



## Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2022

### Dalla rilevazione “porta a porta” al Censimento permanente

Da ottobre 2018 l'Istat ha avviato il Censimento permanente della popolazione e delle abitazioni (che sostituisce il Censimento della popolazione decennale), basato sull'integrazione delle informazioni reperibili dalle fonti amministrative con quelle acquisite dalle indagini campionarie effettuate a rotazione su tutti i comuni italiani.

Nel 2020, a causa del Covid-19, non è stato possibile realizzare le rilevazioni sul campo; tuttavia, l'Istat si è posto comunque l'obiettivo di produrre il conteggio della popolazione residente per genere, età, cittadinanza e grado di istruzione. Attraverso una opportuna metodologia, si sono integrate tutte le informazioni amministrative a disposizione; è stato possibile contare anche su alcune nuove fonti dell'INPS, che consentono di cogliere segnali di vita sulla base delle misure a sostegno di imprese e lavoratori previste proprio nel 2020 per fronteggiare gli effetti economici della crisi indotta dalla pandemia.

Nel 2021 l'Istat ha potuto di nuovo effettuare le rilevazioni censuarie. La ricchezza informativa proveniente dalle fonti amministrative, già studiata ed utilizzata a partire dal 2020 per il conteggio della popolazione, è stata integrata in occasione del 2021 con i risultati delle indagini (da Lista e Areale) attraverso un modello statistico che ha permesso di consolidare la metodologia utilizzata anche per il conteggio della popolazione 2022.

L'utilizzo delle fonti amministrative nella strategia del Censimento permanente era stato immaginato subito dopo il censimento del 2011, quando l'Istat ha iniziato a operare la transizione da un Censimento tradizionale “porta a porta” al Censimento permanente. Nel 2011 infatti, pur nell'ambito di una metodologia ancora sostanzialmente convenzionale (rilevazione sul campo esaustiva e simultanea), ci fu un significativo ricorso all'uso degli archivi per guidare la rilevazione: le Liste Anagrafiche Comunali, per la spedizione postale dei questionari alle famiglie iscritte in anagrafe, e la Lista Integrativa da Fonti Ausiliarie, per rilevare la popolazione abitualmente dimorante non ancora iscritta in anagrafe.

Per il conteggio della popolazione nel 2022 è stato possibile avvalersi, inoltre, dei progressi in termini di qualità e tempestività via via conseguiti dai Registri a supporto della produzione statistica ufficiale. In questo contesto il perno del Censimento permanente della popolazione è costituito dal Registro di Base degli Individui, delle famiglie e delle convivenze (d'ora in avanti RBI) che, insieme ai registri tematici (quali quelli sull'occupazione e sull'istruzione) e alla regolare acquisizione di fonti amministrative, e dal loro trattamento e utilizzo a fini statistici, ha permesso ormai da quasi un decennio di integrare i dati di fonte anagrafica dei Comuni 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 a supporto dei processi produttivi statistici; in particolare, è l'infrastruttura di base per la produzione delle statistiche ufficiali sulla popolazione e il riferimento per l'estrazione dei campioni delle indagini previste per il Censimento permanente (cfr. legge censuaria e Piano Generale di Censimento) e di tutte le indagini campionarie sulle famiglie.

In accordo con l'impianto metodologico del Censimento permanente, e nel rispetto delle norme sulla protezione dei dati personali, il RBI e tutte le fonti amministrative non anagrafiche disponibili in Istat contengono dati anonimi “statistici”, ossia derivanti da un processo di elaborazione e validazione statistica e sono riferiti a un numero limitato di variabili funzionali alla rappresentazione delle principali caratteristiche strutturali della popolazione.

Il RBI è “consolidato” con cadenza annuale con riferimento al 31 dicembre di ciascun anno. Questo consolidamento è realizzato mediante l'integrazione dei flussi individuali anagrafici della dinamica demografica (nascite, decessi, trasferimenti di residenza da e per altro comune o da e per l'estero). L'applicazione del modello di contabilità demografica MIDEA (Micro-DEmographic Accounting) consente di sfruttare le potenzialità della base dati micro (flussi+stock) per produrre indicatori più accurati e innovativi sulla dinamica demografica, tenendo conto della sequenza degli eventi demografici sperimentati dagli individui.

I principali output del sistema MIDEA applicati al RBI “consolidato” sono il bilancio demografico della popolazione residente comunale e la struttura della popolazione residente comunale per genere, età, stato civile e cittadinanza. La produzione tempestiva di questi output anticipa i risultati definitivi annuali del Censimento permanente, con l'obiettivo di assicurare la continuità della produzione statistica ufficiale,



soddisfacendo al tempo stesso le tempistiche stringenti richieste tanto dagli utilizzatori istituzionali dei dati di popolazione, quanto dal regolamento europeo sulle statistiche demografiche.

Questi stessi output vengono rielaborati e diffusi successivamente come definitivi a seguito della correzione del RBI mediante gli esiti del Censimento permanente.

Come per tutto il primo ciclo il conteggio del Censimento permanente del 2022 è finalizzato a correggere gli errori di sovra e sotto-copertura del RBI, individuando le persone presenti nel Registro come residenti ma non trovate negli archivi amministrativi (sovra copertura) e quelle che invece sono state trovate nei dati amministrativi come abitualmente dimoranti ma non risultano tali nel Registro (sotto copertura). Questa correzione, agisce a livello micro, opera infatti attraverso la riclassificazione dei record individuali contenuti nel Registro riconoscendoli come abitualmente dimoranti o meno, sulla base dei segnali di vita amministrativi. Si consolida quindi una innovazione metodologica di rilievo che assicura la corrispondenza in termini di “teste” tra il dato di conteggio e i record di individui abitualmente dimoranti contenuti nel RBI, a differenza di quanto fatto nel 2018 e nel 2019 in cui l’obiettivo della correzione del conteggio veniva conseguito a livello macro, applicando dei pesi (correttori per sovra e sotto copertura) agli individui residenti presenti nel RBI.

