italia non residenti nel RBI;
- gli individui residenti nel RBI privi di segnali di vita diretti

Il passo successivo (Step 5) prevede lo scarico delle fonti fiscali e il confronto dei segnali di vita indiretti (ad esempio, "familiari a carico", dei titolari di automobile o di unità immobiliare) con gli individui residenti nel RBI che, nello Step 4, risultavano privi di segnali di vita diretti.

Infine, nello Step 6, si determina la potenziale sovra copertura anagrafica del RBI, composta dagli individui residenti nel registro ma privi di segnali diretti e indiretti provenienti da tutte le fonti considerate. Per questo gruppo, viene avviata un'ulteriore fase di lavorazione che utilizza le informazioni sulle famiglie anagrafiche identificate nelle Liste anagrafiche comunali (LAC). In questa fase, si individuano, ad esempio, persone registrate come "coniugi" all'interno della stessa famiglia di intestatari che presentano segnali di lavoro, studio o altri segnali di vita diretti. Questo permette di evitare che tali individui, pur privi di segnali diretti e indiretti, vengano automaticamente inclusi nell'insieme delle persone classificate come sovra copertura del RBI.

Questo ha portato a non limitare l'attenzione esclusivamente ai segnali di vita provenienti dalle fonti amministrative disponibili, ma a valorizzare la ricchezza degli archivi amministrativi, integrandoli con i dati delle indagini censuarie. I ricercatori dell'Istat hanno adottato un approccio basato sul *Knowledge Discovery from Databases (KDD)*; utilizzando un processo strutturato e iterativo. In questo processo, parte delle variabili da analizzare vengono costruite progressivamente, durante la lavorazione dei segnali amministrativi. Tale approccio consente di sfruttare al meglio l'ampia mole di dati, sia amministrativi che di indagine, per individuare criteri e profili utili al trattamento di specifiche sottopopolazioni, garantendo una maggiore precisione e affidabilità nell'analisi.

## La definizione finale degli aggregati di sovra e sotto copertura di RBI al 31.12.2023

Nei paragrafi precedenti si è descritto come il processo di integrazione delle fonti sia stato applicato per il conteggio del 2023, con l'obiettivo di identificare due sottogruppi di popolazione che consentono di correggere il Registro Base degli Individui (RBI) al 31 dicembre 2023:

- 1) Individui dimoranti abitualmente in Italia: si tratta di persone che mostrano segnali di vita diretti per almeno un anno negli archivi amministrativi. Poiché non risultano registrati come residenti nel RBI al 31 dicembre 2023, rappresentano la sotto copertura delle anagrafi comunali alla stessa data.
- 2) Individui residenti nel RBI senza segnali di vita diretti o indiretti: sono persone registrate come residenti nel RBI al 31 dicembre 2023, ma prive di segnali di vita nei principali archivi amministrativi. Tali individui rappresentano la sovra copertura delle anagrafi comunali.

L'impianto metodologico adottato si distingue per l'uso degli archivi amministrativi in modo indipendente rispetto alle informazioni anagrafiche comunali. Questo approccio richiama il principio di universalità del censimento tradizionale della popolazione: tutti i Comuni sono analizzati nello stesso momento temporale, applicando tecniche standardizzate sia in termini di metodologia (ad esempio, i profili di continuità) sia nel trattamento dei dati individuali.

Ai fini del conteggio della popolazione 2023 e a valle dei segnali diretti e indiretti sopra descritti sono stati utilizzati, per la prima volta, anche le informazioni sulle utenze domestiche di luce e gas relative all'annualità 2023. L'Acquirente Unico ha fornito all'Istat le informazioni dei consumi di elettricità e gas dei singoli



contatori distribuiti su tutto il territorio nazionale. In particolare, sono stati forniti 28.063.837 record per i consumi elettrici mensili del 2023, articolati in 23.758.216 record riferiti a persone fisiche e 12.878.928 record riferiti a unità giuridiche; invece, per le utenze di gas sono stati forniti 21.883.169 record per i consumi di gas annuali, articolati in 20.634.458 record riferiti a persone fisiche e 7.574.034 record riferiti a unità giuridiche. I dati non contengono l'indirizzo (via e numero civico) del POD o dell'intestatario del contratto, per cui il collegamento con il RBI è di tipo semi probabilistico, ovvero: è deterministico nel caso in cui all'interno della famiglia esiste un solo codice individuo titolare di contratto, ed è invece probabilistico quando nella stessa famiglia ci sono più titolari di contratto.

