

# LA STIMA DELLA CITTADINANZA NEL REGISTRO BASE DEGLI INDIVIDUI TRAMITE INTEGRAZIONE DI FONTI AMMINISTRATIVE.



## **LA STIMA DELLA CITTADINANZA NEL REGISTRO BASE DEGLI INDIVIDUI TRAMITE INTEGRAZIONE DI FONTI AMMINISTRATIVE**

ISBN 978-88-458-1999-5

© 2019

Istituto nazionale di statistica  
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti  
pubblicati sono soggetti alla licenza  
Creative Commons - Attribuzione - versione 3.0.  
<https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire,  
trasmettere e adattare liberamente dati e analisi  
dell'Istituto nazionale di statistica, anche a scopi  
commerciali, a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),  
marchi registrati e altri contenuti di proprietà di  
terzi appartengono ai rispettivi proprietari e non  
possono essere riprodotti senza il loro consenso.



# INDICE

	Pag.
<b>Introduzione</b>	5
<b>1. Analisi del contesto informativo</b>	7
1.1 Identificazione delle esigenze informative e definizione degli obiettivi	7
1.1.1 <i>Cittadinanza: definizione e classificazione Istat</i>	7
1.1.2 <i>Classificazione Istat degli Stati esteri</i>	8
1.1.3 <i>Ricognizione degli archivi amministrativi contenenti informazioni sulla Cittadinanza</i>	8
1.2 Riclassificazione delle variabili: progetto del processo di allineamento delle classificazioni	10
<b>2. Analisi di qualità delle fonti</b>	13
2.1 Analisi di qualità sulle fonti individuate: analisi della copertura, tempestività e concordanza	13
2.1.1 <i>Copertura</i>	13
2.1.2 <i>Tempestività e concordanza</i>	16
2.1.3 <i>Analisi della concordanza degli archivi amministrativi MIUR, INPS, INAIL, CCIAA</i>	17
2.1.4 <i>Analisi della concordanza tra “Ministero dell’Interno - Acquisizione e reiezione della cittadinanza italiana” e LAC</i>	21
2.1.5 <i>Confronto tra paese di nascita in “Anagrafe Tributaria” (AT) e cittadinanza in LAC nel 2015</i>	23
<b>3. Disegno, sperimentazione della procedura di stima della cittadinanza</b>	25
3.1 Specifiche statistiche ed architettura del processo di stima della cittadinanza in RBI	25
3.1.1 <i>Sperimentazione sui dati di Persons &amp; Places</i>	25
3.1.2 <i>Architettura del processo di stima della cittadinanza per gli individui in RBI</i>	27
<b>4. Valutazione della procedura di stima della cittadinanza</b>	31
4.1 Valutazione della procedura di stima della cittadinanza in RBI	31
4.1.1 <i>Analisi dei risultati della procedura applicata ai dati del Registro Base degli Individui</i>	31
4.1.2 <i>Analisi di confronto della cittadinanza nel Registro Base degli Individui con quella di ANVIS</i>	34

4.1.3 <i>Confronto della cittadinanza in RBI con dati sulla cittadinanza della rilevazione "Popolazione residente comunale per sesso, anno di nascita e stato civile"</i>	36
4.1.4 <i>Confronto tra il sottoinsieme di RBI la cui cittadinanza è stata desunta da LAC (01/01/2018) con dati sulla cittadinanza nella rilevazione "Popolazione residente comunale per sesso, anno di nascita e stato civile"</i>	37
<b>Riferimenti bibliografici</b>	39

## INTRODUZIONE<sup>1</sup>

Il sistema di produzione delle statistiche in Istat sta cambiando in favore di un utilizzo sempre maggiore di informazioni amministrative integrate e finalizzate alla costruzione di un sistema di registri che copra i vari universi tematici. Nel Sistema Integrato dei Registri (SIR) (*cf.*, Garofalo, 2016, Di Zio, Falorsi, Fortini, 2018, Radini *et al.*, 2017) sono previsti dei registri di base, ovvero collezioni di osservazioni per le quali è possibile individuare la popolazione statistica di riferimento e informazioni fondamentali denominate variabili *core*. Il Registro Base degli Individui (RBI) è un insieme di individui ottenuti dall'integrazione di varie fonti amministrative. Al suo interno è possibile individuare il sottoinsieme delle unità residenti che costituisce la base del censimento. Per le unità del registro sono state ricostruite le principali informazioni relative alla loro determinazione anagrafica (luogo di nascita, data di nascita, genere). A queste si aggiungono le variabili di localizzazione, in particolare relative al luogo di residenza, e la cittadinanza che vengono ricalcolate annualmente e storicizzate. Le variabili associate alle unità del registro saranno il riferimento per tutte le indagini che le coinvolgono direttamente, come quantità da stimare o come variabili di dominio, o indirettamente, ovvero come informazione ausiliaria. Queste variabili sono caratterizzate da una elevata disponibilità di informazione amministrativa che rende affidabile la loro previsione/ricostruzione a livello di singola unità (livello micro).

Questa pubblicazione presenta i risultati del lavoro svolto nell'ambito del progetto di stima della *cittadinanza* per il Registro Base degli Individui. In particolare, vengono descritte le analisi delle fonti amministrative che includono informazioni sulla cittadinanza, la procedura di stima, la realizzazione dei test e le analisi dei risultati finalizzati alla validazione della procedura proposta.

Il processo di analisi delle fonti è stato condotto sui metadati del Sistema Integrato dei Microdati (SIM). Il test sulla ricostruzione della variabile è stato condotto sui dati della base integrata di microdati Persons & Places (P&P) (*cf.* Vivio *et al.*, 2014, Vivio 2016) che al momento della sperimentazione possedeva le migliori caratteristiche per approssimare il RBI. Dalle analisi condotte sono scaturite le specifiche per la realizzazione dell'algoritmo di ricostruzione della cittadinanza basato esclusivamente su fonti amministrative. L'algoritmo è di tipo gerarchico, ovvero viene stimata la cittadinanza in RBI prendendo le informazioni sulla cittadinanza dalle fonti secondo una opportuna sequenza. La sequenza viene stabilita in base alla qualità ed alla tempestività delle fonti stesse. L'algoritmo è stato implementato in RBI con riferimento a tutte le unità in esso presenti, residenti e non residenti.

La validazione della procedura viene condotta confrontando i risultati ottenuti in RBI con quelli relativi ai principali aggregati pubblicati ufficialmente dall'Istat e con gli esiti della procedura di ricostruzione della cittadinanza implementata nell'Anagrafe Virtuale Statistica (ANVIS<sup>2</sup>) – essenzialmente basata sulle Liste Anagrafiche Comunali (LAC), sul Censimento 2011 e sul codice fiscale. Inoltre, è stata condotta un'analisi di robustezza dell'informazione prodotta rispetto al tempo di rilascio del Registro Base degli Individui. In particolare è stata confrontata la distribuzione della cittadinanza ricostruita a 6 mesi dalla data  $t$  di riferimento del registro base degli individui (ovvero RBI al tempo  $t + 6$ ) con quella ottenuta utilizzando

<sup>1</sup> Il volume è stato curato da Marco Di Zio e Roberta Vivio.

<sup>2</sup> ANVIS è un archivio longitudinale che, a partire dai microdati del Censimento 2011, viene aggiornato annualmente attraverso i flussi demografici relativi a nascite, decessi e trasferimenti di residenza.

tutte le fonti amministrative che sconta un ritardo temporale di circa 24 mesi (RBI al tempo  $t+24$ ), tempo necessario per avere a disposizione tutte le fonti amministrative utilizzate per la ricostruzione della cittadinanza. Questa analisi è importante per capire se è possibile avere una stima preliminare di buona qualità della variabile cittadinanza presente in RBI nel momento in cui questo venga utilizzato come lista di campionamento per il censimento permanente (cfr. Falorsi *et al.*, 2018). Tale lista deve infatti essere disponibile con elevata tempestività. I risultati hanno evidenziato che, per la popolazione residente, la cittadinanza ricostruita a  $t + 6$  mesi ha un buon livello di qualità e dunque può essere utilizzata per la lista di campionamento.

Ricapitolando, il beneficio che si ha utilizzando un elevato numero di fonti amministrative alternative alla LAC per la ricostruzione della cittadinanza risiede fondamentalmente in un raffinamento della stima di tale variabile. Infatti, questa procedura permette di aggiornare e correggere alcuni errori che possono essere presenti nella informazione disponibile nelle Liste Anagrafiche. Grazie all'utilizzo di fonti riguardanti l'acquisizione della cittadinanza si può avere un aggiornamento più tempestivo delle informazioni presenti in Anagrafe. Inoltre le fonti amministrative alternative alla LAC sono un fondamentale ausilio per le eventuali cadute di informazioni nelle LAC stesse, poiché queste possono avere dei valori mancanti nella cittadinanza, oppure possono essere affette da problemi tecnici legati per esempio alla trasmissione di LAC aggiornate.

Infine, nell'ottica di un registro degli individui che includa non solo le persone residenti, le informazioni delle fonti alternative alla LAC sono fondamentali perché tali unità sono generalmente non incluse nelle LAC.

Il lavoro è strutturato come segue. Il Capitolo 1 delinea le esigenze informative e gli obiettivi dello studio. In questo capitolo, vengono anche riportati gli esiti della ricognizione degli archivi amministrativi utilizzabili per la ricostruzione della cittadinanza. In questa prima parte dello scritto viene discusso anche il problema dell'integrazione delle classificazioni adottate nei differenti archivi amministrativi. Sottolineiamo come a tal fine sia stata sviluppata una procedura informatica, che viene descritta in questo paragrafo, atta alla riconciliazione automatica delle differenti classificazioni. Il Capitolo 2 è dedicato alle analisi di qualità delle fonti amministrative; in particolare viene analizzata la loro copertura, la loro concordanza rispetto alle LAC prese come fonte di riferimento e per ultimo viene analizzata la tempestività delle fonti. Il Capitolo 3 è dedicato alla descrizione della procedura di stima e dell'architettura necessaria per la sua realizzazione in fase di produzione. Nello stesso capitolo vengono riportati i risultati dell'applicazione di tale procedura ai dati relativi alla popolazione insistente in P&P, che – vale la pena di ricordare – è composta da residenti e non residenti. Infine, nel Capitolo 4 vengono riportati i confronti condotti con le informazioni sulla cittadinanza disponibili in Istat. La cittadinanza della popolazione residente in RBI viene confrontata con quella ufficialmente pubblicata dall'Istat e con la cittadinanza ricostruita in ANVIS. In questo capitolo viene inoltre analizzata la qualità della cittadinanza con riferimento a una versione anticipata del RBI. Tale analisi è finalizzata alla valutazione di un possibile utilizzo di questa informazione nella lista di campionamento per il censimento permanente.

## 1. ANALISI DEL CONTESTO INFORMATIVO<sup>1</sup>

### 1.1 Identificazione delle esigenze informative e definizione degli obiettivi

La cittadinanza è una delle variabili *core* che RBI ha il compito di produrre per alimentare il SIR con una informazione controllata e condivisa.

Nessuna delle fonti attualmente usate dall'Istat nelle statistiche ufficiali ha una copertura sufficiente a coprire l'universo di tutti gli individui in RBI che contiene, oltre ai cittadini residenti, anche i non residenti presenti nel SIM\_Individui<sup>2</sup> (la base dati integrata costituita dai microdati riguardanti gli individui provenienti da archivi di natura prevalentemente amministrativa).

L'obiettivo è la definizione di una procedura che consenta di stimare la cittadinanza per gli individui (residenti e non residenti) presenti in RBI.

#### 1.1.1 Cittadinanza: definizione e classificazione Istat

Con il termine *cittadinanza*<sup>3</sup> si indica la condizione giuridica di un individuo rispetto ad uno Stato che viene riconosciuta alla fine di un iter amministrativo regolato da normative nazionali. In Italia la concessione della Cittadinanza è di competenza del Ministero dell'Interno.

Per quanto riguarda le definizioni, quella del Ministero dell'Interno recita: *“Cittadinanza è un termine di diritto e indica il rapporto giuridico tra un individuo e lo Stato. In particolare è uno status denominato civitatis, al quale l'ordinamento giuridico ricollega la pienezza dei diritti civili e politici<sup>4</sup>.”*

L'Istat adotta la seguente: *“Cittadinanza è un vincolo di appartenenza a uno stato, richiesto e documentato per il godimento di diritti e l'assoggettamento a particolari oneri<sup>5</sup>”.*

Spesso si usa il termine “nazionalità” come sinonimo di cittadinanza, anche se la nazionalità si riferisce al senso di appartenenza a una nazione per lingua, cultura, tradizione, religione, storia e ha un significato sociologico.

La cittadinanza può cambiare nell'arco della vita ed è un carattere essenziale per la identificazione di ciascun individuo. Anche per questa ragione fa parte dell'insieme di dati individuali riportati negli archivi amministrativi previdenziali, economici, demografi-

1 Hanno contribuito alla redazione del capitolo: Roberta Vivio (paragrafi 1.1, 1.1.1, 1.1.2. e 1.1.3) e Paolo Francescangeli (paragrafo 1.2).

2 Attualmente la cittadinanza utilizzata nelle statistiche ufficiali (in particolare quelle sugli stranieri che sono raccolte nel portale *Immigrati e nuovi cittadini*), è acquisita dall'Anagrafe della popolazione residente (e tutti gli archivi ad essa collegati sul movimento della popolazione) e dall'archivio dei Permessi di soggiorno gestiti dal Ministero dell'Interno.