Questo impianto garantisce, e lo farà in misura sempre maggiore nei prossimi anni, una completa integrazione con gli output delle statistiche demografiche, contribuendo in tal modo a una maggiore stabilità e coerenza di tutto il sistema dei dati che riguardano la popolazione. Il processo censuario vedrà un sempre maggiore utilizzo delle informazioni provenienti dagli archivi amministrativi per stimare il conteggio di popolazione e, laddove 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 2022 della popolazione**

Ai fini del conteggio della popolazione del 2022, il Servizio Censimento della popolazione ha creato un repository che integra le informazioni derivanti dalle fonti amministrative con quelle anagrafiche organizzate nel Sistema Integrato di Microdati (SIM). La base dati del SIM costituisce di fatto una base dati relazionale di dati amministrativi che, garantendo l’anonimato degli individui, è stato costruito con lo scopo di sostenere i processi di produzione statistica, sia per le statistiche sociali che per le statistiche economiche. L’assegnazione di un codice ID univoco e costante permette di identificare ciascun individuo e unità economica all’interno dei diversi archivi e di costruire le relazioni tra le diverse fonti, garantendo al contempo un trattamento di dati privi di identificativi diretti.

Attraverso le indicazioni fornite dagli esperti di fonti amministrative dell’Istat che usano gli archivi per la produzione di statistiche sulla popolazione, il Servizio Censimento permanente della popolazione ha effettuato una selezione delle fonti da prendere in considerazione per il conteggio della popolazione dimorante abitualmente e, attraverso un gruppo di esperti tematici e metodologi, ha stabilito un ordine gerarchico delle fonti nel processo di integrazione che costituisce la base di riferimento per il conteggio della popolazione. Ciò ha permesso poi di osservare i “segnali di vita” degli individui in termini di “dimora abituale in Italia”. Il data base in questione ha integrato, per il conteggio della popolazione 2022, le informazioni anagrafiche individuali di RBI con quelle del Registro Tematico del Lavoro (RTL) dell’Istat, degli archivi dell’istruzione, delle dichiarazioni fiscali e degli archivi della previdenza sociale, nonché del catasto immobiliare dell’Agenzia delle Entrate con un riferimento temporale che va dal primo gennaio 2021 al 31 dicembre 2022.

Nell’ambito delle fonti utilizzate, il Prospetto 1 mostra, per i due anni (2021 e 2022), la denominazione dell’archivio, l’ente titolare fornitore, il numero di record trattati per ciascun anno e la classificazione per tipo di segnale di vita che, secondo la logica gerarchica assegnata dagli esperti del Censimento permanente della popolazione, consente di classificare gli individui ai fini della dimora abituale in Italia.

Complessivamente, i record trattati nel repository di integrazione per ciascun anno sono diverse centinaia di milioni considerando, però, che ciascun individuo può dare luogo in tempi diversi a più di una occorrenza nello stesso archivio, e che lo stesso individuo può apparire allo stesso istante in diversi archivi.

Nel costruire la base dati integrata delle fonti amministrative e anagrafiche si è tenuto conto della periodicità dei dati e della rilevanza rispetto alla definizione di dimora abituale secondo quanto è previsto dal regolamento europeo (regolamento CE n. 1260/2013). Dal Prospetto 1 è possibile osservare che le fonti assicurative dell’INPS relative agli individui che partecipano al mercato del lavoro sono in gran parte disponibili sia per il 2021 che per il 2022. Invece, le fonti del MIUR relative agli Archivi del personale e le fonti previdenziali relative al Casellario dei pensionati e assistenziali, hanno un riferimento data che copre solo il 2021; inoltre, ci sono alcune fonti dell’INPS relative ai sostegni sociali e assistenziali, quali il reddito



di emergenza, i benefici per le indennità Covid-19 e i bonus *baby-sitting*, che sono segnali di vita messi a disposizione dell'Istat per la prima volta per l'anno 2020 e solo alcune (indennità Covid-19) sono confermate solo per il 2021.

**PROSPETTO 1 - ARCHIVI AMMINISTRATIVI E REGISTRI TEMATICI CHE FORNISCONO SEGNALI DI VITA DIRETTI E INDIRETTI RISPETTO ALLA DIMORA ABITUALE IN ITALIA AL 31.12.2022**



SEGNALI DI VITA PER ORDINE GERARCHICO	ENTE TITOLARE	DENOMINAZIONE ARCHIVIO	N. record	
			2021	2022
1) Segnali diretti di lavoro	INPS	<del>UNIEmens</del>	176.871.085	187.035.402
		DMAG (dichiarazione sulla manodopera agricola)	8.325.480	8.105.426
		Lavoratori autonomi in agricoltura	441.854	435.085
		Archivio dei parasubordinati collaboratori	7.655.989	8.167.475
		Rapporti di lavoro domestico	1.249.355	1.186.231
		Gestione Dipendenti Pubblici - Posizioni degli assicurati iscritti	40.197.852	
		Lavoro occasionale (Libretto Famiglia e Prestazioni Occasionali)	2.423.979	1.423.775
		Archivio dei lavoratori autonomi: artigiani e commercianti	3.690.379	3.617.947
		Anagrafe Tributaria delle persone giuridiche e delle persone fisiche con partita IVA	6.272.870	6.758.764
	MIUR	Archivio del personale universitario	458.858	
		Archivio del personale delle scuole statali	1.162.314	
	INAIL	Lavoratori interinali	1.545.488	1.770.348
	MEF	NOIPA - Anagrafica (Cedolini Stipendiali dipendenti PA)	22.672.180	
ISTAT	Registro Tematico del Lavoro (RTL)	26.678.692	27.326.658	
2) Segnali diretti di studio	MIUR	Anagrafe degli studenti delle scuole	17.107.651	16.915.320
		Archivio degli iscritti e delle iscrizioni universitarie	1.974.437	
		Archivio delle lauree e dei laureati	524.187	505.263
		Anagrafe Nazionale Studenti dei corsi Post laurea (iscrizioni)	172.339	
		Anagrafe Nazionale Studenti dei corsi Post laurea (chiusure)	70.542	68.555
3) Segnali diretti di assistenza sociale e contratti di affitto ad uso abitativo	INPS	Beneficiari CIG a pagamento diretto	8.435.339	436.351
		Trattamenti non pensionistici	8.046.956	
		Casellario dei Pensionati e dei trattamenti pensionistici	22.758.714	
		Possessori della Social Card	394.586	
		Reddito di cittadinanza (RDC)	1.771.535	1.690.536
		Indennità Covid-19 (Bonus 2400 euro Decreto sostegni 2021)	1.289.806	
	Bonus <del>baby sitting</del> e centri estivi (emergenza Covid-19)	97.605		
	MEF	Banca dati statistica reddituale	41.496.807	
Agenzia delle Entrate - Contratti di Locazione		4.477.448	4.393.391	
4) Segnali indiretti	MEF	Agenzia delle Entrate - Dichiarazioni Modello Redditi Persone Fisiche	13.686.974	
		Agenzia delle Entrate Dichiarazioni 730	36.418.121	
		Agenzia delle Entrate - Certificazione Unica	49.214.785	
		Agenzia delle Entrate - Catasto delle Unità Immobiliari	111.343.342	
		Agenzia delle Entrate - Catasto delle Unità Immobiliari – Bolzano	877.829	
		Agenzia delle Entrate - Catasto delle Unità Immobiliari – Trento	1.530.517	
	ACI	ACI - Parco veicoli circolanti	50.872.976	48.691.246
Fonti anagrafiche e altre fonti utilizzate	ISTAT	Registro Base degli INDIVIDUI (RBI)		103.208.269
	MIN. INTERNO	Permessi di soggiorno	3.720.669	4.175.975
		Acquisizione e reiezione della cittadinanza italiana	52.605	
	MAE	MAE - Ministero Affari Esteri - Archivio Italiani all'Estero	9.332.437	9.660.885
	MEF	Agenzia delle Entrate - Anagrafe delle Persone Fisiche (AT)	103.004.529	104.585.588