In primo luogo è stata effettuata una distinzione dei diversi profili di utilizzo di luce e gas in modo da caratterizzare ciascun profilo in volumi effettivi di consumo. Per ciascun profilo di utilizzo sono definite distribuzioni mensili di consumo normalizzate, cioè che prescindono dalla quantità, ovvero il volume di energia elettrica e di gas effettivamente utilizzato. Ciascun profilo è stato quindi ulteriormente raggruppato in funzione del volume di consumo ed ogni gruppo di consumo può essere, approssimativamente, ricondotto al numero di utenti. È possibile quindi associare ad ogni gruppo di consumo un numero medio di utenti aspettati. Per utilizzare i consumi all'interno del processo del conteggio censuario, è stata introdotta una metrica di distanza tra la famiglia, il numero di componenti risultanti nel RBI ed i contatori collegabili alla famiglia attraverso il codice identificativo della persona titolare del contratto elettrico.

La metrica di distanza si è basata su tre valori: la distanza tra numero di componenti in famiglia e il numero di componenti predetto dal profilo di consumo; la distanza tra consumi elettrici mensili effettivi del POD e il consumo elettrico mensile medio del gruppo di appartenenza; e, dove possibile, la distanza tra consumo annuo di gas riconducibile alla famiglia di appartenenza e consumo medio di gas dei diversi gruppi di consumo.

Ai fini del conteggio della popolazione al 31.12.2023, a valle dei segnali diretti e indiretti del processo di integrazione tra RBI e i segnali di vita di altre fonti, sono stati confermati nella popolazione come residenti tutti quegli individui che appartengono ad una famiglia la cui utenza si caratterizza per un consumo medio mensile di elettricità (piuttosto che consumo annuale di gas) che corrisponde alla media comunale dei consumi per numero dei componenti della famiglia.

Inoltre, nel 2023, così come negli altri anni precedenti, a valle del processo di integrazione dei segnali di vita amministrativi e dei consumi elettrici e di gas, sono stati individuati alcuni criteri che hanno consentito di considerare i fattori di contesto utili a migliorare l'identificazione della sovra e sotto copertura del RBI. Ad esempio, con riferimento alla sotto copertura, il Prospetto 6 evidenzia che, per tutti gli individui di cittadinanza straniera per i quali sono stati identificati i segnali di vita diretti, il permesso di soggiorno da solo non costituisce una condizione sufficiente per l'identificazione della sotto copertura. Tuttavia se il segnale di vita diretto è localizzato in un Comune frontaliero, gli individui stranieri con tale segnale non vengono considerati dimoranti abitualmente in Italia per ragioni legate ai movimenti transfrontalieri per motivi di lavoro o studio da parte di cittadini di paesi confinanti con l'Italia.