3 La cittadinanza italiana è attualmente regolata dalla legge 5 febbraio 1992, n. 91 (e relativi regolamenti di esecuzione: in particolare il DPR 12 ottobre 1993, n. 572 e il DPR 18 aprile 1994, n. 362).

4 <http://www.interno.gov.it/it/temi/cittadinanza-e-altri-diritti-civili/cittadinanza>.

5 <https://www.istat.it/it/metodi-e-strumenti/glossario>.

ci, ecc. (Prospetto 1.1 e Tavola 1.1) in aggiunta alle informazioni tematiche di pertinenza dell'archivio stesso.

### 1.1.2 Classificazione Istat degli Stati esteri

L'Istat utilizza la Classificazione degli Stati esteri (Istat, 2018) sia per classificare la cittadinanza, sia gli stati esteri. In essa si attribuisce a ogni Stato esistente un codice di tre cifre. *“L'elenco degli Stati con i relativi codici viene aggiornato a seguito delle variazioni che avvengono nella configurazione geopolitica mondiale. Determinano aggiornamenti della classificazione eventi quali ad esempio la nascita di un nuovo Stato, la cessazione di uno Stato preesistente, ecc. Gli Stati sono raggruppati per continente e per area geopolitica di appartenenza. Nel caso dell'Europa, in particolare, si distingue tra Paesi membri dell'Unione Europea e altri Paesi europei.”*

### 1.1.3 Ricognizione degli archivi amministrativi contenenti informazioni sulla Cittadinanza

Si è analizzato il patrimonio informativo a disposizione, costituito dagli archivi amministrativi a copertura nazionale contenuti nel SIM, per verificare la presenza di informazioni sulla cittadinanza.

Una volta verificato questo presupposto si è passati all'analisi della quantità e qualità dei dati.

Sono state esaminate le criticità emerse<sup>6</sup> e predisposte delle soluzioni. Si è condotto un test ed ingegnerizzata la procedura.

La ricognizione sugli archivi ha portato all'individuazione di 34 fonti (riportate nel prospetto 1.1 insieme al dettaglio delle classificazioni adottate).

Le informazioni contenute sono distinguibili in due tipologie:

- dirette: cittadinanza, codice cittadinanza;
- indirette: codice stato estero, codice stato di provenienza, codice stato estero di nascita, codice stato estero di residenza, città di provenienza, nazionalità.

#### Prospetto 1.1 - Archivi che riportano informazioni dirette sulla cittadinanza (cittadinanza, codice cittadinanza, cittadinanza del lavoratore - anno 2017)

ARCHIVI/REGISTRI CONTENENTI CITTADINANZA, CODICE CITTADINANZA, CITTADINANZA DEL LAVORATORE
INPS - Archivio anagrafico lavoratori extracomunitari
INPS - EMens
LAC - Liste anagrafiche comunali
MIUR - Anagrafe degli studenti
INAIL - Lavoratori parasubordinati
INAIL - Lavoratori interinali
INAIL - Denunce Nominative Assicurate
CCIAA - Registro delle imprese e delle unità locali
CCIAA - Archivio Persone d'Impresa - Iscrizione della società nel Registro delle Imprese
Ministero degli Interni - Permessi di soggiorno
CCIAA - Archivio soci delle società di capitale (Banca Dati Soci)
MAE - Ministero Affari Esteri - Archivio Censimento Italiani all'Estero
MAE - Ministero Affari Esteri - Archivio Italiani all'Estero
MIUR - Archivio degli iscritti e delle iscrizioni universitarie
MIUR - Archivio delle lauree e dei laureati
MIUR - Anagrafe degli studenti - Esiti scolastici
MIUR - Dottorati di ricerca
Ministero del Lavoro e delle Politiche Sociali - Comunicazioni Obbligatorie
Ministero dell'Interno - Acquisizione e reiezione della cittadinanza italiana

<sup>6</sup> In questo test non è stata presa in considerazione la doppia cittadinanza.

## 1. Analisi del contesto informativo

Per ogni archivio sono stati esaminati i seguenti elementi: universo di riferimento, unità di rilevazione, dimensione, aggiornamento. Nella tavola 1.1 vengono riportate le seguenti informazioni:

**Tavola 1.1 - Archivi/Registri per contenuto su CITTADINANZA, luogo di nascita, nazionalità. Numero di occorrenze**

CLASSIFICAZIONE/ARCHIVIO	Codice cittadinanza	Cittadinanza	Cittadinanza del lavoratore	Codice stato estero	Codice stato di provenienza	Codice stato estero di nascita	Codice stato estero di residenza	Città di provenienza	Flag cittadinanza italiana	Nazionalità del lavoratore	Nazionalità estera	Nazionalità italiana	Totale
<b>PAESISTAT</b>	<b>4</b>						<b>1</b>						<b>5</b>
LAC - Liste anagrafiche comunali	1												1
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	2							1					3
MIUR - Archivio delle lauree e dei laureati	1												1
<b>NAZIOINPS</b>										<b>1</b>			<b>1</b>
INPS - Rapporti di lavoro domestico										1			1
<b>PAESIASTAT</b>		<b>5</b>											<b>5</b>
MIUR - Anagrafe degli studenti - Bolzano		2											2
MIUR - Anagrafe degli studenti - Esiti scolastici		3											3
<b>PAESIMAE</b>	<b>1</b>				<b>1</b>	<b>1</b>							<b>3</b>
MAE - Archivio Censimento Italiani all'Estero	1				1	1							3
<b>PAESIMIUR</b>		<b>2</b>		<b>1</b>									<b>3</b>
MIUR - Anagrafe degli studenti		2		1									3
<b>S2STA</b>	<b>3</b>			<b>1</b>		<b>1</b>	<b>1</b>						<b>6</b>
CCIAA - Archivio Persone d'Impresa	1					1	1						3
CCIAA - Archivio soci delle società di capitale	1			1									2
CCIAA - Registro delle imprese e delle unità locali	1												1
<b>STATEST</b>		<b>2</b>											<b>2</b>
INAIL - Lavoratori interinali		1											1
INAIL - Lavoratori parasubordinati		1											1
<b>STATEST1</b>				<b>1</b>									<b>1</b>
MEF - Banca dati reddituale				1									1
<b>T_STATI COMUNI</b>		<b>1</b>											<b>1</b>
INAIL - Denunce Nominative Assicurate		1											1
<b>Non specificato nei metadati</b>	<b>3</b>	<b>12</b>	<b>4</b>	<b>2</b>	<b>1</b>		<b>3</b>	<b>2</b>	<b>4</b>		<b>1</b>	<b>1</b>	<b>33</b>
CCIAA - Archivio Persone d'Impresa - Iscrizione della società nel Registro delle Imprese		1											1
CCIAA - Archivio soci delle società di capitale (Banca Dati Soci)		1											1
CCIAA - Registro delle imprese e delle unità locali		1											1
INPS - Archivio anagrafico lavoratori extracomunitari		1											1
INPS - EMens	1												1
MAE - Ministero Affari Esteri - Archivio Censimento Italiani all'Estero								1					1
MAE - Ministero Affari Esteri - Archivio Italiani all'Estero	1				1			1					3
MEF - Archivio statistico delle dichiarazioni 770				1									1
MEF - Banca dati reddituale											1	1	2
Ministero dell'Interno - Acquisizione e reiezione della cittadinanza italiana		3											3
Ministero dell'Interno - Permessi di soggiorno (rilasci, validi in altri stati, variazione)		3											3
Ministero del Lavoro e delle Politiche Sociali - Comunicazioni Obbligatorie			4										4
MIUR - Anagrafe degli studenti				1						1			2
MIUR - Anagrafe degli studenti - Esiti scolastici										3			3
MIUR - Archivio degli iscritti e delle iscrizioni universitarie								1					1
MIUR - Archivio delle lauree e dei laureati	1							2					3
MIUR - Dottorati di ricerca		2											2
<b>Totale</b>	<b>10</b>	<b>22</b>	<b>4</b>	<b>5</b>	<b>2</b>	<b>2</b>	<b>5</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>59</b>

- Per la Cittadinanza: tipo\_dato (codice, descrizione), classificazione utilizzata, grado di pertinenza;
- Mappatura delle Classificazioni negli archivi con riferimento alla classificazione standard Istat;
- Analisi di copertura e qualità della “Cittadinanza” nelle diverse fonti.

## 1.2 Riclassificazione delle variabili: progetto del processo di allineamento delle classificazioni

All'interno del sistema dei metadati coesistono dieci classificazioni diverse tra loro e questo rende necessario un passo di transcodifica per uniformare la cittadinanza alla classificazione ufficiale adottata dall'Istat (PAESISTAT).

Nel prospetto 1.2 si riportano a titolo di esempio, le classificazioni di Arabia Saudita, Australia, Repubblica popolare cinese secondo le diverse alternative.

**Prospetto 1.2 - Esempi di codifica di tre cittadinanze secondo diverse classificazioni**

CLASSIFICAZIONE DEL PAESE DI CITTADINANZA	Arabia Saudita		Australia	Repubblica Popolare Cinese	
	PAESISTAT	302		701	314
PAESI_MIUR	501	724 ARABIA MERIDIONALE FEDERAZIONE (fino al 31/12/1975)	725 ARABIA MERIDIONALE PROTETTORATO (fino al 31/12/1975)	601	524
PAESI_ASTAT	501		601	524	
PAESI_MAE	301		302	307	
STATEST	302		701	314	
S2STAT	SA		AUS	RC	
NAZIOINPS	00484		00678	02133 CINA	
STATEST1	005		007	016 CINA	
T_STATI_COMMUNI	302		701	314 CINA	
CATASTO	Z203		Z700	Z210	

Come premesso, il processo di definizione della variabile cittadinanza in RBI ha richiesto l'integrazione di più fonti amministrative eterogenee rispetto alla classificazione.

Questo ha reso necessario un pretrattamento secondo due funzioni distinte ma strettamente correlate:

1. il riconoscimento dei codici o delle stringhe alfabetiche presenti;
2. l'armonizzazione delle classificazioni per riportarle a una classificazione di riferimento.

Il sistema di riconoscimento delle singole modalità delle classificazioni di cittadinanze si basa sull'utilizzo di funzioni di confronto delle stringhe alfabetiche. Il procedimento che è stato implementato (Funzione CHECK\_PAESE\_NASCITA) esegue il confronto tra due archivi di stringhe alfabetiche generiche, di cui una contiene la classificazione di riferimento, mentre l'altra la classificazione di input da armonizzare. Questa funzione di riconoscimento permette di risolvere anche "stringhe" complesse ed è generalizzabile a una serie di altre aree tematiche, utilizzando uno o più riferimenti tematici appositi. Nel caso della cittadinanza è stata utilizzata come classificazione di riferimento *PAESISTAT*, archiviata nel Sistema Unico dei Metadati - SUM, che riporta anche la serie storica delle nomenclature dei diversi paesi esteri.

Questa classificazione più ampia consente di risolvere i casi di cittadinanze non più esistenti ma ancora citate negli archivi (es. Jugoslavia, Unione delle Repubbliche Socialiste Sovietiche, ecc.).

L'input per la procedura è l'insieme dei dati ottenuto dall'unione di tutte le possibili dizioni e codici contenuti nel Sistema Integrato dei Microdati - SIM.

L'output del confronto individuale tra i dati in tale data set e la classificazione Istat è dato dal codice corrispondente in tale classificazione.

Il processo di transcodifica e armonizzazione della cittadinanza negli archivi è composto da diversi passi come dettagliato di seguito.

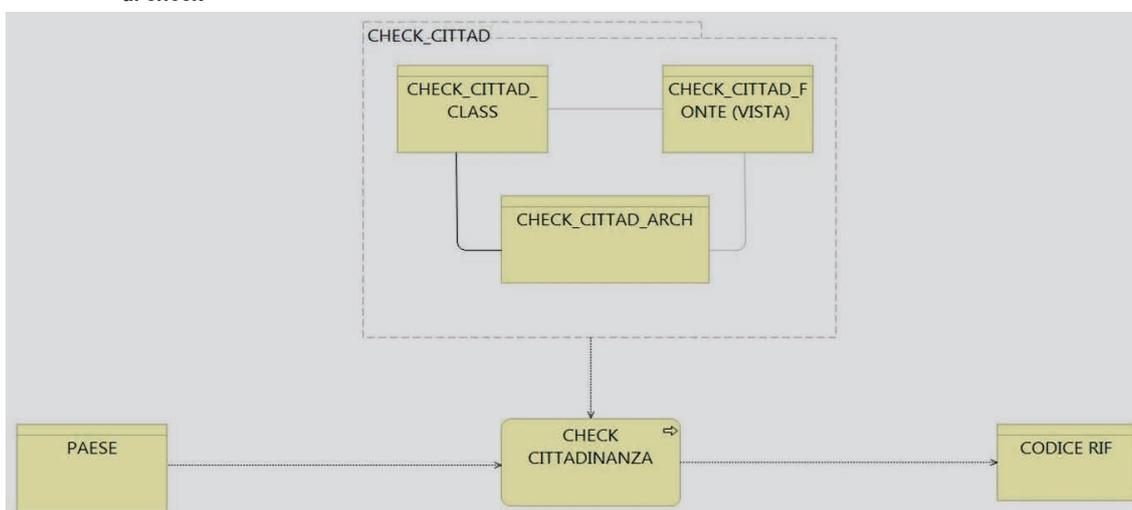
In particolare l'archivio principale che consente l'armonizzazione di tutte le modalità delle diverse classificazioni utilizzate dalle fonti, detto CHECK\_CITTAD, contiene per ogni modalità di classificazione: la fonte, il tipo di classificazione e il tracciato SIM dell'archivio.