Fonte: Istat 2023



Ritardi rispetto alla data di riferimento si registrano per alcune fonti delle dichiarazioni fiscali, in particolare quelle della Banca dati reddituale e le informazioni dei modelli UNICO/730, per i quali si osservano tempi di utilizzazione delle informazioni con un ritardo data di circa 20 mesi. Ciò è dovuto alla natura delle fonti fiscali, che solitamente si rifanno alle dichiarazioni dei redditi dell'anno precedente a quello della data di riferimento.

Nel corso degli ultimi anni il sistema integrato dei dati amministrativi e anagrafici ha conosciuto una notevole evoluzione non solo in riferimento all'impiego di registri tematici, quali RBI e RTL, ma anche dal punto di vista tecnico e informatico consentendo di sfruttare al meglio, e con tempi di lavorazione sempre più rapidi, le informazioni disponibili. Altrettanto significative sono state le attività che hanno portato ad una ingegnerizzazione di processo e di prodotto e ad aggiornare i codici di provincia e comune migliorando la qualità di localizzazione dei segnali delle fonti amministrative.

I ricercatori tematici del Servizio del Censimento permanente della popolazione sono arrivati a una definizione dei segnali di vita (più noti nella letteratura internazionale come “*Signs of administrative life*”) adottata anche nel Piano Generale di Censimento del 2022.

*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.*

Dalla definizione adottata è possibile evincere anche una classificazione gerarchica dei segnali, secondo quanto riportato nella prima colonna del Prospetto 1: i segnali di lavoro e studio, insieme ai contratti di affitto, alle dichiarazioni della Banca dati reddituale<sup>1</sup> e ai sussidi socioassistenziali dell'INPS sono classificabili come segnali di vita diretti rispetto alla dimora abituale in Italia.

Gli archivi dell'INPS relativi al lavoro e quelli del MIUR sulla frequenza di corsi scolastici e universitari sono particolarmente rilevanti ai fini della individuazione della popolazione abitualmente dimorante in Italia. Il Prospetto 2 mostra che queste fonti offrono un notevole dettaglio informativo, legato non solo alla durata dell'attività, ma anche alla localizzazione della stessa (Comune e indirizzo) e ad alcuni attributi specifici dell'attività svolta (tipologia di contratto di lavoro, corso di studi, ecc.) che forniscono elementi rilevanti nel valutare la forza 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, 2023

<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.



Invece, tutte le fonti fiscali (modelli UNICO, modelli 730, ecc.), il possesso di automobili presenti nell'archivio ACI e il possesso di immobili presenti nell'archivio catastale forniscono segnali indiretti rispetto alla dimora abituale. Infatti, a partire dal “Quadro Familiari a carico” dei modelli UNICO Persone Fisiche e 730, si possono individuare le relazioni principali tra “coniugi” e tra il genitore dichiarante e “figlio/i”. Nella base dati dei modelli fiscali sono contenuti, inoltre, anche i dati relativi agli “altri familiari” a carico (il coniuge legalmente ed effettivamente separato; i discendenti dei figli; i genitori, i generi e le nuore; il suocero e la suocera; i fratelli e le sorelle; i nonni e le nonne) a patto che vivano con il dichiarante o che ricevano da lui assegni alimentari non risultanti da provvedimenti dell'Autorità giudiziaria. Dato che questi segnali di presenza in Italia sono desunti “indirettamente” dalla dichiarazione di un percettore di reddito, si è deciso di classificarli come segnali indiretti; così anche il possesso di automobile o di un'unità immobiliare non è considerato un segnale diretto e, in quanto tale, viene associato ai segnali indiretti.

Altri tipi di segnali indiretti sono quelli che si ricavano da RBI e che si riferiscono alla composizione del nucleo familiare registrato in anagrafe. In questo caso, come vedremo, le relazioni principali sono quelle tra intestatario della famiglia e “coniuge” e tra intestatario e “figlio/i”.

La base dati della popolazione residente con cui vengono effettuati i confronti in termini di coerenza con i segnali di vita in Italia o meno è costituita dagli individui contrassegnati come residenti nel RBI. MIDEA (Micro-DEmographic Accounting)/ANVIS (Anagrafe Virtuale Statistica) è il processo statistico che porta con cadenza annuale alla identificazione della popolazione presente nel RBI eleggibile ad essere considerata residente. La base dati ANVIS viene quindi consolidata ogni anno mediante l'aggiornamento a livello micro con i flussi della dinamica demografica (flussi di iscrizione per nascita; cancellazione per decesso; iscrizione/cancellazione da/per l'estero; iscrizione/cancellazione da/per altri Comuni italiani), verificatisi nei periodi dal 9 ottobre 2011 al 31 dicembre di ciascun anno.