#### PROSPETTO 6 - CRITERI PER L'IDENTIFICAZIONE DEGLI INDIVIDUI DIMORANTI ABITUALMENTE IN ITALIA



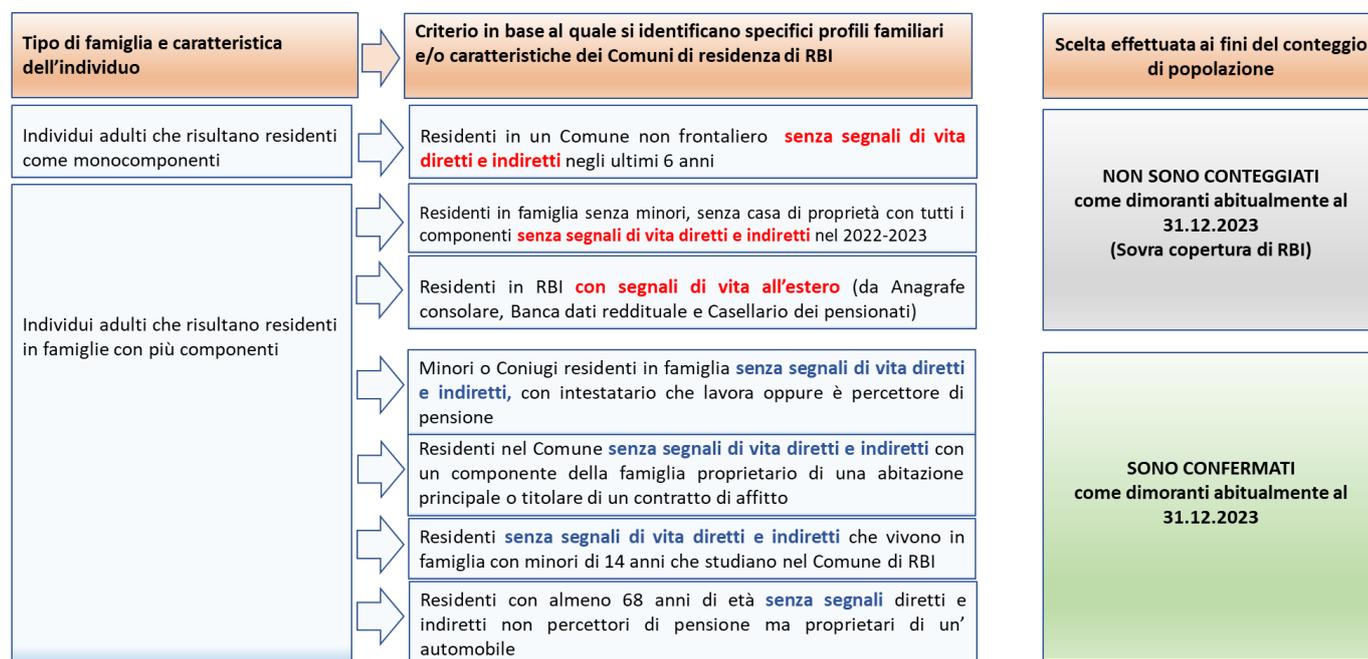
Fonte: Istat, 2024



In ogni caso, è importante sottolineare che, in assenza della valorizzazione del comune di domicilio fiscale, del luogo del contratto di affitto, di lavoro o di studio nella fonte amministrativa di origine, non sarebbe possibile attribuire un Comune di dimora abituale agli individui classificati come appartenenti alla sotto copertura.

Un'analoga analisi è stata condotta in relazione agli individui privi di segnali diretti e indiretti classificati nella sovra copertura di RBI. A valle del processo di integrazione, e grazie alle evidenze emerse dall'applicazione del modello, sono stati definiti ulteriori criteri o profili di individui. Seguendo un approccio iterativo e una logica gerarchica, questi criteri hanno portato alla costruzione di nuove variabili a livello individuale o familiare, come illustrato nel Prospetto 7.

**PROSPETTO 7 - CRITERI DETERMINISTICI IN BASE AI QUALI I RESIDENTI NEL RBI SENZA SEGNALI DIRETTI E INDIRETTI SONO STATI CONFERMATI COME RESIDENTI OPPURE CLASSIFICATI COME INDIVIDUI IN SOVRA COPERTURA**



Fonte: Istat, 2024

Questo passaggio aggiuntivo ha permesso, in alcuni casi, di rafforzare l'assenza di segnali di vita sul territorio già evidenziata dagli output del processo di integrazione classificando così la classificazione come sovra copertura degli individui residenti di RBI senza segnali diretti e indiretti. In altri casi, la creazione di nuove variabili legate al contesto familiare degli individui inizialmente classificati come sovra copertura ha consentito di ricollocarli tra le persone abitualmente dimoranti in Italia, confermandone la presenza nel RBI ai fini del conteggio.