L'informazione è organizzata in una relazione ternaria (Figura 1.1) di tre entità:

- la FONTE amministrativa, ossia il codice che la identifica e il suo tracciato di fonte (CHECK\_CITTAD\_FONTE);
- l'ARCHIVIO delle classificazioni di fonte, ossia la classificazione a cui fa riferimento la singola fonte (CHECK\_CITTAD\_ARCH);
- la CLASSIFICAZIONE, ossia il dettaglio della classificazione usata dalla fonte (CHECK\_CITTAD\_CLASS) e la rispettiva voce di classificazione ufficiale Istat.

In questa struttura, i dati della tavola CHECK\_CITTAD rappresentano un archivio che consente di codificare tutte le fonti di input secondo la classificazione ufficiale Istat.

Figura 1.1 - Relazione ternaria - CHECK\_CITTAD divisa nelle tre relazioni: Fonte, classificazione di fonte e funzione di check



Per semplificare la funzione di transcodifica e armonizzazione delle fonti di input, è stata creata la funzione CHECK\_CITTADINANZA che prende in input i dati di una fonte, recupera in modo automatico dai metadati di fonte, ovvero il tracciato record e la codifica utilizzata dalla fonte, e associa a ogni record il corrispondente codice della classificazione di riferimento, sia nel caso in cui il dato di cittadinanza è espresso in modalità alfabetico, sia se il dato è espresso in forma alfanumerica.

## 2. ANALISI DI QUALITÀ DELLE FONTI<sup>1</sup>

### 2.1 Analisi di qualità sulle fonti individuate: analisi della copertura, tempestività e concordanza

Gli aspetti della qualità presi in considerazione per valutare la variabile cittadinanza nelle fonti riguardano la copertura, l'accuratezza della misurazione e la tempestività. Il test è stato condotto sui dati del 2014. Osserviamo che per gli archivi del Ministero dell'Interno relativi all'acquisizione e alla reiezione della cittadinanza italiana gli anni di riferimento sono stati il 2015 e 2016 poiché per tali fonti non era disponibile il 2014.

#### 2.1.1 Copertura

Il concetto di copertura di una fonte relativamente ad una variabile è collegato alla definizione della popolazione statistica di riferimento. In questo test si prende come popolazione di riferimento la Popolazione insistente, output della base integrata di microdati *Persons & Places* (Vivio *et al.*, 2014, Vivio 2016, Vivio *et al.* 2019).

Essa è un sottoinsieme di RBI. Infatti, la struttura di base del RBI è data dall'insieme degli individui identificati dal SIM (base dati integrata SIM\_Individui) che risultino *eleggibili*. Tale insieme è composto da tutti i soggetti (inclusi gli stranieri e i non residenti in Italia) che compaiono in almeno una delle fonti amministrative che SIM integra.

L'insieme degli individui eleggibili tratti da SIM\_Individui rappresenta l'insieme più ampio possibile su cui operare una selezione per l'identificazione delle popolazioni statistiche di interesse.

La popolazione insistente definita in P&P comprende:

- a livello nazionale gli individui residenti in Italia, gli individui non residenti in Italia (con domicilio fiscale) che lavorano o studiano in Italia. Esclude i residenti all'estero.
- a livello comunale comprende gli individui che risiedono nel comune senza segnali di lavoro o studio (statici), gli individui residenti nel comune con mobilità per studio o lavoro all'interno dello stesso comune di residenza, gli individui che non risiedono nel comune in entrata da un altro comune per studio o lavoro. Infine comprende i non residenti nel comune che sono presenti per motivi occasionali (es. turisti). Esclude i residenti/domiciliati nel comune che lavorano/studiano in altro comune.

Nella popolazione insistente sono identificabili tre strati di popolazione differenti per caratteristiche sociodemografiche ed economiche: studenti delle scuole primarie e secondarie, universitari, lavoratori (Tavola 2.1 e 2.2).

<sup>1</sup> Hanno contribuito alla redazione del capitolo: Marco Di Zio (paragrafi 2.1 e 2.1.2), Roberta Vivio (paragrafi 2.1.1 e 2.1.2), Stefania Di Domenico (paragrafi 2.1.3 e 2.1.5) e Sara Giavante (paragrafo 2.1.4).

Tavola 2.1 - Individui in P&amp;P2014 per cittadinanza e tipo individuo (valori assoluti)

INDIVIDUI ITALIANI E STRANIERI	Totale	Studenti	Universitari	Lavoratori nell'anno	Lavoratori dicembre
Individui italiani e stranieri iscritti in LAC (*)	60.676.993	7.230.555	1.668.418	24.139.943	22.463.201
- <i>Cittadino italiano</i>	55.029.867	6.559.191	1.595.205	21.401.687	20.074.604
- <i>Cittadino straniero residente</i>	4.971.382	556.436	44.760	2.240.549	1.942.466
- <i>Cittadinanza non disponibile</i>	675.744	114.928	28.453	497.707	446.131
Individui non iscritti in LAC	696.091	63.937	28.549	423.720	218.516
<b>Totale</b>	<b>61.373.084</b>	<b>7.294.492</b>	<b>1.696.967</b>	<b>24.563.663</b>	<b>22.681.717</b>

(\*) si fa riferimento al triennio 2012/2014

Tavola 2.2 - Individui in P&amp;P2014 per cittadinanza e tipo individuo (composizione percentuale)

INDIVIDUI ITALIANI E STRANIERI	Totale	Studenti	Universitari	Lavoratori nell'anno	Lavoratori dicembre
Individui italiani e stranieri iscritti in LAC (*)	98,9	99,1	98,3	98,3	99,0
- <i>Cittadino italiano</i>	89,7	89,9	94,0	87,2	88,5
- <i>Cittadino straniero residente</i>	8,1	7,6	2,6	9,1	8,5
- <i>Cittadinanza non disponibile</i>	1,1	1,6	1,7	2,0	2,0
Individui non iscritti in LAC	1,1	0,9	1,7	1,7	1,0
<b>Totale</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>

(\*) si fa riferimento al triennio 2012/2014

Come si evince dai dati, la sottocopertura della cittadinanza nella base dati interessa 1,372 milioni di individui, di cui l'1,1% è relativa ai residenti, e 1,1% ai non residenti o non presenti nelle LAC. Questo accade perché la fonte per la cittadinanza nella Popolazione Insistente al 2014 è soltanto la LAC, e possono verificarsi tre casi:

- l'individuo è iscritto in LAC e la cittadinanza è valorizzata in modo errato
- l'individuo è iscritto in LAC ma la cittadinanza in LAC non è valorizzata (675 mila casi)
- l'individuo non è in LAC e per la cittadinanza manca la fonte di acquisizione (e.g., straniero non residente ma con segnali di lavoro/studio in Italia: 696 mila casi).

E' soprattutto questo terzo segmento di popolazione che rende interessante la Popolazione Insistente ai fini del test sulla possibilità di individuare fonti alternative alla LAC come input per la cittadinanza. Negli strati di popolazione, l'incidenza maggiore degli stranieri è relativa agli studenti universitari, seguiti dai lavoratori.

Nella tavola 2.3 vengono riportati i livelli di copertura delle diverse fonti rispetto alla popolazione insistente.

Notiamo che alcuni strati sono coperti dalla fonte che principalmente li contiene: per gli studenti è l'“Archivio degli iscritti e delle iscrizioni universitarie” e l'“Anagrafe studenti” a coprire maggiormente la variabile; per i lavoratori INPS- EMens” e “INAIL - DNA”. Questo secondo archivio copre anche molto bene gli ‘Altri individui’, quelli che non sono studenti, né lavoratori, (85%).

Gli universitari stranieri - secondo la variabile cittadinanza riportata nell'archivio delle iscrizioni all'a.a. 2014/15 - sono l'8,5% degli iscritti. Gli individui appartenenti a

## 2. Analisi di qualità delle fonti

tale archivio sono presenti anche in altri archivi contenenti la variabile cittadinanza (o informazioni correlate). Tra gli archivi canonici troviamo presenze nell'anagrafe delle persone residenti e nei permessi di soggiorno. Tra quelli meno canonici troviamo che circa 570 mila universitari hanno una fonte di reddito, 264 mila sono in "EMens" e 70 mila tra le "persone d'impresa" da cui emerge che gli studenti-lavoratori sono uno strato nello strato.

**Tavola 2.3 - Individui nella popolazione insistente (PP) per presenza negli archivi amministrativi, tipo di individuo e valorizzazione della cittadinanza - anno 2014** (valori assoluti per 1000)

ARCHIVI	PP		Tipo _Individuo in PP								Unità nell'archivio
	Totale	con cittadinanza null	Lavoratore a dicembre		Universitario		Studente		Altro individuo		
			Totale	con cittadinanza null	Universitari	con cittadinanza null	Studenti	con cittadinanza null	Totale	con cittadinanza null	
<b>PP</b>	<b>61.373</b>	<b>1.372</b>	<b>22.682</b>	<b>665</b>	<b>1.697</b>	<b>57</b>	<b>7.294</b>	<b>179</b>	<b>30.114</b>	<b>480</b>	
INAIL - Denunce Nominative Assicurative	16.678	868	11.392	375	261	6	82	2	2.598	410	17.701
INPS - EMens	14.613	513	13.221	423	265	6	78	2	589	72	14.647
CCIAA - Archivio Persone d'Impresa - Iscrizione della società nel Registro delle Imprese	8.008	148	4.858	111	71	1	16	0	2.531	23	8.535
MIUR - Anagrafe degli studenti	7.288	179	76	2	1	0	7.288	179	0	0	7.295
Ministero dell'Interno - Permessi di soggiorno	2.842	233	1.100	116	50	15	116	8	1.240	68	3.299
CCIAA - Archivio soci delle società di capitale (Banca Dati Soci)	2.421	40	1.487	30	37	1	7	0	766	6	2.600
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	1.697	57	258	5	1.697	57	1	0	0	0	1.697
INAIL - Lavoratori parasubordinati	1.008	25	364	9	49	1	11	0	378	9	1.071
INAIL - Lavoratori interinali	487	15	218	5	16	0	4	0	107	5	488
MIUR - Archivio delle lauree e dei laureati	434	6	108	2	142	3	0	0	175	1	445
MAE - Ministero Affari Esteri - Archivio Italiani all'Estero	277	60	64	24	6	3	17	4	179	26	7.004
MIUR - Dottorati di ricerca	18	1	3	0	0	0	0	0	11	1	20

I "permessi di soggiorno" e i dati relativi alla "acquisizione e reiezione della cittadinanza" (*cf. paragrafi 2.1.3 e 2.1.4*) sono esigui in termini di copertura della popolazione. Questi archivi sono, tuttavia, particolarmente importanti poiché danno informazioni sul sottoinsieme critico degli stranieri e, nel caso specifico, dei permessi di soggiorno di persone non iscritte in LAC.

Inoltre, la variabile cittadinanza è la principale informazione raccolta negli archivi "Concessioni", "Giuramenti" e "Respingimenti" del Ministero dell'Interno in cui vengono registrati il numero dei procedimenti conclusi favorevolmente di acquisizione e concessione della cittadinanza italiana nonché il numero dei respingimenti nell'anno di riferimento (Ministero dell'Interno, 2014).

Le informazioni contenute in questi archivi (in particolare l'archivio sui giuramenti) sono utili per correggere eventuali ritardi nell'aggiornamento della cittadinanza in LAC e per avere ulteriori informazioni di qualità sulla variabile poiché, infatti, nelle richieste di cittadinanza italiana viene osservata la cittadinanza posseduta al momento della richiesta.

Data la natura della fonte, si è ipotizzata una buona qualità dell'informazione relativa alla variabile oggetto di studio.

### 2.1.2 Tempestività e concordanza

Un aspetto importante da tenere in considerazione quando si utilizzano dati amministrativi è la tempistica delle fonti inerente il loro arrivo in Istat (Tavola 2.4) e i relativi tempi di lavorazione necessari per la fruibilità delle informazioni rilevate.

**Tavola 2.4 - Tempistica delle fonti analizzate**

FONTI - ARCHIVIO	SCARTO (IN MESI) RISPETTO ALL'EVENTO REGISTRATO NELL'ARCHIVIO
Ministero dell'Interno - Permessi di soggiorno	T + 2
Comuni - LAC	T + 3
INAIL - Denunce Nominative Assicurative	T + 4
Ministero dell'Interno - Acquisizione/reiezione cittadinanza	T + 9
INAIL - Parasubordinati	T + 10
INPS - EMens	T + 11
CCIAA - Archivio Persone d'Impresa	T + 11
CCIAA - Archivio Soci Titolari	T + 11
MIUR - Archivio delle lauree e dei laureati	T + 12
MIUR - Anagrafe degli studenti	T + 14
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	T + 14

L'analisi di qualità della variabile cittadinanza nelle diverse fonti è basata sulla concordanza tra i valori riportati negli archivi e quelli nella fonte LAC, che ne costituisce il riferimento principale. Il grado di concordanza è stato calcolato relativamente ai due sottoinsiemi degli individui classificati come italiani (ovvero con cittadinanza italiana) e stranieri. Le analisi vengono effettuate solo per gli individui che hanno in entrambi gli archivi la cittadinanza valorizzata.