Le informazioni anagrafiche sono state ottenute integrando le informazioni contenute nei diversi archivi inseriti nel SIM. In particolare, i dati utilizzati per ricostruire il profilo demografico degli individui hanno riguardato:

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

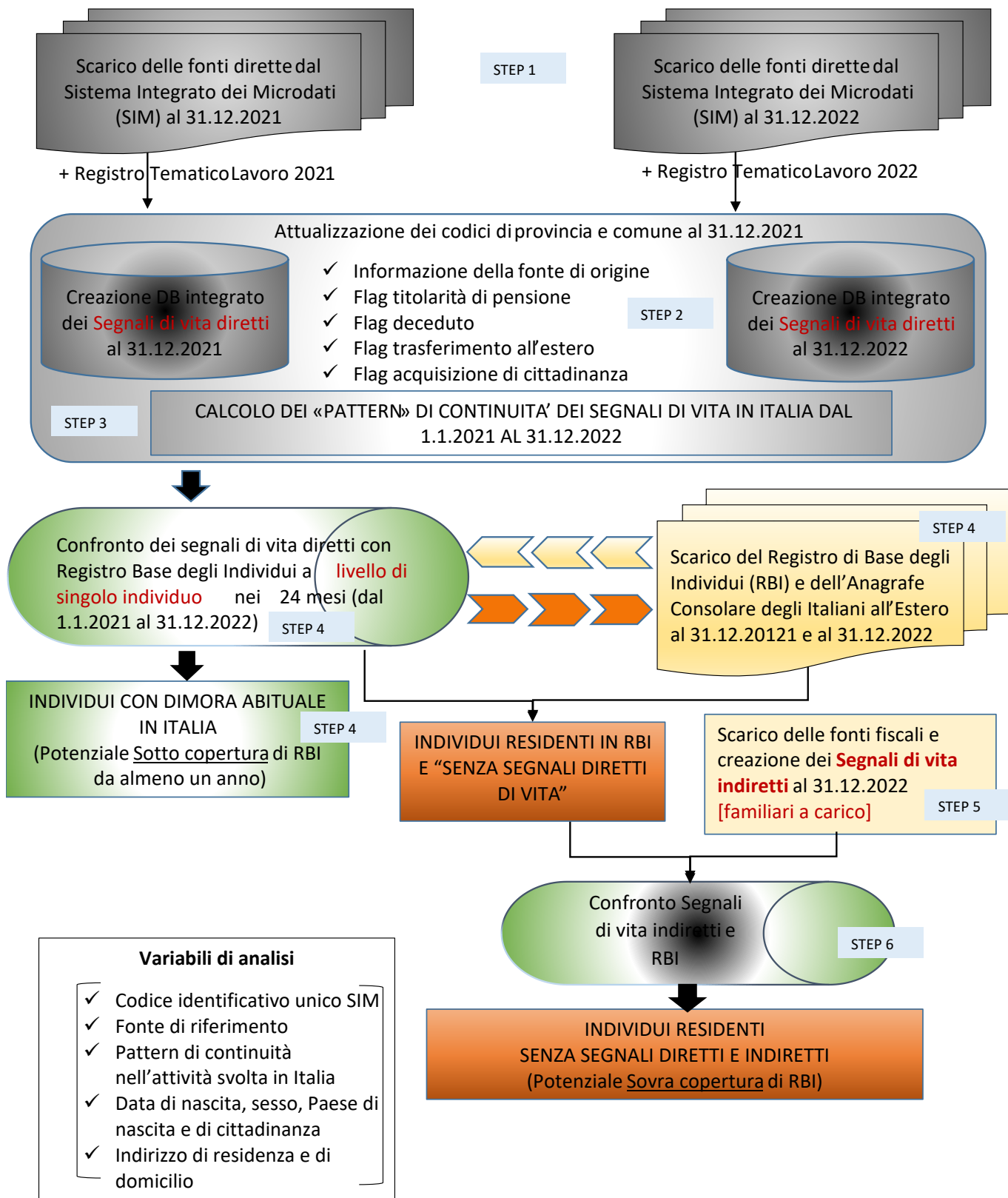
Qualora la Cittadinanza non fosse presente in uno degli archivi considerati (RBI e Permessi di Soggiorno) si è fatto riferimento all'informazione sul Paese di nascita. In presenza di informazioni anagrafiche incongruenti nelle diverse fonti per uno stesso individuo, sono state scelte le informazioni sulla base di una scala gerarchica delle fonti. La fonte dell'Anagrafe Tributaria, insieme a RBI, rappresenta la fonte più completa rispetto alle variabili anagrafiche relative a data di nascita, genere e paese di nascita. Si evidenzia, inoltre, che ai fini dell'analisi della effettiva localizzazione degli individui in termini di dimora abituale, risulta molto utile la presenza in questa fonte anche della variabile “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 mostra il flusso di lavoro, e i relativi passi, con cui è stato seguito il processo di integrazione delle fonti amministrative con RBI, dallo scarico delle fonti dal SIM fino alla determinazione della sotto e sovra copertura del RBI. Successivamente allo scarico dal SIM (*Step 1*), i dati derivanti dalle fonti amministrative prescelte sono stati strutturati in maniera adeguata all'analisi della popolazione dimorante abitualmente in Italia.



**PROSPETTO 3 - FLUSSO DI LAVORO DEI SEGNALI DI VITA PER LA DETERMINAZIONE DEL CONTEGGIO DI POPOLAZIONE DEL CENSIMENTO PERMANENTE DELLA POPOLAZIONE AL 31.12.2022**



Fonte: Istat, 2023



In primo luogo, il processo di integrazione delle fonti amministrative comporta il trattamento di dati provenienti da oltre quaranta archivi amministrativi, ognuno dei quali relativo a più anni, contenenti le informazioni di base sui segnali di vita degli individui. In ciascun archivio sono presenti, per ogni evento amministrativo registrato, le informazioni sulla localizzazione dell'evento stesso, tramite i codici di provincia e comune. Le stesse variabili sono poi utilizzate anche per ulteriori informazioni, quali la residenza, il luogo di nascita o il comune di provenienza. In generale queste informazioni presentano una serie di problematiche:

- ✓ valori validi ma disallineati temporalmente, o comunque non aggiornati alla data di riferimento;
- ✓ valori totalmente o parzialmente mancanti.

Il primo caso è imputabile all'acquisizione degli archivi in momenti diversi e quindi, a seguito delle variazioni amministrative intercorse, uno stesso comune può presentare in differenti annualità codici territoriali diversi. Il fatto poi che il calcolo degli output del processo integrato delle fonti amministrative richieda, come vedremo, l'utilizzo di dati relativi a due anni consecutivi rende il processo di produzione ancora più sensibile a questo tipo di problematica.

Il caso dei valori mancanti risulta invece essere rilevante in particolare per alcune fonti che forniscono segnali diretti di presenza sul territorio (ad es. le posizioni degli assicurati iscritti ex INPDAP), oppure l'informazione è disponibile solo in maniera parziale, con una limitazione all'utilizzabilità delle fonti stesse.