Inoltre, è stata confermata la residenza nel RBI ai fini del conteggio, per tutti gli individui residenti in comuni molto piccoli, nei quali gli indicatori longitudinali di affidabilità delle anagrafi comunali hanno mostrato un'elevata robustezza. Indicazioni in tal senso sono emerse anche da numerose analisi esplorative condotte dai ricercatori dell'Istat nel corso degli ultimi cinque anni.

Sono stati altresì confermati come residenti tutti gli individui che vivono in convivenze anagrafiche in ogni Comune. Infine verifiche analoghe sono state condotte anche per la popolazione senza tetto/senza fissa dimora dei 14 grandi Comuni dell'area metropolitana e per i grandi anziani (persone con almeno 98 anni di età) per i quali si è proceduto a confermare la loro dimora abituale nel RBI.

In sintesi, le persone identificate attraverso i segnali di vita diretti nel prospetto 6 e quelle individuate tramite i segnali di vita indiretti (i "confermati" del prospetto 7) costituiscono gli aggregati fondamentali per il conteggio della popolazione.



## Attribuzione della cittadinanza e dello stato di nascita

Ai fini del conteggio della popolazione 2023, l'attribuzione della cittadinanza viene effettuata a valle del processo di integrazione tra le fonti anagrafiche e gli altri archivi amministrativi. L'informazione sul paese di cittadinanza deriva principalmente dalle Liste Anagrafiche Comunali (LAC) acquisite direttamente dalla Anagrafe Nazionale della Popolazione residente (ANPR). Tuttavia per la produzione del dato di popolazione questa informazione deve essere sottoposta ad un processo di controllo e correzione, in quanto si riscontrano frequentemente valori anomali o non conformi alla classificazione Istat dei paesi di cittadinanza aggiornata annualmente dall'Istituto. Le principali criticità individuate sono le seguenti:

- codici nazionalità mancanti;
- sovrastima dei codici 'Apolidi'
- presenza del codice 'In via di definizione'.

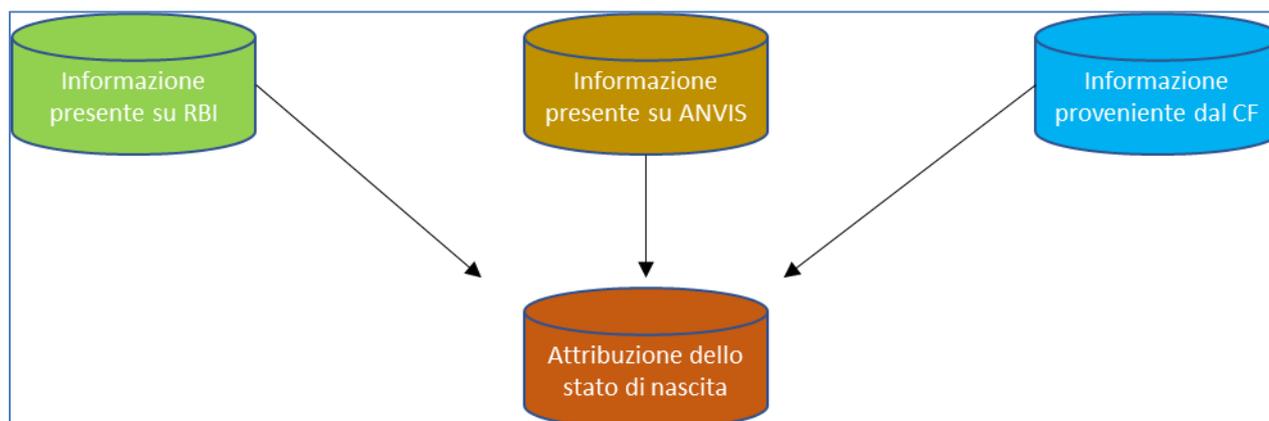
Per gli individui con cittadinanza mancante l'attribuzione avviene utilizzando l'informazione presente nel RBI dell'anno precedente. Tuttavia, questo approccio potrebbe introdurre errori per le persone che hanno acquisito la cittadinanza italiana nell'anno di riferimento. Per evitare tali incongruenze tutti gli individui con cittadinanza straniera vengono confrontati con la fonte dei "Giuramenti di cittadinanza italiana" fornita dal Ministero dell'Interno. Per garantire la massima accuratezza del dato vengono considerati i giuramenti degli ultimi 5 anni.