In una logica previsionale misuriamo il grado di errore (inteso qui come discordanza dalla LAC) connesso alla previsione ottenuta prendendo il valore della fonte alternativa. In questa ottica, indicando con  $CL$  la variabile che indica la cittadinanza in LAC e  $CF$  la cittadinanza nella fonte alternativa, una stima della bontà di previsione è data dalla probabilità condizionata  $Pr(CL=i|CF=i)$  che indica la probabilità di avere la cittadinanza  $i$ -esima in LAC, dato che la cittadinanza nella fonte alternativa è ' $i$ '. Tale probabilità è stimata con la frequenza relativa ottenuta come rapporto del numero totale di individui concordi per una specifica cittadinanza ' $i$ ' nelle due fonti, relativizzato alla numerosità totale del sottoinsieme degli individui della fonte alternativa con cittadinanza ' $i$ '. Più precisamente, indicando con  $N(CL=i, CF=i)$  la frequenza assoluta delle osservazioni che hanno la cittadinanza  $i$ -esima in LAC e fonte alternativa e con  $N(CF=i)$  la numerosità delle osservazioni che hanno la cittadinanza  $i$ -esima nella fonte alternativa si ha che una stima di  $Pr(CL=i|CF=i)$  può essere ottenuta come:

$$\widehat{Pr}(CL=i|CF=i) = N(CL=i, CF=i) / N(CF=i)$$

Una parte delle analisi sono state condotte all'interno di specifici segmenti di popolazione, ovvero studenti, lavoratori, altri, al fine di individuare gli archivi primari per le suddette popolazioni e quindi scegliere il minor numero di archivi per la ricostruzione della cittadinanza. Di seguito al prospetto riepilogativo (Tavola 2.5) sono riportate le singole analisi di concordanza della cittadinanza per le diverse fonti.

Nei paragrafi che seguono, le analisi di concordanza vengono condotte con riferimento a tre gruppi di fonti.

## 2. Analisi di qualità delle fonti

**Tavola 2.5 - Concordezza tra cittadinanza in LAC e nell'Archivio (italiani e stranieri, anno 2014)**  
(composizioni percentuali)

FONTE - ARCHIVIO	Concordezza	
	Italiani	Stranieri
INAIL - Denunce Nominative Assicurative	98,6	96,8
Ministero dell'Interno - Permessi di soggiorno	NA	94,1
INAIL - Parasubordinati	99,5	92,6
INPS - EMens	97,6	89,5
MIUR - Archivio delle lauree e dei laureati	99,9	89,2
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	99,9	90,1
MIUR - Anagrafe degli studenti	99,4	82,2
CCIAA - Archivio Persone d'Impresa	83,6	91,3
CCIAA - Archivio Soci Titolari	99,6	83,3

Nel primo gruppo le informazioni relative alla cittadinanza non sono l'oggetto principale del dato raccolto, come ad esempio le fonti Inps che hanno come obiettivo quello di comunicare all'ente preposto alla gestione delle prestazioni previdenziali 'il soggetto lavoratore' e non le sue caratteristiche quali, tra le altre, la cittadinanza.

Nel secondo gruppo di dati amministrativi la cittadinanza è l'oggetto principale della rilevazione amministrativa, in questo gruppo rientrano le fonti del Ministero dell'Interno con riferimento ai dati relativi all'*Acquisizione e reiezione della cittadinanza italiana*.

Nel terzo gruppo, infine, viene analizzata la concordezza fra la cittadinanza e il luogo di nascita. L'informazione relativa al luogo di nascita può risultare utile nel caso in cui non siano disponibili informazioni amministrative concernenti la cittadinanza, essendo correlata con quest'ultima.

### 2.1.3 Analisi della concordezza degli archivi amministrativi MIUR, INPS, INAIL, CCIAA

In questo paragrafo vengono analizzate le fonti che raccolgono informazioni 'indirette' sulla cittadinanza.

#### Analisi della concordezza tra "MIUR – Anagrafe degli studenti" e LAC

La popolazione di riferimento su cui vengono condotte le comparazioni è composta dagli studenti della scuola dell'obbligo come da classificazione P&P 2014. Nella tavola seguente viene indicata la frequenza assoluta e percentuale relativa alla concordezza tra la variabile cittadinanza osservata nell' "Anagrafe degli studenti" e quella rilevata in LAC (Tavola 2.6).

**Tavola 2.6 - Concordezza della variabile cittadinanza fra la fonte "MIUR - Anagrafe degli studenti" e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Sì	6.444.116	99,4	495.821	82,2
No	42.072	0,6	107.402	17,8
<b>Totale</b>	<b>6.486.188</b>	<b>100,0</b>	<b>603.223</b>	<b>100,0</b>

Nel caso di studenti della scuola dell'obbligo italiani la concordanza con la LAC è pressoché totale (99,4%), mentre in caso di studenti stranieri in MIUR le differenze sono maggiori anche a causa del fatto che in LAC molti di loro risultano italiani (concordanza pari al 82,2%).

#### *Analisi della concordanza tra "MIUR – Archivio degli iscritti e delle iscrizioni universitarie" e LAC*

La popolazione di riferimento su cui vengono condotte le comparazioni è composta dagli "studenti universitari" come da classificazione P&P 2014. Nella tavola seguente viene indicata la frequenza assoluta e percentuale relativa alla concordanza tra la variabile cittadinanza osservata nell' "Archivio degli iscritti e delle iscrizioni universitarie" e quella rilevata in LAC (Tavola 2.7).

**Tavola 2.7 - Concordanza della variabile cittadinanza fra la fonte "MIUR - Archivio degli iscritti e delle iscrizioni universitarie" e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Sì	1.591.489	99,9	42.085	90,1
No	1.743	0,1	4.641	9,9
<b>Totale</b>	<b>1.593.232</b>	<b>100,0</b>	<b>46.726</b>	<b>100,0</b>

Analogamente a quanto osservato per gli studenti della scuola dell'obbligo, la concordanza per gli universitari italiani è quasi totale (99,9%). Per gli universitari stranieri, invece, la discordanza è meno accentuata rispetto a quella riscontrata dagli studenti non universitari (concordanza pari al 90,1%).

#### *Analisi della concordanza tra "MIUR - Archivio delle lauree e dei laureati" e LAC*

La popolazione di riferimento su cui vengono condotte le comparazioni è composta dai laureati come da classificazione P&P 2014. Nella tavola seguente viene individuata la frequenza assoluta e percentuale relativa alla concordanza tra la variabile cittadinanza osservata nell' "Archivio delle lauree e dei laureati" e quella rilevata in LAC (Tavola 2.8).

**Tavola 2.8 - Concordanza della variabile cittadinanza fra la fonte "MIUR - Archivio delle lauree e dei laureati" e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Sì	135.936	99,9	2.533	89,2
No	139	0,1	306	10,8
<b>Totale</b>	<b>136.075</b>	<b>100,0</b>	<b>2.839</b>	<b>100,0</b>

Si riscontrano risultati analoghi a quelli osservati per gli studenti universitari (concordanza al 99,9% per i laureati italiani, all'89,2% per i laureati stranieri) poiché l'archivio dei laureati ne è un sottoinsieme.

*Analisi della concordanza tra “INPS - EMens” e LAC*

La popolazione di riferimento è rappresentata dai lavoratori dipendenti come da classificazione P&P 2014. L'archivio EMens ha una struttura differente rispetto alle altre fonti: mentre nella maggior parte degli archivi viene riportata una sola cittadinanza con riferimento all'anno, in EMens - che riporta delle comunicazioni mensili - possono coesistere più cittadinanze. Inoltre, poiché le comunicazioni mensili possono essere rese da datori di lavoro diversi, non è scontata l'uguaglianza tra le informazioni relative alla cittadinanza registrata per lo stesso individuo.

È questa la ragione per cui, in questa prima analisi, abbiamo preso in considerazione le sole unità che hanno valori concordi nei mesi presenti, essendo le discordanze in numero limitato (Tavola 2.9) rispetto al totale delle osservazioni presenti in EMens.

**Tavola 2.9 - Lavoratori di P&P 2014 con una o più cittadinanze in “INPS - EMens”** (valori assoluti e composizione percentuale)

LAVORATORI DI P&P Valori Assoluti	Valori assoluti	Percentuale
Con un'unica cittadinanza in EMens	13.865.464	98,9
Con più cittadinanze in EMens	149.666	1,1
<b>Totale</b>	<b>14.015.130</b>	<b>100,0</b>

**Tavola 2.10 Concordanza della variabile cittadinanza fra la fonte “INPS - EMens” e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Si	12.178.719	97,6	858.260	89,5
No	302.715	2,4	100.690	10,5
<b>Totale</b>	<b>12.481.434</b>	<b>100,0</b>	<b>958.950</b>	<b>100,0</b>

La cittadinanza rilevata nella fonte EMens concorda per il 97,6% con quella osservata in LAC per quanto riguarda i lavoratori italiani, mentre concorda all'89,5% per i lavoratori stranieri (Tavola 2.10).

*Analisi della concordanza tra “INAIL - Parasubordinati” e LAC*

La popolazione di riferimento è quella composta dai lavoratori parasubordinati come da classificazione P&P 2014. Nella tavola seguente viene individuata la frequenza assoluta e percentuale relativa alla concordanza tra la variabile cittadinanza osservata nell' “Archivio dei parasubordinati” e quella rilevata in LAC (Tavola 2.11).

**Tavola 2.11 - Concordanza della variabile cittadinanza fra la fonte “INAIL - Parasubordinati” e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Si	775.769	99,5	38.127	92,6
No	4.016	0,5	3.043	7,4
<b>Totale</b>	<b>779.785</b>	<b>100,0</b>	<b>41.170</b>	<b>100,0</b>

Questo archivio presenta valori di concordanza elevati sia per i lavoratori parasubordinati italiani (99,5%), sia per i parasubordinati stranieri (92,6%).

*Analisi della concordanza tra “INAIL – Denunce Nominative Assicurative” e LAC*

La popolazione di riferimento è costituita dai cosiddetti NOLUS (individui che NON hanno segnali di Lavoro, Università, Studio) come da classificazione P&P 2014, individuati nell’archivio delle “Denunce nominative Assicurative”. Nella tavola seguente viene rappresentata la frequenza assoluta e percentuale relativa alla concordanza tra la variabile cittadinanza osservata nel sottogruppo dei NOLUS in DNA e quella rilevata in LAC (Tavola 2.12).

**Tavola 2.12 - Concordanza della variabile cittadinanza fra la fonte “INAIL - DNA” dei NOLUS e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Sì	774.368	98,6	283.478	96,8
No	11.000	1,4	9.278	3,2
<b>Totale</b>	<b>785.368</b>	<b>100,0</b>	<b>292.756</b>	<b>100,0</b>

La capacità predittiva dell’archivio DNA dei NOLUS è una delle migliori riscontrate rispetto a quelle osservate per altri archivi relativamente ai cittadini stranieri (96,8%). Si osserva, inoltre, un’ottima percentuale anche per i cittadini italiani (98,6%).

*Analisi della concordanza tra “Ministero dell’Interno - Permessi di soggiorno” e “INAIL – Denunce Nominative Assicurative”*

La fonte relativa ai Permessi di Soggiorno riporta solo cittadinanze non italiane e risulta essere di particolare importanza poiché fa riferimento a uno specifico sottoinsieme non presente in LAC. La qualità di questa fonte è stata analizzata confrontando il suo grado di concordanza con “INAIL - DNA” che ha, a sua volta, un livello di concordanza con la LAC riferita alla popolazione straniera particolarmente elevato. I risultati del confronto sono riportati nella tavola 2.13.

**Tavola 2.13 - Concordanza della variabile cittadinanza fra la fonte “Ministero degli Interni - Permessi di soggiorno” e “INAIL-DNA” (anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Valori assoluti	Percentuale
Sì	901.936	94,1
No	56.208	5,9
<b>Totale</b>	<b>958.144</b>	<b>100,0</b>

La concordanza tra i permessi di soggiorno e DNA è del 94,1 %.

*Analisi della concordanza tra “CCIAA - Camera di Commercio Industria Artigianato e Agricoltura” e LAC*

Gli archivi della fonte CCIAA analizzati sono “CCIAA - Persone Impresa” e “CCIAA - Soci titolari”.

La popolazione di riferimento è costituita dai due archivi CCIAA sopra citati come da classificazione P&P 2014. Nelle tavole seguenti viene rappresentata la frequenza assoluta e percentuale relativa alla concordanza tra la variabile cittadinanza osservata nei due archivi CCIAA e quella rilevata in LAC (Tavole 2.14 e 2.15).

## 2. Analisi di qualità delle fonti

**Tavola 2.14 - Concordanza della variabile cittadinanza fra la fonte “CCIAA-Persone impresa” e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Si	137.936	83,6	378.722	91,3
No	27.043	16,4	36.279	8,7
<b>Totale</b>	<b>164.979</b>	<b>100,0</b>	<b>415.001</b>	<b>100,0</b>

Contrariamente a quanto osservato in altri archivi, nel confronto tra “CCIA - Persone Impresa” e LAC la concordanza è maggiore per gli stranieri (91,3% vs 83,6%).