Ciò comporta che il Servizio Censimento effettui alcune analisi e delle successive operazioni, quali ad esempio:

- ✓ l'analisi di tipo territoriale, anche relativamente a il RBI e il RTL;
- ✓ la definizione di una localizzazione prevalente per i segnali di lavoro e studio;
- ✓ aggregazioni di tipo longitudinale basate su variabili geografiche;
- ✓ confronti temporali su diversi archivi;
- ✓ comparabilità con l'informazione proveniente da altre fonti.

Per gestire queste criticità è stata definita e sviluppata una procedura per l'imputazione e l'allineamento temporale delle variabili di localizzazione, facendo uso sia di informazioni presenti negli stessi archivi, sia di informazioni ausiliarie provenienti da fonti esterne, che è stata introdotta nel processo di integrazione dei dati (*Step 2*). Nella creazione del prodotto integrato dei dati amministrativi, illustrato nel Prospetto 3, sono stati quindi imputati ed aggiornati i codici di provincia e comune di tutte le fonti amministrative adoperate in input dal processo, rendendo così consistenti le informazioni rispetto ad un unico istante di riferimento, che è stato fissato al 31 dicembre 2022. L'imputazione dei valori è avvenuta sfruttando, quando possibile, vincoli per la determinazione univoca dei valori.

Successivamente all'aggiornamento dei codici di provincia e comune, sempre nello *Step 2* del processo di integrazione si procede alla costruzione, sia per il 2021 che per il 2022, della base dati integrata delle fonti di lavoro e studio dell'INPS e del MIUR nonché delle altre fonti con segnali di vita diretti (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, si producono anche i *flag* informativi delle fonti originarie relativi all'eventuale decesso, alla titolarità di una pensione, al trasferimento all'estero (desumibile da Banca dati Reddituale e dal Casellario dei Pensionati) e all'acquisizione di cittadinanza italiana.

Un aspetto rilevante per il conteggio della popolazione è rappresentato dall'aver tenuto conto della definizione di dimora abituale del Regolamento 1260/2013 dell'Unione Europea. Ai fini della "*usually resident population*", il Regolamento europeo specifica come riferimento temporale la permanenza in un luogo per almeno 12 mesi, prima della data di riferimento, oppure almeno 12 mesi dopo la stessa data, come espressione dell'intenzione di vivere stabilmente nello stesso territorio. Sulla base di queste indicazioni, per la definizione dei segnali di vita è stato scelto il periodo che va dal primo gennaio 2021 al 31 dicembre 2022 che, per il conteggio di popolazione in oggetto, rappresenta la data di riferimento del Censimento permanente della popolazione 2022.

L'osservazione longitudinale dei segnali diretti su due anni consente di cogliere specifici profili di presenza degli individui sul territorio (o pattern di continuità nella presenza, attività dello *Step 3* riportata nel Prospetto 3). Questi profili della presenza in Italia in alcuni casi consentono di identificare chiaramente la dimora abituale in Italia, mentre in altri i "segnali di vita" nel nostro Paese sono di bassa intensità, oppure il segnale che si delinea dal 1° gennaio 2021 al 31 dicembre 2022 è quello dei lavoratori stagionali, ovvero di profili che non possono essere associati alla popolazione dimorante abitualmente.

A questo riguardo per dare una idea precisa su come vengono classificati i segnali di lavoro e studio rispetto alla dimora abituale in Italia, il Prospetto 4 può risultare un utile schema di rappresentazione nel configurare i segnali diretti di presenza in Italia. Ad esempio, i profili in blu scuro da 1 a 4 totalizzano almeno 12 mesi di segnali di presenza dal 2021 al 2022, anche se i profili non sono sempre continui per tutto il periodo di



osservazione, come nel caso 3 e 4 dello schema. Invece, i segnali dal numero 5 in poi non solo mostrano profili di discontinuità ma complessivamente non totalizzano, nel corso dei 24 mesi di osservazione, “almeno” 12 mesi di segnali diretti di studio e/o di lavoro in Italia. Questi segnali si configurano con una debole intensità della presenza (profili 5 e 6), oppure in alcuni casi come segnali non utili rispetto alla determinazione della dimora abituale in Italia (profili da 7 a 11). Anche se i profili 10 e 11 potrebbero condurre, nel censimento successivo, alla individuazione di segnali forti di presenza (più o meno equivalenti ai casi 2 e 4)

**PROSPETTO 4 - CONFIGURAZIONE E CLASSIFICAZIONE DEI SEGNALI DIRETTI DI STUDIO E DI LAVORO SECONDO I PROFILI DI CONTINUITÀ DELLA DIMORA ABITUALE IN ITALIA DAL 1.1.2021 AL 31.12.2022**

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 2021-2022
																									2	Continuo, prevalentemente nel 2022
																									3	Continuo, oltre dicembre 2021
																									4	Con pause ma almeno un anno
																									5	Stagionale
																									6	Interrotto sia nel 2021 che nel 2022
																									7	Solo a dicembre 2021
																									8	Solo nei primi mesi del 2021
																									9	Interrotto e solo nel 2021
																									10	Solo negli ultimi mesi del 2022
																									11	Interrotto e solo nel 2022

Fonte: Istat, 2023

Ogni segnale è da associare a uno specifico individuo e a una determinata localizzazione territoriale, secondo quanto è mostrato nel Prospetto 5: ad esempio, se nel periodo considerato per un individuo con codice “0000018” viene rilevato un record in un archivio di lavoro e un altro in una fonte relativa allo studio, entrambi però localizzati nel comune di Agliè, disporremo di un unico segnale localizzato in quel comune specifico. Questo segnale, però, viene contrassegnato sia da un attributo che permette di tracciare l’individuo in entrambi gli archivi di partenza, sia da un attributo relativo alla durata della presenza espressa in termini di attività lavorativa e di studio. Pertanto, la struttura informativa dell’algoritmo che elabora il segnale diretto di lavoro o studio del prodotto integrato delle fonti amministrative si configura con una stringa che sintetizza, in rosso nella terza colonna, un segnale di lavoro nella prima posizione e un segnale di studio nella posizione 9, mentre la classificazione del profilo di continuità è riportato, sempre con una stringa, nell’ultima colonna del prospetto, dove per il caso specifico, il segnale diretto di presenza nella fonte di lavoro è presente per tutti i mesi.