Un aggregato di particolare interesse per gli organismi internazionali come l'Alto Commissariato delle Nazioni Unite per i Rifugiati (UNHCR), che effettua un monitoraggio annuale su questa categoria, è quello degli 'Apolidi'. Nelle LAC, questo aggregato, presenta una numerosità particolarmente sovrastimata. La procedura di correzione prevede l'estrazione di tutti gli individui classificati come 'Apolidi' che risultano avere una relazione di parentela definita come 'figlio dell'intestatario e/o del coniuge/convivente' rispetto all'intestatario della scheda anagrafica. Successivamente, la cittadinanza di questi individui viene confrontata con quella dell'intestatario della scheda anagrafica e del coniuge/convivente. Nel caso in cui né l'intestatario né il coniuge/convivente risultino avere cittadinanza 'Apolide', l'individuo 'eredita' la stessa cittadinanza dell'intestatario di scheda anagrafica. Per quanto riguarda le persone con cittadinanza classificata come 'In via di definizione' queste sono perlopiù individui nati nel corso dell'anno di riferimento. L'attribuzione della cittadinanza avviene in due passaggi: il primo segue lo stesso criterio utilizzato per gli 'Apolidi'; il secondo incrocia i residui del primo passaggio con i dati di cittadinanza derivanti dalla rilevazione 'Iscritti in anagrafe per nascita'. Questo consente di attribuire la cittadinanza anche a quei neonati che non risultino avere una relazione di parentela diretta con l'intestatario della scheda anagrafica o il coniuge/convivente'.

Un ulteriore aggregato rilevante per l'attribuzione della cittadinanza è rappresentato dagli individui in sotto copertura rispetto alle anagrafi comunali, identificati come residenti sulla base di segnali di vita "forti" emersi dal processo di integrazione tra il Registro Base degli Individui (RBI) e le altre fonti amministrative. Per questi individui, il processo di attribuzione della cittadinanza prevede diversi passaggi. In primo luogo, si verifica, tramite il codice identificativo unico, la disponibilità dell'informazione sulla cittadinanza nell'archivio dei permessi di soggiorno. In seconda istanza, si consultano altre fonti amministrative, come l'archivio dei lavoratori domestici dell'INPS. Al termine di questi passaggi, per le poche decine di casi ancora senza un valido la cittadinanza viene attribuita utilizzando lo stato di nascita, che rappresenta una buona proxy della cittadinanza. Per quanto riguarda l'attribuzione dello stato di nascita, è stato effettuato un controllo di coerenza all'interno dell'RBI. Agli individui con provincia di nascita valorizzata è stato automaticamente assegnato lo stato di nascita 'Italia'. I casi anomali sono stati confrontati con i codici della classificazione EUROSTAT e, in caso di mancata conformità, opportunamente corretti. Nei casi in cui le informazioni presenti sul RBI non permettevano di attribuire un paese di nascita valido, i valori anomali sono stati eliminati ('sbiancati'). Per i codici che permettevano una ricostruzione compatibile con la classificazione EUROSTAT è stata effettuata una ricodifica. Concluse queste operazioni di correzione preliminari sul RBI, l'informazione relativa allo stato di nascita è stata confrontata con quella proveniente dall'Anagrafe Virtuale Statistica (ANVIS) e con i dati del codice Belfiore associati al codice fiscale, al fine di garantire la massima accuratezza e coerenza.