**Tavola 2.15 - Concordanza della variabile cittadinanza fra la fonte “CCIAA-Soci titolari” e LAC (italiani e stranieri, anno 2014)** (valori assoluti e composizione percentuale)

CONCORDANZA	Italiani		Stranieri	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Si	2.334.838	99,6	31.217	83,3
No	10.470	0,4	6.261	16,7
<b>Totale</b>	<b>2.345.308</b>	<b>100,0</b>	<b>37.478</b>	<b>100,0</b>

L’archivio “CCIAA - Soci titolari” presenta, analogamente alla maggior parte degli altri archivi, concordanza maggiore per gli italiani rispetto agli stranieri.

### 2.1.4 Analisi della concordanza tra “Ministero dell’Interno - Acquisizione e reiezione della cittadinanza italiana” e LAC

In questo paragrafo vengono riportati i risultati delle analisi condotte su un archivio particolarmente rilevante in termini di contenuti informativi per la variabile cittadinanza. L’archivio è quello del Ministero dell’Interno e fa riferimento ai dati relativi all’*Acquisizione e reiezione della cittadinanza italiana*. Diversamente dagli altri archivi, in questo caso la fonte raccoglie a fini amministrativi informazioni dirette sulla cittadinanza e per tali motivi ci si attende una elevata qualità.

Lo studio ha interessato gli archivi relativi agli anni 2015 e 2016 (il 2014 non era disponibile), confrontati con la Lista Anagrafica Comunale (LAC) rispettivamente del 2015 e del 2016. La variabile cittadinanza viene raccolta negli archivi “Concessioni”, “Giuramenti” e “Respingimenti” del Ministero dell’Interno in cui vengono rilevati il numero dei procedimenti conclusi favorevolmente di acquisizione e concessione della cittadinanza italiana nonché il numero dei respingimenti nell’anno di riferimento (Ministero dell’Interno 2014). Dal punto di vista operativo è stato necessario effettuare una transcodifica per trasformare la classificazione della cittadinanza dei tre archivi considerati in quella adottata dall’Istat. Le analisi condotte separatamente per anno hanno evidenziato risultati pressoché identici. In questo lavoro si riportano solo quelli relativi all’anno 2015.

Come mostrato nella tavola 2.16, l’archivio più numeroso è quello delle “concessioni” di cittadinanza. Vi sono diversi record con codice individuo “null”, le percentuali variano tra il 12% dell’archivio “giuramenti” al 26% dell’archivio “respingimenti”. Inoltre, diversi codici individuo risultano duplicati o triplicati all’interno del singolo archivio (individui di cluster 2 e 3, ovvero che compaiono rispettivamente due o tre volte per anno nell’archivio). Tali numerosità necessitano quindi di ulteriori approfondimenti, prendendo in considerazione anche la possibilità che vi siano errori di linkage nell’attribuzione del codice individuo. Per rendere omogenee le analisi successive si è deciso di studiare

solo gli individui con codice individuo diverso da “null” e appartenenti al cluster 1 (individui che compaiono una sola volta per anno nell’archivio preso in considerazione).

**Tavola 2.16 - Numero di record e numero di individui per archivio (Ministero dell’Interno - acquisizione e reiezione della cittadinanza italiana, anno 2015) (valori assoluti e composizione percentuale)**

NOME ARCHIVIO	Record (v. a.)	Individui (v. a.)	Codice individuo null		Individui di cluster 1 (v. a.)	Individui di cluster 2 (v. a.)	Individui di cluster 3 (v. a.)
			v. a.	%			
Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Concessioni	122.196	106.440	15.705	12,9	106.390	98	3
Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Giuramenti	86.632	76.276	10.316	11,9	76.236	80	0
Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Respingimenti	2.541	1.865	676	26,6	1.865	0	0

Il primo confronto effettuato riguarda gli archivi “concessioni” e “giuramenti” (Tavola 2.17). Il 56,2% delle concessioni di cittadinanza presenti nell’archivio del 2015 si concretizza con l’atto formale del giuramento per l’acquisizione della cittadinanza nello stesso anno (59.745 individui). Il restante 43,9% (46.645 individui) è presente nelle “concessioni” ma non tra i “giuramenti”. La lista di questi individui è stata quindi confrontata con la rispettiva LAC ipotizzando di osservare cittadinanza straniera per coloro che risultano non aver prestato giuramento. Il 91% di questi è nella lista anagrafica comunale, ma contrariamente a quanto ipotizzato, il 35% presenta cittadinanza italiana, mentre tutti gli altri individui hanno cittadinanza straniera.

Gli individui presenti nella lista dei “giuramenti” ma non in quella delle “concessioni” (16.491 individui) sono presenti in LAC e risultano coerentemente con cittadinanza italiana (99%).

**Tavola 2.17 - Presenza/assenza degli individui negli archivi “Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Concessioni” e “Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Giuramenti” (anno 2015) (valori assoluti)**

Presenza in: “Concessioni”	“Giuramenti”		Totale
	Sì	No	
Sì	59.745	46.645	<b>106.390</b>
No	16.491	0	<b>16.491</b>
<b>Totale</b>	<b>76.236</b>	<b>46.645</b>	<b>122.881</b>

Lo studio è proseguito analizzando le numerosità degli individui con cittadinanza straniera nella LAC 2015 per i quali ci si aspettava invece di osservare cittadinanza italiana, perché presenti in “giuramenti”. Queste quantità confrontate con la lista anagrafica comunale del 2016 hanno evidenziato la modifica della cittadinanza da straniera a italiana nell’anno successivo a quello di riferimento.

Il confronto tra “giuramenti” e LAC e tra “respingimenti” e LAC (Tavole 2.18 e 2.19) indica che la lista anagrafica comunale recepisce correttamente le modifiche di cittadinanza e i respingimenti. Il 99% dei giuramenti si trova nella LAC dell’anno di riferimento con cittadinanza italiana e il 99% degli individui osservati nell’archivio “respingimenti” presenta cittadinanza straniera nella LAC dell’anno di riferimento. Inoltre nei casi in cui si osserva cittadinanza straniera invece che italiana, cioè l’1% dei “giuramenti”, vi è concordanza tra il dato presente in “giuramenti” e quello presente in LAC (per il 97% dei casi). L’utilizzo dell’archivio “concessioni” e il rispettivo confronto con la LAC deve sempre essere accom-

pagnato dalle informazioni contenute nel file “giuramenti” perché è in quest’ultimo archivio che si tiene conto dell’atto formale di attribuzione della cittadinanza italiana che poi confluisce anche nella lista anagrafica comunale.

**Tavola 2.18 - Presenza/assenza degli individui negli archivi “LAC” e “Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Giuramenti” (anno 2015) (valori assoluti)**

Presenza in: “LAC”	“Giuramenti”		Totale
	Si	No	
Si	70.815	60.627.098	<b>60.697.913</b>
No	5.421	0	<b>5.421</b>
<b>Totale</b>	<b>76.236</b>	<b>60.627.098</b>	<b>60.703.334</b>

**Tavola 2.19 - Presenza/assenza degli individui negli archivi “LAC” e “Min. Int. - Acquisizione e reiezione della cittadinanza italiana - Respingimenti” (anno 2015) (valori assoluti)**

Presenza in: “LAC”	“Respingimenti”		Totale
	Si	No	
Si	1.638	60.696.275	<b>60.697.913</b>
No	227	0	<b>227</b>
<b>Totale</b>	<b>1.865</b>	<b>60.696.275</b>	<b>60.698.140</b>

Da questo quadro emerge l’esigenza di effettuare opportuni approfondimenti: per sfruttare a pieno le informazioni di questa preziosa fonte, i risultati lasciano supporre che l’archivio “giuramenti” non sia completo o che l’archivio “concessioni” non sia depurato dai giuramenti prestati o che comunque ci sia uno sfasamento temporale tra le informazioni presenti negli archivi; occorre inoltre tenere conto anche del fatto che rispetto all’anno t, gli archivi del Ministero dell’Interno sono disponibili a settembre, mentre la LAC a marzo.

Sulla base dei risultati ottenuti e sulla concordanza delle informazioni tra “giuramenti”, “respingimenti” e LAC, si è quindi deciso di introdurre i tre archivi considerati all’interno della procedura di ricostruzione della cittadinanza, attribuendo una priorità più elevata agli archivi “giuramenti” e “respingimenti”, e una priorità minore all’archivio “concessioni” (cfr. *paragrafo 3.1.1*).

### 2.1.5 Confronto tra paese di nascita in “Anagrafe Tributaria” (AT) e cittadinanza in LAC nel 2015

Come ulteriore passo di analisi si è deciso di confrontare il paese di nascita rilevato in Anagrafe Tributaria con la cittadinanza rilevata in LAC per capire se è possibile utilizzare il paese di nascita come proxy della cittadinanza e ottenere informazioni riguardo gli individui che non presentano segnali in nessun altro archivio al di fuori di AT (Tavola 2.20).

**Tavola 2.20 - Confronto tra paese di nascita in AT e cittadinanza in LAC (nati in Italia e nati all'estero, anno 2014) (valori assoluti e composizione percentuale)**

CONCORDANZA	Nati in Italia in AT		Nati all'estero in AT	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale
Si	54.105.747	98,1	4.069.897	64,6
No	1.049.748	1,9	2.231.417	35,4
<b>Totale</b>	<b>55.155.495</b>	<b>100,0</b>	<b>6.301.314</b>	<b>100,0</b>

Le risultanze ottenute mostrano, in caso di nati in Italia una concordanza pari al 98,1%, mentre per i nati all’estero si evidenzia una maggiore discordanza (35,4%) tra paese di na-

scita e cittadinanza, implicando che un nato all'estero non abbia necessariamente la cittadinanza del paese di nascita. Il dato sul paese di nascita, pertanto, deve essere accompagnato da ulteriori informazioni per poter caratterizzare la cittadinanza dell'individuo.

### 3. DISEGNO, SPERIMENTAZIONE DELLA PROCEDURA DI STIMA DELLA CITTADINANZA<sup>1</sup>

#### 3.1 Specifiche statistiche ed architettura del processo di stima della cittadinanza in RBI

L'algoritmo utilizzato per la ricostruzione della cittadinanza è di tipo gerarchico, ovvero viene stabilita una sequenza con cui entrano nell'algoritmo le fonti da cui viene presa l'informazione relativa alla cittadinanza. Tale gerarchia viene stabilita principalmente in base alla qualità della fonte misurata in termini di concordanza con la LAC, alla tempestività ed alla rilevanza.

La definizione dell'algoritmo da utilizzare per la ricostruzione della cittadinanza in RBI ha seguito diverse fasi di lavorazione. In una prima fase, sono state condotte delle sperimentazioni sui dati P&P. In base ai risultati, è stata proposta una prima versione dell'algoritmo sperimentata sui dati del RBI. I risultati di questa seconda sperimentazione sono stati analizzati e confrontati con le informazioni sulla cittadinanza in ANVIS. A seguito di questi confronti è stata proposta la versione finale dell'algoritmo. Le analisi condotte sui dati P&P sono riportate nel paragrafo 3.1.1. Il paragrafo 3.1.2 è invece dedicato all'illustrazione dell'applicazione della ricostruzione in RBI.

##### 3.1.1 Sperimentazione sui dati di Persons & Places

In questo paragrafo vengono descritti i risultati della sperimentazione condotta sulla popolazione P&P relativamente al 2014 e 2015, ed esplicitata la gerarchia fra le fonti, e le altre specifiche di processo formalizzate per la prima versione dell'algoritmo.

#### **Specifiche per l'Algoritmo:**

*A1. Mettere a missing i valori fuori dominio per le variabili relative alla cittadinanza in:*

1. *Acquisizione e reiezione della cittadinanza italiana - GIURAMENTI*
2. *Acquisizione e reiezione della cittadinanza italiana - REIEZIONI*
3. *LAC*
4. *Ministero dell'Interno - Permessi di soggiorno*
5. *INAIL - Denunce Nominative Assicurative*
6. *INAIL - Parasubordinati*
7. *INPS - EMens*
8. *MIUR - Archivio delle lauree e dei laureati*
9. *MIUR - Archivio degli iscritti e delle iscrizioni universitarie*

<sup>1</sup> Marco Di Zio (paragrafi 3.1 e 3.1.1), Stefania Di Domenico (paragrafi 3.1.1), Roberta Vivio (paragrafo 3.1.1), Roberta Radini (paragrafo 3.1.2).

10. MIUR - Anagrafe studenti
11. Acquisizione e reiezione della cittadinanza italiana - CONCESSIONI
12. CCIAA - Persone impresa
13. CCIAA - Soci titolari

A2. Riclassificare le variabili relative alla cittadinanza nelle diverse fonti in base alla classificazione ufficiale Istat.

A3. A esclusione di EMens e dell'anagrafe degli studenti, prendere la cittadinanza più recente (laddove sia non nulla, altrimenti prendere la cittadinanza non nulla più recente).

A4. Per la variabile cittadinanza in EMens e nell'anagrafe degli studenti, prendere in considerazione solo le unità che non hanno inconsistenze nella variabile cittadinanza nei mesi presenti in EMens, ovvero presentano un unico paese di cittadinanza<sup>2</sup>.