**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}
Esempio	Individuo 18 di Agliè	UniEmens (pos.1)+Università (pos.9)	Tempo indeterminato	Presente tutti i mesi
Dati di esempio	001-001-0000018	1000000010	0----1-----	11111111111111111111

Fonte: Istat, 2023





Riprendendo il flusso di lavoro del Prospetto 3, dopo che si è proceduto con la riduzione dei segnali di lavoro e studio, e degli altri segnali diretti sui singoli individui ed è stato determinato il Comune prevalente dove viene svolta l'attività di studio o di lavoro, si prosegue con l'integrazione degli individui con segnali di vita diretti con il Registro di popolazione RBI. In particolare, nello *Step 4* si effettua lo scarico del RBI al 31.12.2021 e al 31.12.2022 e, per singolo individuo, si realizza il confronto con i segnali di vita diretti che consente di determinare, per ciascun Comune, tutti gli individui con dimora abituale in Italia non residenti nel RBI e gli individui residenti nel RBI senza segnali di vita diretti. Il passo successivo (*Step 5*) consiste nello scarico delle fonti fiscali e nel successivo confronto dei segnali di vita indiretti dei "familiari a carico" e dei titolari di una automobile e di un'unità immobiliare con gli individui residenti nel RBI che erano risultati senza segnali di vita diretti dello *Step 4*.

Infine, l'ultimo passaggio consiste quindi nel determinare, nello *Step 6*, la potenziale sovra copertura anagrafica di RBI che è composta dagli individui privi di segnali diretti e indiretti da tutte le fonti considerate che risultano residenti nel RBI. Su questo contingente di popolazione si procederà, però, con la lavorazione di una ulteriore fase dove viene utilizzata l'informazione sulle famiglie anagrafiche identificate nelle LAC, in modo da riconoscere le persone registrate come "coniugi", all'interno della stessa famiglia, di intestatari della famiglia che hanno segnali di lavoro e studio o altri segnali di vita diretti. Questo contingente di popolazione in quanto privo di segnali diretti e indiretti delle fonti prese in esame finirebbe, altrimenti, nell'insieme degli individui classificati come sovra copertura di RBI.

Ciò ha indotto a non limitare l'attenzione dei segnali di vita solo ed esclusivamente alle fonti amministrative disponibili, ma di valorizzare la ricchezza degli archivi amministrativi, utilizzando contemporaneamente i dati delle indagini censuarie. I ricercatori dell'Istat hanno operato secondo un approccio di tipo *Knowledge Discovery from Databases*; si sono avvalsi, in definitiva, di un processo strutturato e iterativo, in cui una parte delle variabili da analizzare sono costruite in itinere, man mano che si procede con la lavorazione dei segnali amministrativi. Questo tipo di approccio consente, all'interno di una grande mole di dati amministrativi e di indagine, di identificare criteri e profili utili per il trattamento di specifiche sottopopolazioni.

### **La definizione finale degli aggregati di sovra e sotto copertura di RBI al 31.12.2022**

Nei paragrafi precedenti si è visto che il processo di integrazione delle fonti è stato utilizzato per il conteggio del 2022 al fine di identificare due sotto gruppi di popolazione che consentono di correggere il RBI al 31.12.2022:

- 1) gli individui dimoranti abitualmente in Italia perché mostrano segnali di vita diretti" di almeno un anno dagli archivi amministrativi; questi individui, dato che non sono presenti nel RBI come residenti al 31.12.2022, rappresentano la sotto copertura delle anagrafi comunali alla stessa data;
- 2) gli individui residenti nel RBI al 31.12.2022 senza segnali di vita diretti e indiretti negli archivi amministrativi, che rappresentano la sovra copertura delle anagrafi comunali.

L'impianto sviluppato si caratterizza per l'uso degli archivi amministrativi in modo "indipendente" rispetto alle informazioni anagrafiche comunali. Questo approccio richiama molto da vicino il carattere di "universalità" del Censimento tradizionale della popolazione: tutti i Comuni sono osservati nello stesso istante di tempo e sono sottoposti a tecniche standardizzate sia in riferimento alla metodologia utilizzata (i profili di continuità) che al trattamento dei dati individuali.

Inoltre, anche nel 2022, grazie alla disponibilità dei dati di indagine, e attraverso i risultati del modello a classi latenti studiato per il conteggio del 2021, e ad opportune analisi e valutazione da parte degli esperti tematici dell'Istat si è arrivati a identificare, a valle del processo di integrazione, alcuni criteri in base ai quali è stato possibile considerare anche i fattori di contesto che consentono di migliorare il processo di identificazione della sovra e sotto copertura di RBI.

Ad esempio, in riferimento alla sotto copertura, il Prospetto 6 mostra che, per tutti gli individui di cittadinanza straniera per i quali sono stati identificati i segnali di vita diretti, il solo permesso di soggiorno non rappresenta una condizione sufficiente ai fini della identificazione della sotto copertura; mentre se il segnale di vita diretto è localizzato in un Comune frontaliere tutti gli individui stranieri con tale segnale non sono considerati dimoranti abitualmente in Italia per le ovvie ragioni connesse ai movimenti frontalieri per lavoro o studio da parte di cittadini dei paesi esteri confinanti con l'Italia. In ogni caso, vale la pena considerare che se la fonte amministrativa di origine non avesse valorizzato il comune del domicilio fiscale,



il luogo del contratto di affitto, di lavoro o di studio non sarebbe possibile attribuire un Comune di dimora abituale agli individui classificati in sotto copertura.