## PROSPETTO 8 - SCHEMA DI SINTESI ATTRIBUZIONE STAO DI NASCITA



Fonte: Istat, 2024

Nei casi in cui le tre fonti (RBI, ANVIS, Codice Fiscale) forniscano informazioni discordanti, è stata attribuita l'informazione che aveva la maggiore frequenza di occorrenze valide. Quando tutte e tre le fonti presentavano valori discordanti, si è considerata come corretta l'informazione proveniente dal RBI. Questo approccio ha permesso di costruire una variabile temporanea che, tuttavia, non tiene conto della data di scissione o unione degli stati, ma solo del codice dello stato valido secondo le classificazioni ufficiali di Istat ed Eurostat relative ai codici degli stati di nascita.

### Attualizzazione dei confini degli stati di nascita

Per garantire la conformità con i requisiti di Eurostat per i dati censuari, i codici degli stati relativi al luogo di nascita devono essere attualizzati ai confini internazionali dell'anno di riferimento del censimento. Pertanto, tutti i codici degli stati non più esistenti (ad esempio, URSS, Cecoslovacchia, Jugoslavia) devono essere aggiornati. Mentre per i paesi che si sono aggregati non si riscontrano criticità, per gli stati che si sono scissi è necessaria una procedura articolata in due fasi:

**Prima fase:** Per gli individui nati in paesi interessati da una scissione, si verifica se almeno una delle tre fonti (RBI, ANVIS o Codice Fiscale) contiene un codice stato di nascita congruente con i confini attuali. In caso affermativo, viene assegnato lo stato di nascita trovato. Se nessuna delle fonti dispone di un'informazione congruente, si passa alla seconda fase.

**Seconda fase:** Si adotta una procedura probabilistica che imputa a ciascun individuo uno degli stati attuali derivanti dal paese non più esistente. La probabilità è calcolata attraverso due possibili approcci:

- A livello comunale: Si calcolano le percentuali di distribuzione degli individui nati nei nuovi stati derivati (ad esempio, Repubblica Ceca e Slovacchia nel caso della Cecoslovacchia) rispetto al totale comunale.
- A livello nazionale: Se non sono disponibili distribuzioni a livello comunale, la probabilità viene calcolata considerando la distribuzione degli stati a livello nazionale.

Le percentuali così determinate costituiscono la probabilità di assegnazione dello stato di nascita (Prospetto 9). Successivamente, a ciascun individuo viene attribuito uno stato di nascita tramite un numero pseudocasuale (compreso tra 0 e 1), che determina in modo univoco l'imputazione. A titolo esemplificativo, si riporta una tabella che illustra i tre stati non più esistenti che influenzano maggiormente la ricodifica probabilistica, con le relative percentuali di distribuzione e i nuovi stati derivati:



**PROSPETTO 9 - ESEMPIO DI RICODIFICA STATO ESTERO DI NASCITA**

Stato non più esistente	Stati attuali derivati	Percentuale (esempio comunale)	Percentuale (esempio nazionale)
Cecoslovacchia	Repubblica Ceca, Slovacchia	60% - 40%	65% - 35%
URSS	Russia, Ucraina, altri	50% - 30% - 20%	55% - 25% - 20%
Jugoslavia	Serbia, Croazia, altri	45% - 35% - 20%	50% - 30% - 20%

Fonte: Istat, 2024

Questa procedura garantisce un'assegnazione coerente e standardizzata degli stati di nascita, rispettando i confini internazionali aggiornati (Prospetto 10).

**PROSPETTO 10 - RICODIFICA DEGLI STATO ESTERO DI NASCITA**

Stato non più esistente	Stato esistente
Ex Cecoslovacchia	Repubblica Ceca Slovacchia
Ex Jugoslavia	Bosnia-Erzegovina Croazia Kosovo Macedonia del Nord Montenegro Serbia Slovenia
Ex URSS	Armenia Azerbaijan Bielorussia Estonia Federazione russa Georgia Kazakhstan Kirghizistan Lettonia Lituania Moldova Tagikistan Turkmenistan Ucraina Uzbekistan

Fonte: Istat, 2024