A5. Ricostruire la variabile cittadinanza in base ad un approccio gerarchico, ovvero prendere il valore presente secondo il seguente ordine:

1. Acquisizione e reiezione della cittadinanza italiana - GIURAMENTI
2. Acquisizione e reiezione della cittadinanza italiana - REIEZIONI
3. LAC
4. Ministero dell'Interno - Permessi di soggiorno
5. INAIL - Denunce Nominative Assicurative
6. INAIL - Parasubordinati
7. INPS - EMens
8. MIUR - Anagrafe studenti
9. MIUR - Archivio delle lauree e dei laureati
10. MIUR - Archivio degli iscritti e delle iscrizioni universitarie
11. Acquisizione e reiezione della cittadinanza italiana - CONCESSIONI
12. CCIAA - Persone impresa
13. CCIAA - Soci titolari

L'applicazione della procedura ha messo in evidenza che non tutte le modalità possono essere ricostruite tramite l'algoritmo così definito. Per queste modalità residue la cittadinanza è stata ricostruita utilizzando il "luogo di nascita". Dopo questo passo rimangono ancora da ricostruire le unità per le quali il luogo di cittadinanza non è più compatibile con lo stato al tempo di riferimento, per esempio per le persone nate nella ex-Jugoslavia. Per queste si è deciso di utilizzare - ove presente - il dato del censimento del 2011 nel caso in cui lo stato al censimento 2011 risulti compatibile. Per i casi rimanenti è stato stimato lo stato in base alla distribuzione di frequenza delle persone aventi una determinata cittadinanza non più attuale.

Nella tavola 3.1 sono riportate le numerosità delle unità ricostruite con le sole fonti amministrative. Osserviamo che la LAC permette di valorizzare circa il 98% nel 2014 e il 99% nel 2015. La differenza è dovuta al fatto che nel 2014 ci sono stati dei problemi con alcune regioni che hanno inviato con ritardo la LAC.

<sup>2</sup> Le altre unità non verranno usate per la ricostruzione poiché non è possibile stabilire con certezza quale sia l'ultima cittadinanza valida.

### 3. Disegno, sperimentazione della procedura di stima della cittadinanza

**Tavola 3.1 - Individui per cittadinanza ricostruita, per fonte (anno 2015)** (valori assoluti e composizione percentuale)

ARCHIVI	Valori assoluti	Percentuale
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Giuramenti	72.092	0,12
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Respingimenti	1.660	0,00
LAC	60.601.582	98,81
Ministero degli Interni - Permessi di soggiorno	186.869	0,30
INAIL - Denunce Nominative Assicurative	180.194	0,29
INAIL - Parasubordinati	7.886	0,01
INPS - EMens	100.544	0,16
MIUR - Anagrafe degli studenti	51.257	0,08
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	11.528	0,02
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Concessioni	97	0,00
CCIAA - Archivio Persone Impresa	5.552	0,01
CCIAA - Archivio Soci Titolari	8.059	0,01
Non disponibili	105.720	0,17
<b>Totale popolazione</b>	<b>61.333.040</b>	<b>100,00</b>

Le fonti alternative alla LAC contribuiscono per l'1,7% alla ricostruzione della cittadinanza della popolazione totale di P&P. Il loro contributo è rilevante soprattutto per la popolazione straniera (Tavola 3.2) non in LAC. Dopo l'integrazione con le fonti ausiliarie rimangono 105.720 unità senza cittadinanza valorizzata (NA), pari allo 0,2% della popolazione considerata nella sperimentazione (Tavola 3.1).

**Tavola 3.2 - Individui per cittadinanza ricostruita, per fonte (italiani e stranieri, anno 2015)**  
(valori assoluti e composizione percentuale)

ARCHIVI	Valori assoluti		Percentuale	
	Italiani	Stranieri	Italiani	Stranieri
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Giuramenti	72.092	0	100,0	0,0
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Respingimenti	0	1.660	0,0	100,0
LAC	55.594.006	5.007.576	91,7	8,3
Ministero degli Interni - Permessi di soggiorno	0	186.869	0,0	100,0
INAIL - Denunce Nominative Assicurative	49.728	130.466	27,6	72,4
INAIL - Parasubordinati	3.892	3.994	49,4	50,6
INPS - EMens	49.620	50.924	49,4	50,6
MIUR - Anagrafe degli studenti	19.730	31.527	38,5	61,5
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	4.326	7.202	37,5	62,5
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Concessioni	0	97	0,0	100,0
CCIAA - Archivio Persone Impresa	1.381	4.171	24,9	75,1
CCIAA - Archivio Soci Titolari	7.844	215	97,3	2,7
<b>Totale</b>	<b>55.802.619</b>	<b>5.424.701</b>	<b>91,1</b>	<b>8,9</b>

#### 3.1.2 Architettura del processo di stima della cittadinanza per gli individui in RBI

Il processo di produzione della variabile cittadinanza all'interno del Registro Base degli Individui (RBI) viene applicato dopo avere definito tutte le unità statistiche del registro e dopo aver costruito la versione di riferimento del registro. Il processo in oggetto è schematizzato nella Figura 6.1 ed è composto di varie fasi che di seguito vengono descritte.

La prima fase del processo è l'acquisizione delle fonti che contengono l'informazione relativa alla cittadinanza, e la transcodifica di tale informazione in modo che risulti omoge-

nea tra le diverse fonti. Il processo di transcodifica utilizza gli algoritmi descritti nel paragrafo 4 per associare ad ogni codifica proprietaria di ciascuna fonte il valore della codifica corrispondente rispetto alla classificazione ufficiale Istat della cittadinanza (Istat, 2018). In questa prima attività viene considerata una classificazione delle cittadinanze che comprende anche le variazioni storiche di questa.

Quindi il processo si avvia acquisendo per ogni fonte e per ogni individuo l'informazione sulla variabile cittadinanza con associato il codice di "Priorità" legato alla qualità della fonte come descritto in precedenza. La qualità delle fonti che determina l'indice di priorità è stato definito a valle della sperimentazione descritta in Capitolo 2.

Alla fine di questa prima fase si crea un archivio che riporta per ogni individuo l'insieme di valori di cittadinanza transcodificati acquisiti dalle varie fonti. A questo archivio viene applicato un algoritmo di selezione e integrazione secondo la priorità della fonte e altri criteri di seguito descritti.

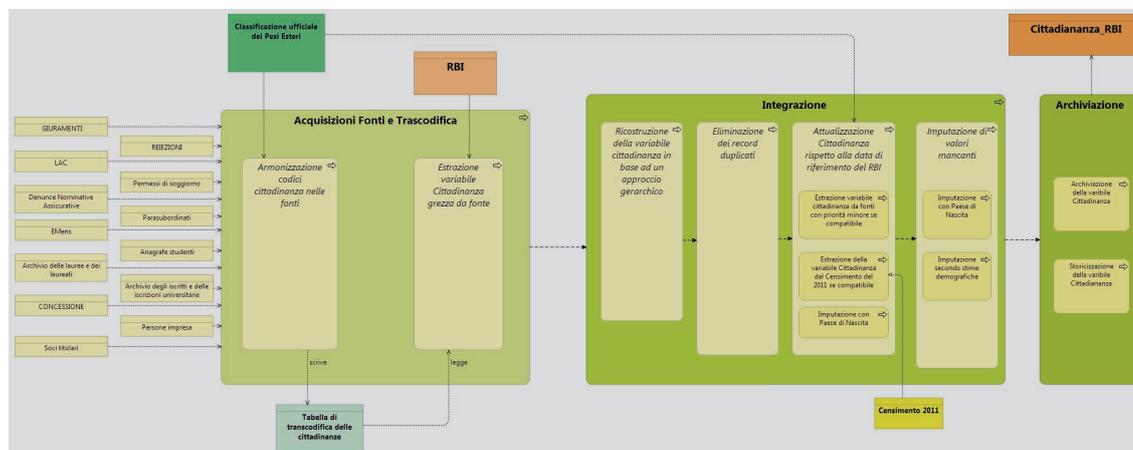
Questa seconda fase del processo è l'integrazione, ovvero la scelta del valore della variabile cittadinanza per ogni individuo. In particolare:

- Acquisizione, per ogni individuo della cittadinanza associata alla fonte con priorità maggiore;
- Deduplicazione dei record doppi presenti nelle fonti, soprattutto in LAC;
- Attualizzazione della cittadinanza rispetto alla data di riferimento del RBI. Questo sotto processo segue le seguenti regole di selezione della cittadinanza da fonti con priorità minore ma compatibile. In particolare per alcuni tipi di cittadinanza, come ex Jugoslavia, il valore attualizzato ad oggi corrisponde a 5 cittadinanze distinte. Si utilizza questo metodo per selezionare una cittadinanza valida rispetto alla data del registro. I passi sono:
  - Acquisizione della variabile cittadinanza da fonti con priorità minore se compatibile con quella indicata dalla fonte con priorità maggiore
  - Acquisizione della variabile Cittadinanza dal Censimento del 2011 se compatibile
  - Imputazione della variabile Cittadinanza con il Paese di Nascita se compatibile
- Imputazione di valori mancanti, solo nei casi in cui le procedure già descritte non abbiano prodotto un valore, variabile cittadinanza non valorizzata:
  - Imputazione con Paese di Nascita, se valorizzato e ad esso corrisponde un valore di cittadinanza valido alla data di riferimento del registro;
  - Imputazione secondo stime demografiche per record non risolti con i passi precedenti utilizzando un donatore casuale all'interno dello strato definito dalla cittadinanza (e.g., Jugoslavia). Questa procedura riproporrà la distribuzione osservata delle cittadinanze tra quelli che erano definiti come Jugoslavia.

La terza fase del processo è l'**archiviazione**, ovvero la memorizzazione della variabile Cittadinanza all'interno del RBI e storicizzazione delle modifiche di cittadinanza. Questa variabile a differenza delle altre variabili *core* del registro si considera variabile nel tempo.

### 3. Disegno, sperimentazione della procedura di stima della cittadinanza

Figura 3.1 - Processo di definizione della variabile cittadinanza nel Registro Base degli Individui





## 4. VALUTAZIONE DELLA PROCEDURA DI STIMA DELLA CITTADINANZA<sup>1</sup>

### 4.1 Valutazione della procedura di stima della cittadinanza in RBI

Per valutare la procedura di stima della cittadinanza in RBI, vengono analizzati i risultati relativamente alle fonti utilizzate, ed al loro impatto sul sottoinsieme dei residenti e dei non residenti. La distribuzione della cittadinanza ottenuta con la procedura proposta viene confrontata con quella relativa ad altre procedure, ovvero con la distribuzione stimata nell'Anagrafe Virtuale Statistica (ANVIS) e con quella ufficiale Istat pubblicata su Demo. Istat Il confronto con ANVIS è finalizzato non solo alla validazione della procedura, ma ad un allineamento delle procedure utilizzate da RBI e ANVIS. Questo allineamento è necessario poiché ANVIS stesso è una fonte di informazione fondamentale per la costruzione della popolazione residente in RBI, di conseguenza un disallineamento in una delle variabili principali nei due insiemi di dati potrebbe portare ad inconsistenze nel sistema delle statistiche demografiche e censuarie.

In questo capitolo la distribuzione della cittadinanza viene inoltre confrontata con la distribuzione ottenuta con una versione preliminare di RBI in cui si prendono in considerazione le fonti disponibili a 6 mesi di ritardo dal tempo di riferimento (t+6). Tale analisi è resa necessaria dal fatto che una versione preliminare di RBI al tempo t+6 viene utilizzata come lista di estrazione del campione per il censimento.

#### 4.1.1 Analisi dei risultati della procedura applicata ai dati del Registro Base degli Individui

L'algoritmo applicato al RBI 2017 ricostruisce la cittadinanza per il sottoinsieme dei residenti e per ulteriori 1.127.098 individui classificati come non residenti, di cui 1.028.552 sono stranieri.

La loro distribuzione per area geografica è riportata nella tavola 4.1.

**Tavola 4.1 - Popolazione non residente - cittadini stranieri per area di appartenenza** (composizione percentuale)

AREE GEOGRAFICHE	% Cittadinanza
Ue_28	28,0
Africa settentrionale	14,3
Europa centro-orientale	13,6
Asia orientale	10,2
Asia centro-meridionale	10,0
Africa occidentale	9,4
America centro-meridionale	5,2
America settentrionale	3,4
Altri Paesi europei	1,8
Asia occidentale	1,4
Africa orientale	1,4
Africa centro-meridionale	0,6
Altro	0,4
Oceania	0,2
Apolide	0,0

<sup>1</sup> Hanno contribuito alla redazione del capitolo: Marco Di Zio (paragrafo 4.1), Raffaella Maria Aracri (paragrafo 4.1.1), Mario Basevi e Gianni Corsetti (paragrafo 4.1.2) e Stefania Di Domenico (paragrafi 4.1.3 e 4.1.4).

In questo sottoinsieme le cittadinanze più frequenti sono riportate nella tavola 4.2.

**Tavola 4.2 - Popolazione non residente - cittadini stranieri per area di appartenenza** (composizione percentuale)

PAESE	% Cittadinanza
Romania	17,7
Cina	7,8
Marocco	6,8
Egitto	3,8
Albania	3,7
Stati Uniti d'America	3,2
Ucraina	2,9
Tunisia	2,8
Pakistan, Repubblica Islamica del	2,8
Bangladesh	2,7
Nigeria	2,7
India	2,5
Polonia	2,3
Senegal	1,9
Bulgaria	1,6
Moldavia, Repubblica di	1,5
Filippine	1,3
Kosovo	1,3
Ghana	1,3
Macedonia, Repubblica di	1,2
Sri Lanka	1,2
Brasile	1,1

La tavola 4.3 riporta la frequenza di utilizzo delle fonti per la ricostruzione della LAC nel sottoinsieme dei residenti di RBI. Nella stessa tavola si evidenzia anche la nazionalità ricostruita per fonte. Com'era da attendersi, la cittadinanza viene ricostruita principalmente dalla LAC (99,9%). I rimanenti casi ammontano a circa 70,000 unità.