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



Fonte: Istat, 2023

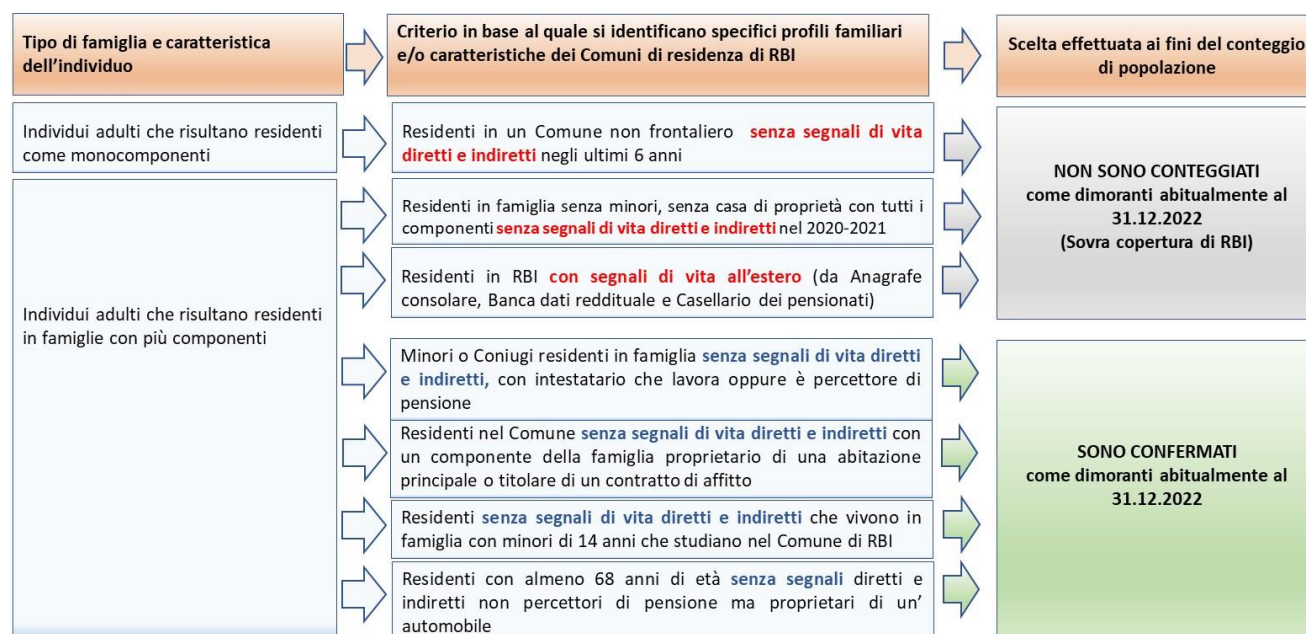
Un discorso analogo è stato fatto anche in riferimento agli individui senza segnali diretti e indiretti della sovra copertura di RBI. In particolare, a valle del processo di integrazione e grazie alle evidenze emerse dall'applicazione del modello sono stati definiti ulteriori criteri o profili di individui che, seguendo l'approccio iterativo e la logica gerarchica, hanno portato alla costruzione di nuove variabili a livello individuale o familiare, come si può evincere dal Prospetto 7. In base a questo ulteriore passaggio è stato possibile, in alcuni casi, rafforzare l'assenza del segnale di vita sul territorio già risultante dall'output del processo di integrazione e, quindi, avvalorare l'esito di sovra copertura degli individui residenti di RBI senza segnali diretti e indiretti; invece in altri casi la creazione di nuove variabili associate al contesto familiare degli individui, in prima istanza classificati come sovra copertura di RBI, ha reso possibile il loro recupero tra le persone abitualmente dimoranti in Italia e in quanto tali confermare la loro presenza nel RBI ai fini del conteggio.

Inoltre, è stata confermata la residenza nel RBI ai fini del conteggio di tutti gli individui dei comuni molto piccoli per i quali gli indicatori longitudinali di tenuta delle anagrafi comunali è risultata molto robusta. Indicazioni in tal senso sono emerse anche da numerose analisi esplorative condotte dai ricercatori dell'Istat nel corso degli ultimi 5 anni.

Sono stati anche confermati come residenti anche tutti gli individui che vivono nelle convivenze anagrafiche di tutti i Comuni. Questo criterio è stato definito a seguito di una attività di verifica effettuata dall'Istat nei primi mesi del 2022, nel corso della quale sono stati verificati, presso tutte le anagrafi comunali, sia gli indirizzi sia gli aggregati di popolazione corrispondenti precaricati dall'Istituto sul Sistema di Gestione della Rilevazione. Verifiche molto simili sono state effettuate anche per la popolazione dei grandi anziani (persone con almeno 98 anni di età) per i quali si è proceduto alla loro conferma della dimora abituale nel RBI.



**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, 2023

In sintesi, le persone individuate con i segnali di vita diretti del prospetto 6 e quelle individuate con i segnali di vita indiretti (i CONFERMATI del prospetto 7) rappresentano, in considerazione delle conferme e delle aggiunte derivate dall'individuazione della sotto copertura, gli aggregati fondamentali per il conteggio della popolazione.

### Attribuzione della cittadinanza e dello stato di nascita

Ai fini del conteggio della popolazione 2022, l'attribuzione della cittadinanza è una attività che 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) che vengono acquisite direttamente dalla Anagrafe Nazionale della Popolazione residente (ANPR). Ai fini della produzione del dato di popolazione la cittadinanza deve essere comunque sottoposta ad un processo di controllo e correzione dato che per alcuni individui si riscontrano valori anomali e comunque non in linea con la classificazione Istat dei paesi di cittadinanza che annualmente viene prodotta dall'Istituto. In particolare le criticità che vengono riscontrate nel dato di popolazione sono le seguenti:

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

Per gli individui con cittadinanza mancante si procede, come primo step, ad attribuire il valore l'informazione disponibile nel Registro Base degli Individui (RBI) dell'anno precedente. Tuttavia, al fine di tenere conto anche dei processi di naturalizzazione verificatisi da un anno all'altro, tutti gli individui con cittadinanza straniera vengono confrontati con i "Giuramenti di cittadinanza italiana" (per maggiore robustezza del dato vengono considerati i giuramenti di cittadinanza italiana degli ultimi 5 anni dall'archivio del Ministero dell'Interno) e per coloro che si abbinano con questa fonte viene assegnata la cittadinanza Italiana.

Un aggregato su cui è riposta molta attenzione da parte degli organismi internazionali (ad esempio, l'Alto Commissariato delle Nazioni Unite per i Rifugiati - UNHCR che effettua su questo aggregato un monitoraggio annuale) è quello degli 'Apolidi' che nelle LAC risulta con una numerosità particolarmente sovrastimata. Nella procedura di correzione sono estratti tutti gli individui 'Apolidi' che risultano avere come relazione di parentela 'figlio dell'intestatario e/o del coniuge/convivente' rispetto all'intestatario di scheda anagrafica