**Tavola 4.3 - Popolazione residente - distribuzione delle fonti per priorità, ricostruzione della cittadinanza e cittadinanza residenti (italiani e stranieri)** (composizione percentuale)

FONTE	Priorità della fonte	% Ricostruzione cittadinanza	Totale residenti per fonte	
			% Italiani	% Stranieri
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Giuramenti	100	0,078	100,0	0,0
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Respingimenti	90	0,004	0,0	100,0
LAC	80	99,893	91,5	8,5
Ministero dell'Interno - Permessi di soggiorno	70	0,015	0,0	100,0
INAIL - Denunce Nominative	60	0,005	0,1	99,9
INAIL - Parasubordinati	50	0,000	0,0	100,0
MIUR - Archivio delle lauree e dei laureati	20	0,002	95,7	4,3
MIUR - Archivio degli iscritti e delle iscrizioni universitarie	10	0,000	96,9	3,1
Ministero dell'Interno - Acquisizione e reiezione cittadinanza italiana - Concessioni	9	0,000	100,0	0,0
CCIA - Persone impresa	8	0,001	37,5	62,5
CCIA - Soci titolari	7	0,000	0,0	100,0

Poiché il Registro Base degli Individui è utilizzato in una versione preliminare per l'estrazione di campioni, e dato il ritardo di alcune fonti rispetto ad un tempo di riferimento  $t$ , si è analizzata la differenza della cittadinanza in una versione del registro ottenuta con le informazioni disponibili al tempo  $t + 6$  mesi, con la versione che sfrutta tutti gli archivi, secondo l'algoritmo proposto. Questa valutazione legata alla tempestività delle fonti si rende necessaria poiché nel periodo di maggio di ogni anno  $t$  il registro RBI deve essere disponibile e aggiornato con riferimento data al primo gennaio di ogni anno  $t$  per fornire la lista di campionamento per il censimento permanente della popolazione.

In particolare, nella versione al tempo  $t + 6$  di RBI – per le unità per le quali non sono

#### 4. Valutazione della procedura di stima della cittadinanza

disponibili informazioni aggiornate – vengono utilizzate quelle dell’anno precedente.

Più in dettaglio, l’algoritmo di definizione della cittadinanza della popolazione residente al 01/01/2018 è stato eseguito in due versioni:

- la prima, utilizzando la fonte LAC del 2018, “Permessi di soggiorno” e “Denunce Nominative Assicurative” riferite al 2017 e tutte le altre fonti riferite al 2016;
- la seconda, utilizzando la fonte LAC del 2018 e tutte le altre fonti riferite al 2017.

Nelle tavole che seguono la versione 1 è riferita al calcolo della cittadinanza svolto a maggio 2018 e la versione 2 è riferita al calcolo della cittadinanza svolto a dicembre 2018.

La popolazione italiana e straniera si distribuisce:

- versione 1: Popolazione italiana - 91,503%, Popolazione straniera - 8,497%
- versione 2: Popolazione italiana - 91,500%, Popolazione straniera - 8,500%

La cittadinanza straniera distinta in aree di appartenenza si sintetizza in tavola 4.4.

**Tavola 4.4 - Popolazione residente - cittadini stranieri per area di appartenenza e versione (Maggio 2018 Ver. 1, Dicembre 2018 Ver. 2) (composizione percentuale)**

AREE GEOGRAFICHE	% Cittadinanza Ver.1	% Cittadinanza Ver.2	% Differenza tra Ver.1 e Ver.2
Ue_28	30,388	30,377	0,010
Europa centro-orientale	20,339	20,362	-0,023
Africa settentrionale	12,730	12,726	0,004
Asia centro-meridionale	10,076	10,074	0,002
Asia orientale	9,386	9,383	0,003
Africa occidentale	7,379	7,377	0,001
America centro-meridionale	6,921	6,919	0,002
Asia occidentale	0,937	0,937	0,000
Africa orientale	0,772	0,772	0,000
Africa centro-meridionale	0,482	0,482	0,000
America settentrionale	0,334	0,334	0,000
Altri Paesi europei	0,204	0,204	0,000
Oceania	0,041	0,042	0,000
Apolide	0,011	0,010	0,001

Le cittadinanze più numerose tra gli stranieri sono quelle riportate nella tavola 4.5.

**Tavola 4.5 - Popolazione residente - cittadini stranieri per paese di cittadinanza più frequente (Maggio 2018 Ver. 1, Dicembre 2018 Ver. 2) (composizione percentuale)**

PAESE	% Cittadinanza Ver.1	% Cittadinanza Ver.2
Romania	23,164	23,157
Albania	8,570	8,568
Marocco	8,095	8,093
Cina	5,649	5,648
Ucraina	4,613	4,612
Filippine	3,277	3,276
India	2,859	2,858
Bangladesh	2,565	2,564
Moldavia, Repubblica di	2,549	2,548
Egitto	2,328	2,327
Pakistan, Repubblica Islamica del	2,197	2,197
Sri Lanka	2,112	2,111
Nigeria	2,106	2,105
Senegal	2,060	2,061
Perù	1,897	1,896
Polonia	1,853	1,852
Tunisia	1,828	1,828
Ecuador	1,567	1,566
Macedonia, Repubblica di	1,260	1,260
Bulgaria	1,148	1,147

Come era da attendersi, le variazioni al livello Italia sono molto piccole, poiché la maggior parte delle cittadinanze viene ricostruita dalla LAC che è già disponibile nell’arco di

tempo  $t + 6$  mesi. Questo risultato conferma la possibilità di utilizzare la cittadinanza in RBI in una sua versione preliminare.

#### 4.1.2 Analisi di confronto della cittadinanza nel Registro Base degli Individui con quella di ANVIS

Con data di riferimento al 1° gennaio viene prodotto annualmente l'elenco delle persone residenti iscritte in anagrafe attraverso il confronto a livello individuale tra l'Anagrafe Virtuale Statistica (ANVIS) e le Liste Anagrafiche Comunali (LAC). A partire da queste fonti di dati e grazie anche al supporto delle rilevazioni di fonte anagrafica a livello aggregato - bilancio demografico annuale (P2/P3) e struttura per età, sesso e cittadinanza dei comuni (Posas/Strasa) - e di altri archivi amministrativi a livello individuale<sup>2</sup> si definisce una unica popolazione residente statistica denominata ANVIS++, che è una lettura trasversale (stock) e validata di ANVIS.

L'obiettivo principale di ANVIS++ è quello di fornire a RBI la localizzazione in termini di comune di residenza per ciascun individuo presente.

Le operazioni che portano alla costruzione di questa lista permettono anche di stimare e recuperare la sotto-copertura dei flussi anagrafici individuali non inviati dai comuni con riferimento all'ultimo anno considerato. Attraverso il reinserimento della sotto-copertura nel sistema longitudinale ANVIS viene garantito l'allineamento tra questo e RBI. Per tale ricostruzione è necessario stimare alcune informazioni, fra queste la cittadinanza.

La cittadinanza degli individui di ANVIS è valorizzata inizialmente attraverso i dati del Censimento 2011 se l'individuo era presente in quella occasione; viene invece valorizzata attraverso i dati dei flussi demografici se l'individuo entra (nascita o immigrazione dall'estero) per la prima volta in ANVIS dopo il Censimento 2011; gli stessi flussi demografici vengono anche utilizzati per aggiornare la cittadinanza degli individui.

ANVIS non gestisce tuttavia i microdati provenienti dalla fonte sulle acquisizioni di cittadinanza (Ministero dell'Interno) e quindi, per gli individui non soggetti ad eventi demografici, si utilizza l'informazione presente in LAC, poiché più aggiornata. Pertanto, in caso di discordanza della variabile cittadinanza su un individuo presente in LAC ed in ANVIS, viene presa la cittadinanza presente in LAC.

Tuttavia, l'informazione LAC riguardo alla cittadinanza non è sempre presente o valida<sup>3</sup>. In questi casi la cittadinanza in ANVIS++ è valorizzata attraverso il valore ANVIS, ovvero attraverso un processo di stima che fa ricorso al codice Belfiore presente nel codice fiscale e all'informazione sul luogo di nascita presente sia in LAC che in ANVIS. Nel complesso ANVIS++ al 1° gennaio 2018 contiene circa 70 mila individui con cittadinanza stimata.

Va inoltre considerato che prima dell'integrazione tra ANVIS e LAC, queste vengono sottoposte ad un processo di validazione che mira ad individuare ed eliminare le doppie iscrizioni in anagrafe mantenendo soltanto il record corrispondente all'iscrizione più recente indipendentemente dal codice cittadinanza presente.

Come è chiaro i due processi di ricostruzione della variabile cittadinanza in ANVIS++ ed RBI risultano essere scollegati e possono fornire risultati diversi sullo stesso individuo, pertanto è stato effettuato un confronto a livello di microdato e sui casi discordanti sono state definite delle regole per portare a coerenza i due stock.

Per quanto riguarda i dati riferiti al 1° gennaio 2018 il numero di righe discordanti è

<sup>2</sup> Anagrafe Consolare e Anagrafe Tributaria.

<sup>3</sup> Per codice cittadinanza valido si intende un valore riferito ad un paese estero esistente al tempo dell'analisi.

#### 4. Valutazione della procedura di stima della cittadinanza

risultato pari a 11.852.

Nella tavola 4.6 viene riportata la distribuzione di frequenza delle discordanze per tipologia di attribuzione in ANVIS e l'eventuale presenza nei record con almeno un doppione eliminato.

**Tavola 4.6 - Casi discordanti tra ANVIS++ e RBI per fonte di provenienza in ANVIS++ ed eventuale presenza di doppione in LAC (anno 2017) (valori assoluti)**

FONTE DELLA VARIABILE CITTADINANZA IN ANVIS++	Individuo presente in LAC con doppione		Totale
	No	Sì	
Codice cittadinanza valido e concordante in ANVIS e LAC	949	1.319	<b>2.268</b>
Codice cittadinanza valido da LAC	402	219	<b>621</b>
Codice cittadinanza valido da ANVIS	8.754	9	<b>8.763</b>
Codice cittadinanza non valido in LAC e ANVIS e stimato da codice fiscale o luogo di nascita	198	2	<b>200</b>
<b>Totale</b>	<b>10.303</b>	<b>1.549</b>	<b>11.852</b>

Si evidenzia che la maggior parte dei casi discordanti si ha in corrispondenza di codici cittadinanza provenienti esclusivamente da ANVIS (8.763 su 11.852) perché non presenti o non validi in LAC. Mentre la maggior parte dei casi sottoposti al trattamento di eliminazione dei doppi in LAC si trovano in corrispondenza di un codice cittadinanza valido e concordante in ANVIS e LAC (1.319 su 1.549).

Va evidenziato che la presenza di due righe presenti più di una volta nelle LAC con stesso codice SIM è causata da due problemi:

- doppie iscrizioni in anagrafe della stessa persona;
- falsi positivi SIM<sup>4</sup>.

I record con stesso codice SIM presente più di una volta vengono trattati in ANVIS++ e RBI con due algoritmi diversi. In ANVIS++, come detto sopra, si mantiene la riga con la data di iscrizione in anagrafe più recente mentre in RBI si mantiene la riga con codice cittadinanza minore.

Per portare a coerenza i due stock sono stati analizzati in modo integrato i due processi di valorizzazione della cittadinanza. Nei casi di doppi SIM per comprendere l'origine delle differenze si è di nuovo fatto ricorso alle LAC originarie.

La tavola 4.7 mostra, in sintesi, il risultato delle analisi e le scelte effettuate per allineare le cittadinanze in ANVIS++ e RBI.

**Tavola 4.7 - Selezione finale della fonte (con motivazione) per tipologia di valorizzazione in ANVIS++ (anno 2017) (valori assoluti)**

FONTE FINALE SELEZIONATA PER LA CITTADINANZA	FONTE DELLA VARIABILE CITTADINANZA IN ANVIS++				Totale
	Codice cittadinanza vali- do e concordante in ANVIS e LAC	Codice cittadinanza valido da LAC	Codice cittadinanza valido da ANVIS	Codice cittadinanza non valido in LAC e ANVIS e stimato da codice fiscale o luogo di nascita	
ANVIS++ - in RBI cittadinanza stimata	243	21	4.508	-	<b>4.772</b>
RBI - cittadinanza da fonte più robusta	209	65	6	-	<b>280</b>
RBI - cittadinanza da fonte più recente	1.422	389	1.405	76	<b>3.292</b>
ANVIS++ - in RBI cittadinanza non valida	18	13	2.752	117	<b>2.900</b>
ANVIS++ - cittadinanza da fonte più robusta	176	100	90	7	<b>373</b>
ANVIS++ - dopo ritorno LAC su falsi positivi SIM	200	33	2	-	<b>235</b>
<b>Totale</b>	<b>2.268</b>	<b>621</b>	<b>8.763</b>	<b>200</b>	<b>11.852</b>

4 Per falso positivo si intende l'assegnazione dello stesso codice SIM a due persone diverse durante il processo di integrazione SIM.

Nel complesso la numerosità dei casi risolti attribuendo la cittadinanza da ANVIS++ è pari a 8.280, mentre per i rimanenti 3.572 record la cittadinanza è rimasta quella presente in RBI. Come si evince, si è scelto di limitare la numerosità delle cittadinanze stimate favorendo le fonti più robuste o recenti.

Questo lavoro ha inoltre permesso di individuare e correggere un errore nella procedura di attribuzione della cittadinanza in ANVIS++ che causava una sovrastima degli individui “apolidi”.

#### 4.1.3 Confronto della cittadinanza in RBI con dati sulla cittadinanza della rilevazione “Popolazione residente comunale per sesso, anno di nascita e stato civile”

Una valutazione del processo è stata condotta confrontando la distanza tra la distribuzione comunale della popolazione residente per cittadinanza ricostruita dall’algoritmo nel RBI – Registro base degli Individui, e l’omologa distribuzione ufficiale Istat.

In particolare, è stato calcolato l’errore assoluto medio (AARE) percentuale:

$$\text{Average Absolute Relative Error} = \text{AARE}(\hat{\theta}) = \frac{1}{D} \sum_{d=1}^D \left| 1 - \frac{\hat{\theta}_d}{\theta_d} \right|$$

dove  $\hat{\theta}$  è la stima del parametro di riferimento  $\theta$ .

Il Registro Base degli Individui preso in considerazione, è riferito al 31 dicembre 2017. Presenta al suo interno 7.960 comuni<sup>5</sup> e una popolazione totale di 60.539.031 residenti: 55.402.699 italiani e 5.136.303 stranieri distribuiti in 7.926 comuni. Vi sono soltanto 29 casi in cui la cittadinanza non è valida.

La popolazione residente considerata è al 01/01/2018<sup>6</sup>. I comuni presenti in tale rilevazione sono quelli della classificazione ufficiale Istat al 31/12/2017<sup>7</sup>, e parzialmente sono disallineati con quelli del 1 gennaio (es. il comune di Sappada il 01/01/2018 è passato dal Veneto al Friuli).

La popolazione con cittadinanza italiana non è diffusa: per effettuare il confronto è stata calcolata come differenza tra la popolazione totale residente (rilevazione POSAS<sup>8</sup>) e la popolazione straniera residente (rilevazione STRASA<sup>9</sup>).

**Tavola 4.8 - Comuni per classe di differenza relativa tra popolazione residente nel Registro Base degli Individui (RBI) e popolazione residente comunale (POSAS) - 31/12/2017** (valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione RBI		Popolazione POSAS	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	842	10,6	1.370.981	2,3	1.370.981	2,3
]0-1]	6.504	81,8	56.581.343	93,5	56.543.041	93,5
]1-5]	581	7,3	2.519.368	4,2	2.504.552	4,1
]5-10]	15	0,2	29.659	0,0	28.008	0,1
>10	4	0,1	3.428	0,0	3.157	0,0
<b>Totale</b>	<b>7.946</b>	<b>100,0</b>	<b>60.504.779</b>	<b>100,0</b>	<b>60.449.739</b>	<b>100,0</b>

5 L’ammontare dei comuni in RBI coincide col numero di comuni ufficiale del 01/01/2018 poiché il Registro è costruito sulla base della LAC al 01/01/2018.

6 La Popolazione residente comunale per sesso e anno di nascita viene calcolata al 31 dicembre di ogni anno e diffusa al 1° gennaio dell’anno successivo.

7 Elenco comuni, codici, denominazioni ufficiale. Istat. <https://www.istat.it/it/archivio/67899>.

8 Istat “Popolazione residente comunale per sesso anno di nascita e stato civile – POSAS”. <http://demo.istat.it/pop2018/index.html>.

9 Istat “Popolazione residente comunale straniera per sesso anno di nascita e stato civile – STRASA” <http://demo.istat.it/strasa2018/index.html>.

#### 4. Valutazione della procedura di stima della cittadinanza

Per queste ragioni è stato possibile effettuare solo il confronto sui comuni presenti in entrambi gli archivi (pari a 7.946 comuni). In riferimento agli stranieri, i comuni in cui gli stranieri non sono presenti sono raggruppati sotto la modalità “*non presenti*”. La modalità “*non calcolabile*” fa riferimento invece ai comuni per il quale l’indice presenta il denominatore pari a zero.

**Tavola 4.9 - Comuni per classe di differenza relativa tra popolazione italiana residente nel Registro Base degli Individui (RBI) e popolazione italiana residente comunale desunta da POSAS - 31/12/2017**  
(valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione italiana RBI		Popolazione italiana POSAS	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	950	11,9	1.341.157	2,4	1.341.157	2,4
]0-1]	6.359	80,0	51.674.613	93,4	51.626.399	93,4
]1-5]	617	7,8	2.336.739	4,2	2.322.667	4,2
]5-10]	16	0,2	13.264	0,0	12.951	0,0
>10	4	0,1	5.312	0,0	4.779	0,0
<b>Totale</b>	<b>7.946</b>	<b>100,0</b>	<b>55.371.085</b>	<b>100,0</b>	<b>55.307.953</b>	<b>100,0</b>

**Tavola 4.10 - Comuni per classe di differenza relativa tra popolazione straniera residente nel Registro Base degli Individui (RBI) e popolazione straniera residente comunale desunta da STRASA - 31/12/2017**  
(valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione straniera RBI		Popolazione STRASA	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	1.801	22,7	169.876	3,3	169.876	3,3
]0-1]	1.788	22,5	3.257.418	63,5	3.262.823	63,5
]1-5]	2.932	36,9	1.470.577	28,7	1.475.746	28,7
]5-10]	818	10,3	161.700	3,1	162.588	3,1
>10	572	7,2	74.094	1,4	70.734	1,4
Non presenti	32	0,4	0	0,0	0	0,0
Non calcolabile	3	0,0	0	0,0	19	0,0
<b>Totale</b>	<b>7.946</b>	<b>100,0</b>	<b>5.133.665</b>	<b>100,0</b>	<b>5.141.786</b>	<b>100,0</b>

Analizzando i risultati ottenuti per il totale della popolazione residente (Tavola 4.8), le differenze relative tra il registro RBI e il benchmark riguardano soprattutto la classe in cui la differenza è compresa tra 0-1%, in cui ricadono l’81,9% dei comuni e, soprattutto, il 93,5% della popolazione residente. Questo dato risulta confermato se si considerano i soli residenti italiani (Tavola 4.9): l’80% dei comuni e il 93,3% della popolazione residente. Per quanto riguarda gli stranieri (Tavola 4.10), invece, il maggior numero di comuni confluisce nella classe di differenza 1-5%, valida per il 36,9% dei comuni italiani, che però interessa solo il 28,6% della popolazione straniera in RBI e il 18,7% di quella del benchmark: verosimilmente si tratta di comuni a bassa densità abitativa. La percentuale più alta della popolazione, ovvero il 63,5% fa riferimento alla classe di differenza 0-1%, coerente, quindi, sia col risultato totale, sia col risultato nazionale.

#### 4.1.4 Confronto tra il sottoinsieme di RBI la cui cittadinanza è stata desunta da LAC (01/01/2018) con dati sulla cittadinanza nella rilevazione “Popolazione residente comunale per sesso, anno di nascita e stato civile”

I comuni presenti in entrambi gli archivi sono, in questo caso, 7.942 (in RBI ve ne sono alcuni in cui la cittadinanza è stata desunta da archivi alternativi alla LAC). Valgono, inoltre, tutte le osservazioni fatte in precedenza relativamente agli stranieri.

**Tavola 4.11 - Comuni per classe di differenza relativa tra popolazione residente, con cittadinanza desunta da LAC, nel Registro Base degli Individui (RBI) e popolazione residente comunale (POSAS) - 31/12/2017**  
(valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione RBI		Popolazione POSAS	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	688	8,7	898.402	1,5	898.402	1,5
]0-1]	6.549	82,5	56.385.882	93,4	56.423.608	93,4
]1-5]	667	8,4	2.810.728	4,7	2.811.078	4,7
]5-10]	28	0,3	251.798	0,4	262.855	0,4
>10	10	0,1	17.285	0,0	30.718	0,0
<b>Totale</b>	<b>7.942</b>	<b>100,0</b>	<b>60.364.095</b>	<b>100,0</b>	<b>60.426.661</b>	<b>100,0</b>

**Tavola 4.12 - Comuni per classe di differenza relativa tra popolazione italiana residente, con cittadinanza desunta da LAC, nel Registro Base degli Individui (RBI) e popolazione italiana residente comunale desunta da POSAS e STRASA - 31/12/2017** (valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione RBI italiana		Popolazione italiana	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	810	10,2	985.569	1,8	985.569	1,8
]0-1]	6432	81,0	51.661.503	93,5	51.678.583	93,5
]1-5]	667	8,4	2.559.290	4,6	2.567.013	4,6
]5-10]	23	0,3	23.476	0,1	23.877	0,0
>10	10	0,1	18.008	0,0	30.488	0,1
<b>Totale</b>	<b>7942</b>	<b>100,0</b>	<b>55.247.846</b>	<b>100,0</b>	<b>55.285.530</b>	<b>100,0</b>

**Tavola 4.13 - Comuni per classe di differenza relativa tra popolazione straniera residente, con cittadinanza desunta da LAC, nel Registro Base degli Individui (RBI) e popolazione straniera residente comunale (STRASA) - 31/12/2017** (valori assoluti e composizione percentuale)

CLASSE DI DIFFERENZA RELATIVA %	Comuni		Popolazione RBI straniera		Popolazione STRASA	
	Valori assoluti	Percentuale	Valori assoluti	Percentuale	Valori assoluti	Percentuale
0	1.751	22,0	164.367	3,2	164.367	3,2
]0-1]	1.681	21,2	2.732.933	53,4	2.737.485	53,2
]1-5]	2.991	37,6	1.929.319	37,7	1.943.069	37,8
]5-10]	823	10,4	195.856	3,8	200.441	3,9
>10	657	8,3	93.774	1,9	95.319	1,9
Non presenti	32	0,4	0	0,0	0	0,0
Non calcolabile	7	0,1	0	0,0	450	0,0
<b>Totale</b>	<b>7.942</b>	<b>100,0</b>	<b>5.116.249</b>	<b>100,0</b>	<b>5.141.131</b>	<b>100,0</b>

Anche in questo caso, per quanto riguarda il totale della popolazione residente (Tavola 4.11), le differenze relative tra il registro RBI e il benchmark riguardano soprattutto la classe 0-1%, valida per l'82,5% dei comuni e per il 93,4% della popolazione residente. Analogamente, se si considerano i soli residenti italiani (Tavola 4.12) la classe 0-1% comprende l'81% dei comuni e il 93,5% dei residenti. Per quanto riguarda gli stranieri (Tavola 4.13), il maggior numero di comuni fa riferimento alla classe 1-5%, valida per il 37,7% dei comuni italiani, riguardante il 37,7% della popolazione straniera RBI e il 37,8 di quella del benchmark: sono comuni un po' più popolosi rispetto al caso precedente. La percentuale più alta della popolazione, ovvero il 53,4% di RBI e il 53,2 della Rilevazione, invece, fa riferimento alla classe di differenza relativa 0-1%.

## RIFERIMENTI BIBLIOGRAFICI

- Di Zio M., Falorsi P.D., Fortini M. 2016. *Un sistema di registri statistici basato sull'integrazione di informazioni da archivi amministrativi, indagini statistiche e stime di popolazione*. MILeS2016 Milano - Impresa, Lavoro e Società 2016
- Falorsi P.D., Radini R., Scannapieco M., Tosco. 2018. *Modernizing Data Integration Systems at Istat*. European Conference on Quality in Official Statistics, Q2018, Krakov, Poland
- Garofalo G. 2016. *Il Sistema dei registri come strumento di integrazione e miglioramento della qualità dei processi statistici*. 12° Conferenza Nazionale di Statistica.
- Istat, 2018. Codice delle unità territoriali estere. <https://www.istat.it/it/archivio/6747>
- Ministero dell'Interno - Dipartimento per le Libertà Civili e l'Immigrazione, Ufficio Centrale di Statistica. 2014. *Acquisto, concessione e reiezione della cittadinanza italiana*, [ucs.interno.gov.it/FILES/AllegatiPag/1263/DIR\\_INT00018\\_DLC\\_2013.pdf](https://ucs.interno.gov.it/FILES/AllegatiPag/1263/DIR_INT00018_DLC_2013.pdf)
- Radini R., Scannapieco M. e Garofalo G. 2017. *The Italian Integrated System of Statistical Registers: On the Design of an Ontology-based Data Integration Architecture*. In NTTS 2017 Conference.
- Vivio R., Barberis, P., De Gabrieli, MD. 2014. *Il sottoprogetto sperimentale: popolazione insistente*. In Garofalo: Il Progetto ARCHIMEDE obiettivi e risultati sperimentali. Istat Working Paper, 19-35, 9/2014. <https://www4.istat.it/it/files/2014/11/IWP-n.-9-2014.pdf>
- Vivio R. 2016. *Il Progetto Archimede. Le potenzialità per le analisi territoriali*. XII Conferenza nazionale di Statistica, 22-24 giugno 2016, Roma. <https://www.istat.it/it/dodicesima-conferenza/programma/032-progetto-archimede>
- Vivio, R., S. Casacci, S. Di Domenico, M. Ferraro. 2019. City users and daytime population. An approach with administrative data. In NTTS 2019 Conference, 11-16 march 2019, 162, Bruxelles. [https://ec.europa.eu/eurostat/cros/system/files/ntt2019\\_book\\_of\\_abstracts.pdf](https://ec.europa.eu/eurostat/cros/system/files/ntt2019_book_of_abstracts.pdf)