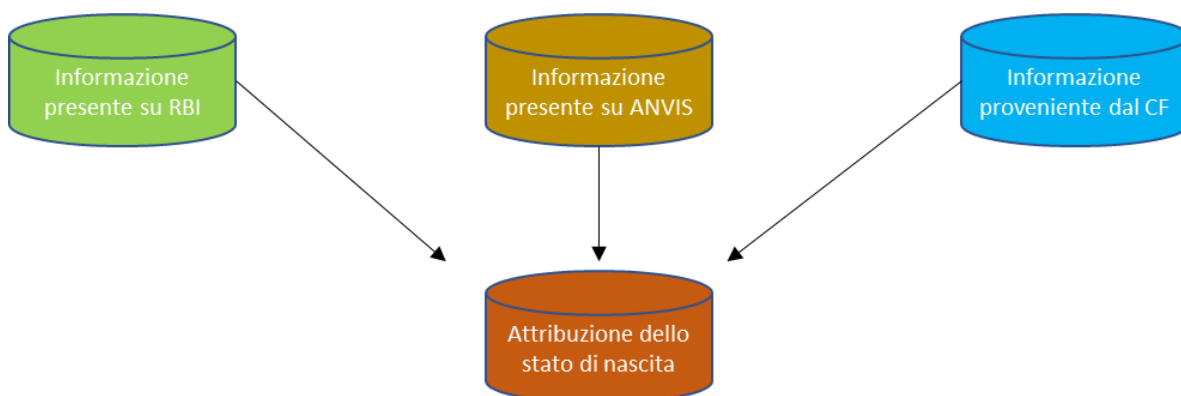
e successivamente la cittadinanza di questi individui viene confrontata per l'appunto con la cittadinanza dell'intestatario di scheda anagrafica e del coniuge/convivente. Nel caso in cui né l'intestatario né il coniuge/convivente risultino avere cittadinanza 'Apolide', l'individuo 'eredita' la stessa cittadinanza dell'intestatario di scheda anagrafica.

Le persone con cittadinanza 'In via di definizione' sono, nella quasi totalità dei casi, individui nati nel corso dell'anno di riferimento, pertanto per assegnare la cittadinanza si effettuano due passaggi: il primo è lo stesso della procedura usata per gli 'Apolidi'; invece, il secondo step incrocia i residui del primo passaggio con i dati di cittadinanza provenienti dalla fonte dalla rilevazione 'Iscritti in anagrafe per nascita' che consente di attribuire la cittadinanza anche agli individui che, nati nel corso dell'anno di riferimento, non risultino avere una relazione di parentela 'figlio dell'intestatario e/o del coniuge/convivente'.

Un ulteriore aggregato che viene trattato ai fini dell'attribuzione della cittadinanza è quello degli individui che emergono come residenti in base ai segnali di vita "forti" del processo di integrazione tra il RBI e le altre fonti amministrative, vale a dire la componente di popolazione in sotto copertura rispetto alle anagrafi comunali. Questi individui subiscono diversi passaggi nella procedura di assegnazione della cittadinanza: in primo luogo viene effettuata, attraverso il codice identificativo unico, una ricerca se l'informazione sulla cittadinanza è disponibile nell'archivio dei permessi di soggiorno e, in seconda istanza, si passa ad una verifica negli altri archivi disponibili (ad esempio, nell'archivio dei lavoratori domestici da parte della fonte INPS). Al termine di questi passaggi, si verifica solitamente che poche decine di individui risultano ancora avere il valore mancante o non valido e, per questi casi residuali, si procede all'assegnazione della cittadinanza utilizzando direttamente 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 interna al Registro Base degli Individui (RBI) e agli individui che avevano l'informazione sulla provincia di nascita valorizzata è stato automaticamente posto lo stato di nascita 'Italia'. Inoltre, i casi anomali sono stati confrontati con i codici della classificazione di EUROSTAT e in caso di mancata conformità sono stati corretti. Nei casi in cui dalle altre informazioni presenti su RBI non è stato possibile attribuire nessun paese di nascita, i valori anomali sono stati sbiancati, mentre i codici per i quali si è riusciti a ricostruire il paese di nascita compatibile con la classificazione di EUROSTAT si è provveduto ad una ricodifica.

Eseguita questa prima fase di correzione sul RBI, l'informazione dello stato di nascita è stato confrontata con il medesimo valore proveniente dall'Anagrafe Virtuale Statistica (ANVIS) e con le informazioni provenienti dal codice Belfiore del codice fiscale.



Nei casi in cui le tre fonti abbiano informazioni discordanti, è stata attribuita l'informazione che aveva l'occorrenza valida maggiore. Invece, per i casi in cui tutte e tre le fonti avessero avuto valori discordanti l'informazione proveniente da RBI è stata assunta come quella corretta.

In questo modo è stato possibile costruire una variabile temporanea che però non tiene in considerazione la data della scissione/unione degli stati ma solo il codice dello stato che risultava valido per entrambe le classificazioni Istat e Eurostat sui codici degli stati di nascita.

Inoltre, ai fini del rilascio dei dati censuari per EUROSTAT, gli stati del luogo di nascita devono essere aggiornati ai confini dell'anno di riferimento del censimento. Pertanto tutti i codici riguardanti stati non



più esistenti devono essere modificati (URSS, Cecoslovacchia, Jugoslavia etc.). Per i paesi esteri che si sono aggregati, l'attualizzazione non crea problemi, mentre per tutti gli stati che si sono scissi si segue una procedura che si compone in due fasi. Nella prima fase, per gli individui nati in uno dei paesi per i quali si è verificata una scissione si controlla se, in una delle fonti (RBI, ANVIS o Codice Fiscale) usate nella fase precedente, è presente un codice stato di nascita congruente rispetto ai confini dello stato non più esistente;; in caso affermativo gli viene attribuito lo stato di nascita trovato in una delle tre fonti; in caso contrario se in nessuna delle fonti è presente uno stato congruente secondo i confini attuali si passa alla seconda fase che utilizza una procedura probabilistica che imputa all'individuo uno dei paesi di nascita congruente con lo stato non più esistente.

In questa fase vengono calcolate tutte le percentuali parziali delle distribuzioni di frequenza a livello comunale degli stati afferenti allo stato non più esistente (ad esempio, per la Cecoslovacchia vengono calcolate le percentuali degli individui nati in Repubblica Ceca e nella Slovacchia rispetto al totale dei due paesi). La percentuale così calcolata diventa la probabilità che ad un individuo venga attribuito uno degli stati esteri attuali derivato dalla nazione non più esistente. Invece, se non è possibile calcolare le distribuzioni di frequenza a livello comunale, la probabilità viene calcolata attraverso il medesimo processo che tiene conto, però, della distribuzione dello stato di nascita a livello nazionale. Infine per ogni individuo viene estratto un numero pseudocasuale (che varia tra 0 e 1) grazie al quale è possibile attribuire lo stato di nascita.

A titolo esemplificativo si riporta la tabella con i tre stati non più esistenti che impattano maggiormente la ricodifica probabilistica.

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